



# Traffic-aware scheduling and feedback reporting in wireless networks

Matha Deghel

## ► To cite this version:

Matha Deghel. Traffic-aware scheduling and feedback reporting in wireless networks. Autre. Université Paris Saclay (COMUE), 2017. Français. NNT : 2017SACLC030 . tel-01614259

**HAL Id: tel-01614259**

**<https://theses.hal.science/tel-01614259>**

Submitted on 10 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT  
DE  
L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE À  
CENTRALESUPÉLEC

ECOLE DOCTORALE N°580

Sciences et technologies de l'information et de la communication (STIC)

Spécialité : Réseaux, Information et Communications

Par

**M. Matha Deghel**

Traffic-Aware Scheduling and Feedback Reporting in Wireless  
Networks

Thèse présentée et soutenue à Gif-sur-Yvette, le 22 Mai 2017 :

**Composition du jury :**

<b>M. Michel Kieffer,</b>	Université Paris-Sud	Examineur, Président du Jury
<b>M. Vincent K. N. Lau,</b>	Hong Kong University	Rapporteur
<b>M. Jean-Marie Gorce,</b>	INSA-Lyon	Rapporteur
<b>M. Anthony Ephremides,</b>	University of Maryland	Examineur
<b>M. Jianwei Huang,</b>	Chinese University of Hong Kong	Examineur
<b>Mme Laura Cottatellucci,</b>	Eurecom	Examinatrice
<b>M. Mérouane Debbah,</b>	CentraleSupélec, Huawei	Directeur de Thèse
<b>M. Mohamad Assaad,</b>	CentraleSupélec	Co-Directeur de Thèse

**Title:** Traffic-aware scheduling and feedback reporting in wireless networks

**Keywords:** Queueing stability, scheduling, feedback allocation, wireless networks, 5G

**Abstract:** Demand of wireless communication systems for high throughputs continues to increase, and there are no signs this trend is slowing down. Three of the most prominent techniques that have emerged to meet such demands are OFDMA, cooperative relaying and MIMO. To fully utilize the capabilities of systems applying such techniques, it is essential to develop efficient scheduling algorithms and, more generally, efficient resource allocation algorithms. Classical studies on this subject investigate in much detail settings where the data requests of the users are not taken into consideration or where the perfect and full CSI is assumed to be available for the scheduling mechanism. In practice, however, different limitations may result in not having perfect or full CSI knowledge, such as limited feedback resources, probing cost and delay in the feedback process.

Accordingly, in this thesis we examine the problems of scheduling and feedback allocations under realistic considerations concerning the CSI knowledge. Analysis is performed at the packet level and considers the queueing dynamics in the systems with arbitrary arrival processes, where the main performance metric we adopt is the stability of the queues. The first part of the thesis considers a multi-point to multi-point MIMO system with TDD mode under limited backhaul capacity and taking into account the feedback probing cost. Regarding the interference management technique, we apply interference alignment (IA) if more than one pair are active and SVD if only one pair is active. The second part of the thesis considers a multiuser multichannel OFDMA-like system where delayed and limited feedback is accounted for. Two scenarios are investigated, namely the system without relaying and the system with relaying. For the latter one, an additional imperfection we account for is that the users have incomplete knowledge of the fading coefficients between the base-station and the relay.

# Acknowledgments

This thesis represents the culmination of a lot of work and effort, the majority of which would not have been possible without the advice, guidance and support of many people.

First and foremost, I would like to express my sincerest gratitude to my supervisor, Mohamad Assaad. To me, he is an ideal supervisor, and I feel extremely fortunate to be able to work with him. He is passionate, generous, rigorous and highly supportive. He tried his best to help me get around the many difficulties I faced. I cannot thank him enough.

The second person I want to thank is my co-supervisor, Mérouane Debbah. He is one of the most generous and enthusiastic persons I have ever met. He has given me enormous encouragement and help throughout the years. My gratitude to him is beyond expression.

I would also like to thank all the jury members, Vincent K. N. Lau, Jean-Marie Gorce, Anthony Ephremides, Jianwei Huang, Michel Kieffer and Laura Cottatellucci, for their participation in my PhD defense and for their valuable feedback on several aspects of it. Additionally, I would like to thank my coauthors, Anthony Ephremides and Ejder Bastug, with whom I had many technical discussions that helped progress my research.

For providing a great environment to come to every day, a huge thanks to all my friends and colleagues in CentraleSupélec. I would furthermore like to thank all my friends outside of Centrale-Supélec, who have stayed in touch with me and supported me during this challenging journey.

Finally, none of this would have been possible at all without the constant love and support from all the members of my family. They have been believing in me and gave me the courage to try out things that I myself was not sure if I would be able to accomplish. I am eternally grateful for them.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgments</b>	<b>ii</b>
<b>Acronyms</b>	<b>vii</b>
<b>General Notation</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background and Motivation . . . . .	1
1.2 Thesis Outline and Contributions . . . . .	7
1.3 Publications . . . . .	8
<b>2 Mathematical Tools for Queueing Analysis</b>	<b>11</b>
2.1 Different Forms of Queueing Stability . . . . .	12
2.2 Stability Region and Optimal Scheduling Policy . . . . .	12
2.3 Stability Optimal Scheduling Algorithms . . . . .	14
2.3.1 Max-Weight Policy . . . . .	15
2.3.2 Strong Stability Proof for the Max-Weight Policy . . . . .	16
2.3.3 Scheduling with Known Mean Arrival Rates . . . . .	19
2.4 Suboptimal or Approximate Scheduling . . . . .	19
<b>3 Queueing Stability and CSI Probing of a TDD Wireless Network with Interference Alignment</b>	<b>21</b>
3.1 Overview . . . . .	21
3.2 System Model . . . . .	24
3.2.1 Interference Alignment Technique . . . . .	25
3.2.2 CSIT Sharing Over Limited Capacity Backhaul Links . . . . .	27
3.2.3 Rate Model and Impact of Training . . . . .	28
3.2.4 Queue Dynamics, Stability and Scheduling Policy . . . . .	30
3.3 Success Probabilities and Average Rates . . . . .	32

## Contents

---

3.3.1	Average Rate Expression for IA . . . . .	32
3.3.2	SVD Scheme and its Average Rate Expression . . . . .	34
3.4	Stability Analysis for the Symmetric Case . . . . .	36
3.4.1	Average Rate Expressions and their Variation . . . . .	37
3.4.2	Stability Regions and Scheduling Policies . . . . .	39
3.4.3	Conditions under which IA Provides a Gain in terms of Queueing Stability	41
3.4.4	Comparison between the Imperfect and Perfect Cases in terms of Queueing Stability . . . . .	43
3.4.5	Impact of the Number of Bits $B$ on the Stability Region . . . . .	45
3.5	Algorithmic Design and Performance Analysis for the General Case . . . . .	45
3.5.1	Average Rate Expressions . . . . .	46
3.5.2	Stability Regions and Scheduling Policies . . . . .	47
3.5.3	$\beta_A$ -Approximate Scheduling Policy . . . . .	48
3.5.4	Compare the Imperfect and Perfect Cases in terms of Queueing Stability .	51
3.6	Stability Analysis with Multiple Rate Levels . . . . .	52
3.6.1	Average Rate Expressions and their Variation . . . . .	53
3.6.2	Stability Regions and Scheduling Policies . . . . .	55
3.6.3	Conditions under which IA Provides a Gain in terms of Queueing Stability	56
3.6.4	Compare the Imperfect and Perfect Cases in terms of Queueing Stability .	57
3.7	Validation of the Proposed Model . . . . .	57
3.8	Closing Remarks . . . . .	60
3.9	Appendix . . . . .	61
3.9.1	Proof of Proposition 1 . . . . .	61
3.9.2	Proof of Theorem 4 . . . . .	61
3.9.3	Proof of Proposition 2 . . . . .	63
3.9.4	Proof of Lemma 2 . . . . .	64
3.9.5	Proof of Lemma 3 . . . . .	65
3.9.6	Proof of Theorem 5 . . . . .	67
3.9.7	Proof of Proposition 4 . . . . .	68
3.9.8	Proof of Theorem 7 . . . . .	70
3.9.9	Proof of Theorem 8 . . . . .	72
3.9.10	Derivation of Rate Approximation 1 . . . . .	78
3.9.11	Derivation of Rate Approximation 2 . . . . .	79
3.9.12	Proof of Theorem 9 . . . . .	80
3.9.13	Proof of Theorem 10 . . . . .	84
3.9.14	Proof of Lemma 6 . . . . .	86
<b>4</b>	<b>Opportunistic Feedback Reporting and Scheduling Scheme for Multichannel Wireless Networks</b>	<b>87</b>
4.1	Overview . . . . .	87
4.2	System Model . . . . .	90
4.2.1	System without Relaying . . . . .	91
4.2.2	System with Relaying . . . . .	92

## Contents

---

4.2.3	Limited and Delayed CSI Knowledge Scheme . . . . .	95
4.2.4	Queueing Model . . . . .	96
4.3	Proposed Algorithms and Stability Analysis for the System without Relaying . . . .	98
4.3.1	Algorithm FSA and its Stability Performance . . . . .	99
4.3.2	Algorithm FMA and its Stability Performance . . . . .	104
4.4	Proposed Algorithms and Stability Analysis for the System with Relaying . . . . .	108
4.4.1	Case without Delay . . . . .	108
4.4.2	Case with Delay . . . . .	113
4.5	Validation of the Proposed Model . . . . .	114
4.6	Closing Remarks . . . . .	119
4.7	Appendix . . . . .	120
4.7.1	Proof of Theorem 14 . . . . .	120
4.7.2	Proof of Theorem 15 . . . . .	129
4.7.3	Proof of Corollary 1 . . . . .	135
<b>5</b>	<b>Conclusions and Perspectives</b>	<b>141</b>
5.1	Conclusions . . . . .	141
5.2	Perspectives . . . . .	142
	<b>Bibliography</b>	<b>143</b>

## Contents

---



# Acronyms

AF	Amplify-and-Forward. 92, 93
BC	broadcast channel. 22
BS	base station. 1–5, 7, 8, 87–114, 117–119, 121, 124, 126, 132, 141–143
CDF	cumulative distribution function. 28
CS	central scheduler. 23, 27, 31
CSI	channel state information. 2, 3, 5–8, 22, 23, 26–31, 33, 59, 87–90, 95–101, 105, 109, 110, 119, 121, 141, 142
DF	Decode-and-Forward. 92, 93
DoF	degree-of-freedom. 22, 25
FDD	frequency-division duplex. 3, 87, 88, 91
IA	interference alignment. 2, 5, 7, 21–23, 25–29, 32, 33, 35, 36, 40–46, 52, 56, 57, 59, 60, 68, 69, 141, 143
IC	interference channel. 22
ISI	inter-stream interference. 26, 28, 29, 34, 35
IUI	inter-user interference. 26, 28, 29
LTE	long term evolution. 87, 88
MGF	moment-generating function. 33, 34, 61–63
MIMO	multiple-input multiple-output. 1, 2, 6, 7, 21–25, 34, 35, 60, 141
MRC	maximum ratio combining. 92
OFDM	orthogonal frequency division multiplexing. 88
OFDMA	orthogonal frequency-division multiplexing access. 2, 3, 5, 6, 87
PDF	probability distribution function. 61, 62, 137

## Acronyms

---

SINR	signal-to-interference-plus-noise ratio. 29, 30, 33, 36–38, 45, 46, 52–54, 57, 64
SNR	signal-to-noise ratio. 7, 25, 26, 29, 35, 89, 91, 94–96
SVD	singular value decomposition. 2, 5, 7, 22, 23, 25, 30, 32, 35, 36, 38, 40, 42, 46, 52, 54, 57, 59, 60, 68, 69, 141
TDD	time-division duplex. 2, 3, 7, 21–23, 27, 30, 60, 87
TDMA	time-division multiple access. 57
ZF	zero forcing. 2, 22, 23

# General Notation

$\mathbb{P}$	Probability measure
$\mathbb{E}$	Expectation operation
$\mathbf{I}_N$	Identity matrix of size $N$
$ \cdot $	Absolute value for scalars and cardinality for sets or subsets
$\ \cdot\ _1$	Norm of the first degree
$\ \cdot\ $	Norm of the second degree
$(\cdot)^T$	Transpose of $(\cdot)$
$(\cdot)^H$	Conjugate transpose of $(\cdot)$
$\mathcal{CH}$	Convex hull
$\mathcal{CN}(\mu, \sigma^2)$	Complex normal random variable with mean $\mu$ and variance $\sigma^2$
$\mathbb{R}_+^N$	The space of $N$ -dimensional vectors with nonnegative real entries
$\Gamma(\cdot)$	Gamma function: $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$
$\Gamma(\cdot, \cdot)$	Upper incomplete Gamma function: $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$
$\text{Bet}(\cdot, \cdot)$	Beta function: $\text{Bet}(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$
$\Psi(\cdot, \cdot; \cdot)$	Kummer function: $\Psi(a, b; x) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt} t^{a-1} (1+t)^{b-a-1} dt$
${}_2F_1(\cdot, \cdot; \cdot; \cdot)$	Hypergeometric function: ${}_2F_1(a, b; c; z) = \sum_{n=0}^\infty \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$ , where $(q)_n$ is the Pochhammer symbol: $(q)_n = 1$ for $n = 0$ , $(q)_n = q(q+1) \dots (q+n-1)$ for $n > 0$
$\mathbb{1}(\cdot)$	Indicator function
$\limsup$	Limit superior
$\lceil x \rceil$	Least integer greater than or equal to $x$
$\otimes$	Kronecker product
$\mathcal{O}(\cdot)$	Big O notation
$\prec$	Component-wise inequality operator
$\triangleq$	Defined as
$\sim$	Distributed as



# List of Figures

2.1	The convex hull of the set of points $\{(0, 0), (0, 0.5), (0.5, 0), (0.25, 0.5), (0.5, 0.25)\}$ represents the system stability region of the example. . . . .	14
3.1	A sketch of an $N$ -user MIMO interference channel with limited backhaul. . . . .	25
3.2	Global CSI knowledge at transmitter $k$ under the imperfect case. . . . .	28
3.3	Stability region (in gray) for the symmetric system under the imperfect case, where $N = 2$ . (a) IA provides a queueing stability gain, and (b) IA does not provide a queueing stability gain. . . . .	42
3.4	Stability regions of the perfect (dotted region) and imperfect (gray region) cases for the symmetric system, where $N = 2$ . (a) IA provides a queueing stability gain, and (b) IA does not provide a queueing stability gain. . . . .	44
3.5	Total average queue length vs. mean arrival rate $a$ . Here, $\zeta_c = 0.2$ and $\tau = 1$ . . . . .	58
3.6	Total average queue length vs. mean arrival rate $a$ . Here, $\zeta_c = 0.4$ and $\tau = 1$ . . . . .	58
3.7	Achievable fraction $\frac{r(N, B')}{r(N, B)}$ vs. number of bits $B'$ . Here, $\zeta_c = 1$ and $B = 40$ . . . . .	59
3.8	Achievable fraction $\beta_P$ vs. number of bits $B$ . . . . .	60
3.9	Example that illustrates the result of Lemma 4. . . . .	67
3.10	A tree that shows the vectors in $S_2$ that yield $\mathbf{s}_{i,4} = (1, 1, 1, 1)$ . Here, $N = 4$ and $n = 2$ . . . . .	71
4.1	Multi-channel multi-user queueing system. . . . .	90
4.2	System model without relaying. . . . .	91
4.3	System model with relaying. . . . .	92
4.4	Total average queue length vs. mean arrival rate $a$ . Here, the velocity $v = 1$ km/hr. . . . .	116
4.5	Total average queue length vs. mean arrival rate $a$ . Here, the velocity $v = 5$ km/hr. . . . .	116
4.6	Total average queue length vs. mean arrival rate $a$ . Here, the velocity $v = 10$ km/hr. . . . .	116
4.7	Total average queue length vs. mean arrival rate $a$ . Here, the velocity $v = 50$ km/hr. . . . .	116
4.8	Minimum achievable fraction $\beta_r$ vs. threshold $\delta$ . . . . .	117
4.9	Total average queue length vs. mean arrival rate $a$ . Here $\tau = 5$ . . . . .	118

## List of Figures

---

# List of Tables

3.1	List of simulation parameters. . . . .	58
4.1	List of simulation parameters for the system without relaying . . . . .	114
4.2	Possible rates used for the simulations. . . . .	115
4.3	List of simulation parameters for the system with relaying . . . . .	117

## List of Tables

---



# Chapter 1

## Introduction

### 1.1 Background and Motivation

Wireless and cellular systems currently experience an exponential growth in the demand for network traffic as more and more consumers want to have constant wireless connectivity and access to streaming services; global mobile data traffic grew 63% percent in 2016, and this traffic is expected to increase sevenfold between 2016 and 2021 [1]. Next generation networks are expected to provide significantly higher data rates, spectral efficiency, and reliability than the current systems. In order to meet the increasing demand in data rates (that are currently being supported by high speed wired networks), and more generally to provide performance improvement, it is essential to fully exploit the capacity available in wireless and cellular systems, as well as to develop robust strategies for integrating these systems into a large scale, heterogeneous data network.

To this end, a vital step is to better understand and exploit the physical layer capabilities and to consider the effects that different physical layer characteristics have upon network design and control. In this regard, multiple physical-layer strategies have emerged and are shown to provide substantial performance improvement, where four of the most prominent ones are detailed next.

#### Multiple Antennas

The use of multiple antennas offers many benefits in practical wireless communication systems [2, 3]. For example, single user multiple-input multiple-output (MIMO) systems exploit multiple transmit and receive antennas to improve capacity, reliability, and resistance to interference. Such benefits can also be offered in a multiuser MIMO (MU-MIMO) system, where a base station (BS) communicates with multiple users or where there are multiple transmitters each of which communicates with its respective receiver. MU-MIMO on the downlink is especially interesting because the MIMO sum capacity can scale with the minimum of the number of BS antennas and the sum of the number of users times the number of antennas per user. This means that MU-MIMO can achieve MIMO capacity gains with a multiple antenna BS and a bunch of single antenna users. To fully exploit the

benefits offered by using multiple antennas, accurate knowledge of the channel state information (CSI) of the users is necessary at the BS. For systems employing time-division duplex (TDD) mode, one prominent technique, which exploits the uplink-downlink reciprocity of radio propagation, is to let the users send orthogonal training sequences to the BS so that it can estimate their CSIs. This technique has the important advantage that the length of the training sequences is independent of the number of antennas. This advantage has led to the concept of massive MIMO (also called large-scale MIMO), which is a special case of MU-MIMO where the BS is equipped with a large number of antennas (on the order of hundreds to thousands) [4, 5]. On the other hand, the length of the training sequences depends on the number of active users in the system and it scales proportionally with this number. Recalling that systems are slotted and that the period of a slot is limited (since the coherence time is limited), one can notice the trade-off that arises here: the more users are scheduled (i.e. more active users) in a slot, the less time is reserved for transmission. Thus, it is of great importance to design an efficient scheduling policy to select the subset of active pairs at each time-slot, where obviously one of the things this policy should exploit efficiently is the trade-off outlined before.

### Interference Management Techniques for MIMO Systems

The use of multiple antennas allows transmissions in the same time-frequency block. However, interference is a major issue in this case, which can substantially decrease the system performance unless treated properly. In this regard, several interference management techniques (i.e. transmission schemes) have been proposed and investigated. For instance, for a single user MIMO system, singular value decomposition (SVD) technique was shown to provide the best performance among other techniques such as zero forcing (ZF) and matched filtering (MF) [6]. In the context of a multi-user MIMO system, especially the one that considers multiple transmitters each of which communicating with its respective receiver, interference alignment (IA) was shown to be an efficient interference management technique [7]. IA uses the spatial dimension offered by multiple antennas for alignment. The key idea is to design the transmission scheme in such a way as to ensure that the interference signal lies in a reduced dimensional subspace at each receiver. Other efficient transmission schemes exist and can be considered in the multi-user MIMO context, such as ZF and regularized zero-forcing (RZF) (which is also sometimes called minimum mean square error (MMSE) precoding) [8]. Obviously, the interference management technique affects the scheduling decisions. This observation results from the facts that (i) the achievable service rates depends on the transmission scheme that is adopted, and (ii) any good user-scheduling policy should take these rates into consideration.

### Parallel Channels

In the frequency domain, creating parallel channels can be achieved by using orthogonal frequencies. Such a technique can be found in the latest standards for cellular communication and is known under the name of orthogonal frequency-division multiplexing access (OFDMA) [9]. In the space domain, for a MIMO system we can create parallel channels by using orthonormal beamforming. This technique can also be found in the latest cellular communication standards [9]. Under both cases, namely OFDMA and MIMO with orthonormal beamforming, acquiring the CSI can be done under

the TDD mode, where more time resources are necessary if more CSIs are decided to be reported. This acquisition can be also done under the frequency-division duplex (FDD) mode, where more frequency resources are necessary if more CSIs need to be fed back. Note that under the latter mode there is no channel reciprocity, so to acquire CSI, at the first stage the BS sends training sequences that permit the users to estimate their channel states, then the users (that are selected for the feedback) can then report their states to the BS. Since reporting more feedback requires more resources (i.e. more overhead in the system), which leaves less resources for transmissions, an important trade-off arises here. Clearly, efficient user scheduling and feedback reporting policies should exploit this trade-off.

### Cooperative Relaying

Cooperative communication techniques have emerged as a means of improving the performance of wireless networks. These techniques take advantage of the broadcast nature of wireless communications and spatial diversity to improve the system performance. In particular, cooperative relaying was introduced as a very promising approach to provide throughput gains as well as coverage extension [10, 11]; relaying is one of the features being proposed for the 4G LTE Advanced system [12]. In this approach, relay nodes are used to forward the replica of packets from the source node, and the destination node can combine multiple copies of the signal to better decode the original message. Several cooperative relaying protocols are proposed in the literature, among which the most prominent are: (i) Amplify-and-Forward, under which the relay acts as an analog repeater, (ii) Decode-and-Forward, where the relay decodes, encodes and retransmits the received message, and (iii) Decode-and-Reencode, for which the main idea is that the relay decodes the received message, but constructs a codeword differing from the source codeword. Combining cooperative relaying and OFDMA was shown to deliver high data rates requirements [13, 14], particularly for users at the cell edge, however the feedback process would result in a prohibitive overhead in such systems. In this regard, it is important to investigate the performance of cooperative relaying under incomplete feedback information conditions and to develop scheduling and resource allocation algorithms that adapt to these conditions.

### Interaction between the physical layer and the MAC layer

As alluded earlier, another important step towards providing better system performance is a careful cross-layer design. The importance of such a design can be seen, for example, in multi-node wireless communication systems with interference properties. In such systems, the communication links between pairs of nodes cannot be viewed independently but rather as interacting entities where the service rate of one pair is a function of choices and decisions for the access and physical layer parameters of the others. Another complexity of wireless communication systems is the fact that the channel might be changing in time, where this variation might be happening at the scale of milliseconds in the case of fast fading. Designs of different layers need to take this variability into account in such a way as to make sure that the system optimally compensates it.

In this thesis, the interaction between the physical layer and media access control (MAC) layer is of particular interest.

## 1.1. Background and Motivation

---

The MAC layer is responsible for controlling how devices in a network gain access to medium and permission to transmit, that is to say, it determines the scheduling decision at each time-slot; such a decision is part of the network control mechanism which determines the access control vector and the traffic forwarding decisions. As alluded earlier, for a scheduling policy to be efficient, it is essential that it takes into consideration physical layer parameters and properties. Typical physical layer functions include power control, precoding, and selection of the modulation constellation.

Motivated by the above observations, in this thesis we develop policies for dynamic scheduling and resource allocation for wireless systems. Analysis is done at the packet level and considers the complete dynamics of stochastic arrivals, i.e. the traffic patterns are taken into consideration. Arrivals for each user are stored in a respective queue (i.e. buffer) at the controller (e.g. base-station). Dynamics of *queue lengths* (also sometimes called *queue backlogs*) are also considered in the analysis. Indeed, it is shown that queue length information is important in the design of robust scheduling policies, or more generally robust network algorithms, which yield high data rates and low packet delays in the presence of time varying channels and changing user demands.

The primarily goal of the scheduling mechanism in this thesis is to stabilize the system and thereby achieve maximum throughput and maintain low packet delay. In broad terms, a system is called *stable* if all the queue lengths are finite (in the mean) [15, 16]. Stability performance of a scheduling policy can be characterized by the *stability region* it can achieve. This region is defined as the collection of all traffic load matrices that are sustainable by the policy, i.e. set of mean arrival rate matrices for which the system stays stable under the policy. The stability region of a specific policy should be distinguished from the *system stability region* (also called network capacity region). The latter being (i) sometimes defined as the union of all the individual policy stability regions, taken over all possible control policies, such as scheduling and power allocation, and (ii) other times defined as the union of all the individual policy stability regions, taken (only) over all possible scheduling allocation policies. It is worth mentioning that the stability region of a system is distinct from the information theoretic capacity of this same system, where the latter includes mainly, but not only, optimization over all possible coding schemes [17]. Furthermore, a policy is called *throughput optimal* (or just optimal) if its stability region coincides with the system stability region.

To fully exploit a wireless network, we need to investigate its fundamental performance and develop efficient and low-complexity control policies. The key challenge is to accommodate time variations of channels (due to mobility, multi-path fading, etc.) and to adapt to other constraints such as channel interference or limited feedback resources. It is well known that *opportunistic scheduling* [18] is an efficient approach to improve the network performance, where, under this scheduling, service priority is given to users with favorable channel conditions. One example is the *Max-Weight* policy, which is, under some assumptions, known to be throughput optimal [19–22]. This policy uses instantaneous queueing information and channel state information to schedule data transmissions, and is shown to support any achievable network throughput without the knowledge of data arrival statistics. Many previous contributions on opportunistic scheduling make some idealistic assumptions; for example, in a time-slotted wireless network, channel states are assumed to be fully and instantaneously available at the controller (e.g. BS) with negligible overheads. These assumptions, however, may not be realistic. Moreover, Max-Weight policies may not be throughput optimal without these assumptions. Motivated by the above considerations and observations, in this thesis we adopt more realistic assump-

## 1.1. Background and Motivation

---

tions and study their impact on the design and performance of the scheduling policy. Assumptions regarding the feedback process and the overhead this process may incur are of particular interest. Under such assumptions, one can see the importance of designing efficient feedback allocation algorithms. Indeed, such algorithms directly impact the scheduling mechanism since generally a user cannot be scheduled for transmission unless its channel state was reported to the BS. Therefore, in this thesis we are also interested in developing feedback allocation algorithms and to investigate their impacts on the scheduling mechanism.

The first system we consider in this thesis is a multi-point to multi-point MIMO system with TDD mode, where each transmitter communicates with its corresponding receiver. Two realistic limitations we account for are: (i) the backhaul that connects the transmitters to each other is of limited capacity, and (ii) probing a channel consumes a fraction of the time-slot. From the set of transmitter-receiver pairs present in the system, at each time-slot only a subset of these pairs will be scheduled (i.e. active), and this subset is the result of a certain scheduling policy. If more than one pairs are scheduled, we use IA as an interference management technique, while we use SVD technique if only one pair is scheduled. In the former case, applying IA requires that the transmitters share their local CSI knowledge with each other over the backhaul. Since the backhaul is of limited capacity, each transmitter should quantize its local CSI to be able to send it to other transmitters, and this leads to imperfect global CSI knowledge at each transmitter. Note that all these considerations make the physical layer of the system a very complex entity.

For the above system, designing an efficient scheduling policy that, in addition to taking into consideration the traffic patterns, exploits and accounts for the physical layer properties is significantly important (but very challenging). Specifically, the trade-off between having more scheduled pairs, which implies higher probing cost (thus less time dedicated for transmission), and less scheduled pairs, which implies lower probing cost (hence more time for transmission), should be exploited by this policy in an optimal manner. Furthermore, characterizing the conditions under which applying IA can deliver performance improvements with respect to using SVD is also very important.

The second system we consider is a multichannel multiuser OFDMA-like system where delayed and limited feedback is accounted for. Two scenarios are considered, namely one without relaying and the other with relaying. For the latter scenario, an additional limitation we account for is that the fading coefficients between the BS and the relay are not perfectly known at the users. It should be noted that the delay in the feedback process depends on the amount of feedback resources.

An important trade-off arises in this system between having more reported feedback but which is less accurate (since there is more delay in the feedback process) and having less reported feedback but which is more accurate (since there is less delay in the feedback process). When selecting the amount of feedback resources, this trade-off should be exploited efficiently. Note that the feedback reporting decisions directly impact the scheduling mechanism. Thus, designing efficient joint feedback reporting and scheduling schemes that, in addition to the traffic patterns, account for the different limitations in the system is of particular interest. Additionally, because of these limitations, any such schemes cannot be optimal, so it is important to characterize the performance that the scheme can guarantee.

## Related Work

In this thesis, scheduling and resource allocation decisions are made in the presence of time varying channel conditions and stochastic packet arrivals. The main objective of the scheduling policy design is to achieve a maximum stability region. A scheduling policy that achieves the entire system stability region is called throughput optimal (or simply optimal). The concept of stability-optimal operation comes originally from the control and automation theory [23–26]. It was applied to the wireless communication systems first in [19,20], where the authors have proposed the opportunistic Max-Weight scheduling strategy and have shown that it can guarantee stability of the user buffers (whenever this is possible). This scheme was then extended by some bounds and generalized to adapt to various network scenarios in [27–29]. Since then, this concept has been the subject of extensive research in the wireless communication framework under various network and traffic scenarios.

For wireless systems where complete and perfect channel state information is available at the scheduler and where traffic patterns are taken into consideration, there has been much work in developing scheduling policies for various performance metrics that include stability, packet-delay guarantees and utility maximization [20,27,30–36]. In practice, however, several limitations may result in not having complete or perfect CSI knowledge at the scheduler, such as limited feedback resources, delay in the feedback process and estimation error. Because of such limitations, it may not be possible to develop optimal scheduling policies, hence the importance of characterizing the stability region the proposed policy guarantees to achieve with respect to the system stability region [15,16].

In the context of limited feedback resources, in addition to the design of an efficient scheduling policy, the feedback algorithm needs to be carefully designed since it directly impacts the scheduling mechanism. For single-channel networks where at most one user can be scheduled on the channel, joint feedback and scheduling policies have been proposed and analyzed (from a queueing stability point of view) in [37–42]. Applying the feedback-scheduling algorithms for single-channel networks directly to a multichannel network by treating the multichannel network as multiple single-channel networks is not a good approach in general, especially under the context of limited feedback resources. There is a paucity of literature addressing the design of feedback-scheduling algorithms and studying their stability performance for multichannel networks with such a feedback limitation [43–45]. We note that, however, many studies have addressed the problem of scheduling policies for multichannel systems under the assumption of full (and perfect) feedback knowledge [46–50]. It is worth noting that for multichannel systems where the impact of bursty traffic is not accounted for, resource allocation problems have been the subject of extensive research [51–55]. On the other hand, taking into consideration another kind of feedback limitation, [56–58] study scheduling problems under delayed channel state information.

Analogous to wireless systems where there is only one transmission over a channel (e.g. OFDMA-based systems), queue length-based policies stabilizing the buffers can be designed for MIMO systems where multiple transmissions are allowed in the same time-frequency block [59–68]. In the latter systems, we note that the bit-rate of each active user depends on the channel states of all other active users, thus this user cannot simply estimate its bit rate using its current channel state; this makes the stability analysis highly complicated for these systems.

Further, identifying scheduling strategies for multi-hop systems have been addressed in [69–74]. Such studies are usually made under the assumption that full and perfect CSI is available for the scheduling mechanism, since considering any kind of feedback limitation would add a considerable complexity to the stability analysis for these systems.

Finally, decentralized scheduling algorithms have been investigated in [75–79]. Most of these studies adopt a simple physical layer so as to keep the stability analysis tractable. Analysis considering a decentralized scheduling with a relatively more involved physical layer can be found in [64, 68].

## 1.2 Thesis Outline and Contributions

The main contributions and the outline of the thesis are as follows.

**In Chapter 2**, we introduce the notion of queue stability and develop the theoretic tools necessary to analyze wireless networks with time varying rates and bursty traffic.

**In Chapter 3**, we characterize the performance in terms of queueing stability of a network composed of multiple MIMO transmitter-receiver pairs taking into account the dynamic traffic pattern and the probing/feedback cost. We adopt a centralized scheduling scheme that selects a number of active pairs in each time-slot. We consider that the system applies IA technique if two or more pairs are active, whereas SVD is used in the special case where only one pair is active. A TDD mode is adopted where transmitters acquire their CSI by decoding the pilot sequences sent by the receivers. Since global CSI knowledge is required for IA, the transmitters have also to exchange their estimated CSIs over a backhaul of limited capacity (i.e. imperfect case). Under this setting, we characterize the stability region of the system under both the imperfect and perfect (i.e. unlimited backhaul) cases, then we examine the gap between these two resulting regions. Further, under each case we provide a centralized probing policy that achieves the max stability region. These stability regions and scheduling policies are given for the symmetric system, where all the path loss coefficients are equal to each other, as well as for the general system. For the symmetric system, we provide the conditions under which IA yields a queueing stability gain compared to SVD. Under the general system, the adopted scheduling policy is of a high computational complexity for moderate numbers of pairs, consequently we propose an approximate policy that has a reduced complexity and that guarantees to achieve a fraction of the system stability region. A characterization of this fraction is also provided. Finally, in the same vein as the symmetric case with single rate level, a stability analysis for the symmetric case under a multiple rate levels scheme is given.

**In Chapter 4**, we address the problem of feedback allocation and scheduling for multiuser multichannel downlink cellular network under limited and delayed feedback. We consider two scenarios: (i) system without relaying, and (ii) system with relaying. For the latter system, an additional imperfection is taken into consideration: the users have incomplete knowledge of the fading coefficients between the BS and the relay. More specifically, we assume that the only information a user knows about any of these fading coefficients is if its corresponding signal-to-noise ratio (SNR) is higher than or equal to a certain threshold. For each system, we propose an efficient joint feedback allocation

and scheduling algorithm, in which the decisions are made at the users side and where the required amount of feedback resources is exactly equal to the number of channels in the system. This algorithm is shown to achieve good stability performance with respect to the ideal system, however it is suitable for a continuous-time contention scheme. We thus propose a second algorithm that is adapted for a discrete-time contention scheme, which adopts a threshold-based concept and is designed in such way as to imitate the first one as much as possible. Under the second algorithm, the feedback decision is done at the users side, and the BS uses this feedback to perform scheduling. Regarding the choice of the amount of feedback resources for this algorithm, we find the best trade-off between having more feedback resources (thus, more knowledge at the scheduler) but longer delay (hence, less accurate CSI) and having less feedback resources but low delay; these results are given for various system setups, i.e. different values of users velocity. For the system with relaying, we further investigate the special case where the delay in the feedback process is not accounted for. Specifically, we develop a joint feedback and scheduling algorithm, we analyze its performance and we characterize the fraction this algorithm guarantees to achieve with respect to the stability region of the ideal system.

In Chapter 5, we conclude the thesis and present possible extensions of the results and future research directions.

## 1.3 Publications

The following publications were produced during the course of this thesis.

### Journal Articles

- [80] M. Deghel, M. Assaad, M. Debbah, and A. Ephremides, "Queueing Stability and CSI Probing of a TDD wireless Network with Interference Alignment", IEEE Transactions on Information Theory, 2017.
- [81] M. Deghel, M. Assaad, M. Debbah, and A. Ephremides, "Traffic-Aware Scheduling and Feedback Allocation in Multichannel Wireless Networks", (*submitted to*) IEEE Transactions on Wireless Communications, 2017.
- [82] M. Deghel, M. Assaad, M. Debbah, and A. Ephremides, "Traffic-Aware Scheduling and Feedback Reporting in Multichannel Wireless Networks with Relaying", (*submitted to*) IEEE Transactions on Information Theory, 2017.

### Conference Papers

- [83] M. Deghel, M. Assaad, and M. Debbah, "Opportunistic Feedback Reporting and Scheduling Scheme for Multichannel Wireless Networks", 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 2016, pp. 1-7.



### 1.3. Publications

---

- [84] M. Deghel, M. Assaad, and M. Debbah, "Queueing Stability and CSI Probing of a TDD Wireless Network with Interference Alignment", 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, 2015, pp. 794-798.
- [85] M. Deghel, M. Assaad, and M. Debbah, "System Performance of Interference Alignment under TDD Mode with Limited Backhaul Capacity", 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 1673-1678.
- [86] M. Deghel, E. Baştuğ, M. Assaad, and M. Debbah, "On the Benefits of Edge Caching for MIMO Interference Alignment", 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Stockholm, 2015, pp. 655-659.



## Chapter 2

# Mathematical Tools for Queueing Analysis

In this chapter we provide the queueing theoretic tools used throughout the thesis.

To begin with, we define a system with one scheduler and  $N$  users, where the incoming data for each user is stored in a respective queue (at the scheduler) until transmission. Let  $q_k(t)$  denote the queue length of user  $k$ , which can be seen as a real valued stochastic process that evolves in discrete time over time-slots  $t \in \{0, 1, 2, \dots\}$ ;  $q_k(t)$  can also be called the *backlog* of user  $k$  at time-slot  $t$ , as it represents an amount of work that needs to be done. Let  $A_k(t)$  denote the arrival rate for user  $k$  at time-slot  $t$ , and let  $a_k = \mathbb{E}\{A_k(t)\}$  be the corresponding mean arrival rate. We assume that the arrival process is finite, i.e.  $A_k(t) < A_{\max}$  where  $A_{\max}$  is a finite positive constant. In addition, we define  $D_k(t)$  to be the amount of data that is decided to be transmitted to user  $k$  at time-slot  $t$ ; in other words,  $D_k(t)$  is the *service rate* of user  $k$  at time-slot  $t$ .

Note that in the context of wireless communications,  $D_k(t)$  depends on the state of the wireless channels, the scheduling policy, and the resource allocation algorithms employed. We also note that the units of  $q_k(t)$ ,  $A_k(t)$ , and  $D_k(t)$  depend on the context of the system under consideration. For example, these quantities might be integers with units of *packets*. Alternatively, they might be real numbers with units of *bits*, *kilobits*, or some other unit of unfinished work relevant to the system. In this thesis, we assume that the queue lengths are measured in bits and arrival and service processes in bits per slot, unless specified otherwise.

Under the above definitions, the queue lengths evolve according to the following *dynamic equation*:

$$q_k(t+1) = \max\{q_k(t) - D_k(t), 0\} + A_k(t), \quad \forall k \in \{1, \dots, N\}, \forall t \in \{0, 1, \dots\}. \quad (2.1)$$

From the above equation, it can be noticed that for user  $k$  the actual work processed on slot  $t$  is defined as  $\min\{q_k(t), D_k(t)\}$ , which may be less than the offered amount  $D_k(t)$ ; this happens if there is little or no backlog for user  $k$  on time-slot  $t$ . We point out that here we assume that the arrivals occur at the end of time-slot  $t$ , so that they cannot be transmitted during that slot.

An alternative way to express the dynamic equation is by considering the case where the arrivals of

time-slot  $t$  can be transmitted during the same slot, hence we have

$$q_k(t+1) = \max \{q_k(t) - D_k(t) + A_k(t), 0\}, \quad \forall k \in \{1, \dots, N\}, \forall t \in \{0, 1, \dots\}. \quad (2.2)$$

We point out that the stability analysis is the same under either of the dynamic equations given above.

## 2.1 Different Forms of Queueing Stability

The different forms of stability of queueing systems are as follows.

**Definition 1** (Rate Stability). *The system is rate stable if*

$$\lim_{T \rightarrow \infty} \frac{q_k(T)}{T} = 0 \text{ with probability } 1, \forall k \in \{1, \dots, N\}. \quad (2.3)$$

**Definition 2** (Mean Rate Stability). *The system is mean rate stable if*

$$\lim_{T \rightarrow \infty} \frac{\mathbb{E}\{|q_k(T)|\}}{T} = 0 \text{ with probability } 1, \forall k \in \{1, \dots, N\}. \quad (2.4)$$

**Definition 3** (Steady State Stability). *The system is called steady state stable if*

$$\lim_{\eta \rightarrow \infty} \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{P}\{|q_k(t)| \geq \eta\} = 0, \text{ for } \eta \geq 0, \forall k \in \{1, \dots, N\}. \quad (2.5)$$

**Definition 4** (Strong Stability). *The condition for strong stability of the system can be expressed as*

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{q_k(t)\} < \infty, \quad \forall k \in \{1, \dots, N\}. \quad (2.6)$$

In this thesis, the focus will be mainly on *strong stability*. From its definition, strong stability implies that the mean queue length of every queue in the system is finite and (by *Little's theorem*) it also implies finite average delays. It should be noted that under some mild assumptions strong stability implies all of the other three forms of stability [16].

## 2.2 Stability Region and Optimal Scheduling Policy

In this section, we present the definitions of the concepts of stability region and optimal scheduling policy.

**Definition 5** (Stability Region of a Scheduling Policy). *The stability region of a scheduling policy is defined as the set of mean arrival rate vectors for which the system stays stable under this policy.*

The stability region of a specific scheduling policy should be distinguished from the *system stability region*. For the definition of this latter region, we assume that each channel has  $M$  possible states.

## 2.2. Stability Region and Optimal Scheduling Policy

---

Under state  $m$ , we define  $c_{k,m}$  to be the amount of data that can be transmitted to user  $k$  at time-slot  $t$ . Note that  $D_k(t) = c_{k,m}$  if the channel of user  $k$  is at state  $m$  and this user is scheduled for transmission. Let  $p_{k,m}$  be the probability that user  $k$  is scheduled under state  $m$  and  $\pi_m$  be the probability that state  $m$  occurs.

**Definition 6** (Stability Region of a System). *The system stability region consists of all the mean arrival rate vectors  $\mathbf{a} = (a_1, \dots, a_N)$  such that there exists a set of probabilities  $\{p_{k,m}\}$  yielding*

$$a_k \leq \sum_{m=1}^M p_{k,m} c_{k,m} \pi_m, \quad \forall k, \quad (2.7)$$

with

$$\sum_k p_{k,m} \leq 1, \quad \forall m. \quad (2.8)$$

The concept of stability region leads us to the definition of optimal scheduling policy.

**Definition 7** (Optimal Scheduling Policy). *A scheduling policy that achieves the entire system stability region is called throughput optimal.*

It is noteworthy to mention that when describing and characterizing stability regions, we implicitly mean that the system is strongly stable in the *interior* of the characterized region. For the points on the boundary of the region, the system is known to have at least a weaker form of stability [16]. Note that throughout this manuscript we shall often use the term “stability” to refer to “strong stability”, unless specified otherwise. We finally point out that a common feature for stability regions is that they all are *convex*, i.e. any point inside a stability region can be written as a convex combination of the vertices (i.e. corner points) of this region.

### Example of the construction of the system stability region

To provide some insight on how to characterize the stability region of a system, consider the following example. Consider a system with a single base-station (i.e. central scheduler) and two mobile users, and where each user has two possible channel states {ON, OFF}. We suppose that  $D_k(t) = 1$  if at time-slot  $t$  the channel state of user  $k$  is ON and this user gets scheduled for transmission; otherwise,  $D_k(t) = 0$ . Let states ON and OFF be denoted by indexes  $m = 1$  and  $m = 2$ , respectively. We assume that the probability of the channel state to be ON is the same for both users and given by  $1/2$ . We also assume that the base-station has full knowledge of the channel states at each time-slot. If only one user’s channel state is ON, the base-station schedules the user with ON channel state; otherwise, the base-station schedules user  $k$  with probability  $p_{k,1}$ . We note that  $p_{1,1} + p_{2,1} \leq 1$ .

### 2.3. Stability Optimal Scheduling Algorithms

Based on the above, the rate allocated to user  $k$  ( $k = 1, 2$ ) is

$$\begin{aligned} & \mathbb{P} \{ \text{the channel state of user } k \text{ is ON, the channel state of the other user is OFF} \} + \\ & p_{k,1} \mathbb{P} \{ \text{the channel states of both users are ON} \} \\ &= \frac{1}{4} + p_{k,1} \frac{1}{4}. \end{aligned} \tag{2.9}$$

Hence, the stability region of the system is characterized by the following set of equations

$$a_1 \leq \frac{1}{4} + p_{1,1} \frac{1}{4}, \quad a_2 \leq \frac{1}{4} + p_{2,1} \frac{1}{4}, \quad \text{and } p_{1,1} + p_{2,1} \leq 1. \tag{2.10}$$

Based on the above equations, and by varying  $p_{1,1}$  and  $p_{2,1}$ , we can plot the system stability region, as shown in Figure 2.1.

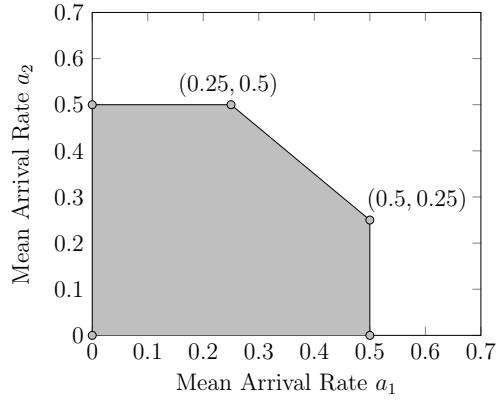


Figure 2.1: The convex hull of the set of points  $\{(0,0), (0,0.5), (0.5,0), (0.25,0.5), (0.5,0.25)\}$  represents the system stability region of the example.

## 2.3 Stability Optimal Scheduling Algorithms

In this section, we first present a stability optimal algorithm (i.e. policy) for stabilizing networks without requiring knowledge of the statistics of the arrival processes, where we note that such statistics are not available in general. Then, for the special case where the arrival rates are known, we discuss and present a scheduling policy the system can adopt in order to stabilize the network.

We consider a general transmission scheme where multiple users can be scheduled for transmission at the same time. In addition, we assume a single-hop system. We define  $\mathcal{L}(t)$  to be the subset of scheduled users at time-slot  $t$ . Let  $\Lambda$  denote the system stability region. We next provide an optimal scheduling policy that selects the subset of users to schedule at each time-slot based on the queue lengths and the available channel state information at the scheduler.

### 2.3.1 Max-Weight Policy

Here, we present the Max-Weight scheduling policy and we show that, under some conditions, this policy is throughput optimal. To this end, we define  $C_k(t)$  to be the amount of data that can be transmitted to user  $k$  at time-slot  $t$ . It is worth pointing out that  $D_k(t) = C_k(t)$  if user  $k$  is scheduled for transmission at time-slot  $t$ ; otherwise,  $D_k(t) = 0$ . Recall that the scheduler has instantaneous knowledge of all the queue lengths. Based on the above, the Max-Weight scheduling policy can be defined as follows.

**Definition 8** (Max-Weight Policy). *In every time-slot, e.g.  $t$ , the Max-Weight algorithm schedules a subset of users,  $\mathcal{L}(t)$ , according to the following*

$$\mathcal{L}(t) \in \arg \max_{\mathcal{L}_i} \mathbb{E} \left\{ \sum_{k \in \mathcal{L}_i} q_k(t) C_k(t) \right\}, \quad (2.11)$$

where (i) the expectation is over the randomness of the channel state, (ii)  $\mathcal{L}_i$  represents any possible subset of users, and (iii) breaking ties at random. In the special case where the scheduler also has complete knowledge of the  $C_k(t)$ , the definition of the Max-Weight policy reduces to

$$\mathcal{L}(t) \in \arg \max_{\mathcal{L}_i} \left\{ \sum_{k \in \mathcal{L}_i} q_k(t) C_k(t) \right\}. \quad (2.12)$$

It is worth mentioning that for systems where at most one user can be scheduled per channel, e.g. OFDMA system, and where complete knowledge of the  $C_k(t)$  is available at the scheduler, the definition of Max-Weight reduces to find the user that yields  $\max_k \{q_k(t)C_k(t)\}$ . Note that the Max-Weight (which is sometimes also called *Min-Drift*) is a special case of a more general scheduling rule called *Back-Pressure* algorithm [15], which is used in multi-hop systems and which schedules packets by looking at the differential backlogs (i.e. queue-length differences from a one-hop downstream node).

In the following, we provide the basic tools for proving stability via *Lyapunov function* techniques, which we then use for proving the stability of the Max-Weight policy. To this end, denote  $\mathbf{q}(t) = (q_1(t), \dots, q_N(t))$ . The idea is to define a non-negative function, called Lyapunov function, as a scalar measure of the aggregate congestion of all queues in the system. Scheduling algorithms are then evaluated in terms of how they affect the change in the Lyapunov function from one slot to the next.

**Definition 9** (Lyapunov Function). *A function  $Ly : \mathbb{R}^N \rightarrow \mathbb{R}$  is called Lyapunov function if it satisfies the following properties: (i)  $Ly(\cdot)$  is non-negative, i.e.  $Ly(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^N$ , (ii)  $Ly(\mathbf{x}) \rightarrow +\infty$  as  $\mathbf{x} \rightarrow +\infty$ , (iii) non-decreasing in any of its arguments, and (iv) differentiable.*

**Theorem 1** (Foster-Lyapunov Theorem). *Let  $\{\mathbf{q}(t)\}$  be an irreducible discrete time Markov chain with a countable state space. Suppose that there exists a Lyapunov function  $Ly(\cdot)$  and a finite set  $\mathcal{X} \in \mathbb{R}_+^N$  satisfying the following conditions:*

- $\mathbb{E} \{Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t) = \mathbf{x}\} < +\infty$  if  $\mathbf{x} \in \mathcal{X}$
- $\mathbb{E} \{Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t) = \mathbf{x}\} < 0$  if  $\mathbf{x} \in \mathcal{X}^c$

### 2.3. Stability Optimal Scheduling Algorithms

---

then the Markov chain  $\{\mathbf{q}(t)\}$  is positive recurrent.

It is worth mentioning that the expression  $\mathbb{E}\{Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t) = \mathbf{x}\}$  is called the (one step) *Lyapunov drift*, representing the expected change in the Lyapunov function from one slot to the next. To make the link between the Foster-Lyapunov theorem and the stability of the adopted system, we consider the quadratic Lyapunov function defined as the following

$$Ly(\mathbf{q}(t)) \triangleq \frac{1}{2} (\mathbf{q}(t) \cdot \mathbf{q}(t)) = \frac{1}{2} \sum_{k=1}^N (q_k(t))^2. \quad (2.13)$$

Based on the above Lyapunov function, we next present an extension of the Foster-Lyapunov Theorem, which represents a sufficient condition for stability.

**Theorem 2** (Extension of Foster-Lyapunov Theorem). *Under the quadratic Lyapunov function defined as  $Ly(\mathbf{q}(t)) = \frac{1}{2} \sum_{k=1}^N (q_k(t))^2$ , if there exist constants  $E < \infty$  and  $\epsilon > 0$ , such that for all time-slots the following inequality holds*

$$\mathbb{E}\{Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t)\} \leq E - \epsilon \sum_k q_k(t), \quad (2.14)$$

then the Markov chain  $\{\mathbf{q}(t)\}$  is positive recurrent. Furthermore, the system  $\mathbf{q}(t)$  is strongly stable, meaning that

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{q_k(t)\} < \infty, \quad \forall k \in \{1, \dots, N\}. \quad (2.15)$$

#### 2.3.2 Strong Stability Proof for the Max-Weight Policy

Using the tools given in the previous subsection, we now provide a detailed proof of the stability of the Max-Weight policy. Specifically, we want to prove that the following statement holds.

**Theorem 3.** *The Max-Weight policy is throughput optimal. In other words, it can stabilize the system for every mean arrival rate vector that is inside the system stability region.*

*Proof.* Let us use  $\Delta^*$  to denote the Max-Weight policy. Define the quadratic *Lyapunov function*  $Ly(\mathbf{q}(t)) \triangleq \frac{1}{2} (\mathbf{q}(t) \cdot \mathbf{q}(t)) = \frac{1}{2} \sum_{k=1}^N q_k(t)^2$ . Note that  $q_k(t)^2$  is nothing but  $q_k(t)(t)^2$ ; the same remark holds for  $A_k(t)$  and  $D_k(t)$ , i.e.  $A_k(t)^2$  and  $D_k(t)^2$  are used to denote  $(A_k(t))^2$  and  $(D_k(t))^2$ , respectively. To describe the evolution of the queue lengths, we use the following dynamic equation

$$q_k(t+1) = \max\{q_k(t) - D_k(t), 0\} + A_k(t), \quad \forall k \in \{1, \dots, N\}, \forall t \in \{0, 1, \dots\}. \quad (2.16)$$



### 2.3. Stability Optimal Scheduling Algorithms

---

From the evolution equation given above we have

$$\begin{aligned}
Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) &= \frac{1}{2} \sum_{k=1}^N [q_k(t+1)^2 - q_k(t)^2] \\
&= \frac{1}{2} \sum_{k=1}^N [(\max\{q_k(t) - D_k(t), 0\} + A_k(t))^2 - q_k(t)^2] \\
&\leq \sum_{k=1}^N \frac{[A_k(t)^2 + D_k(t)^2]}{2} + \sum_{k=1}^N q_k(t) [A_k(t) - D_k(t)], \tag{2.17}
\end{aligned}$$

where in the final inequality we have used the fact that for any  $q \geq 0$ ,  $A \geq 0$ ,  $D \geq 0$ , we have

$$(\max\{q - D, 0\} + A)^2 \leq q^2 + A^2 + D^2 + 2q(A - D).$$

Now define  $Dr(\mathbf{q}(t))$  as the *conditional Lyapunov drift* for time-slot  $t$

$$Dr(\mathbf{q}(t)) \triangleq \mathbb{E}\{Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t)\}. \tag{2.18}$$

From (2.17), we have that  $Dr(\mathbf{q}(t))$  for a general scheduling policy satisfies

$$Dr(\mathbf{q}(t)) \leq \mathbb{E}\left\{\sum_{k=1}^N \frac{A_k(t)^2 + D_k(t)^2}{2} \mid \mathbf{q}(t)\right\} + \sum_{k=1}^N q_k(t) a_k - \mathbb{E}\left\{\sum_{k=1}^N q_k(t) D_k(t) \mid \mathbf{q}(t)\right\}, \tag{2.19}$$

where we have used the fact that arrivals are independent and identically distributed (i.i.d.) over slots and hence independent of current queue backlogs, so that  $\mathbb{E}\{A_k(t) \mid \mathbf{q}(t)\} = \mathbb{E}\{A_k(t)\} = a_k$ . Now define  $E$  as a finite positive constant that bounds the first term on the right-hand-side of the above drift inequality, so that for all  $t$ , all possible  $q_k(t)$ , and all possible control decisions that can be taken, we have

$$\mathbb{E}\left\{\sum_{k=1}^N \frac{A_k(t)^2 + D_k(t)^2}{2} \mid \mathbf{q}(t)\right\} \leq E. \tag{2.20}$$

Note that  $E$  exists since  $A_k(t) < A_{\max}$  and the amount of data that can be transmitted  $D_k(t)$  is bounded, i.e. there exists a finite positive constant  $D_{\max}$  such that  $D_k(t) < D_{\max}$ .

Using the expression in (2.19) yields

$$Dr(\mathbf{q}(t)) \leq E + \sum_{k=1}^N q_k(t) a_k - \mathbb{E}\left\{\sum_{k=1}^N q_k(t) D_k(t) \mid \mathbf{q}(t)\right\}. \tag{2.21}$$

The conditional expectation at the right-hand-side of the above inequality is with respect to the randomly observed channel states. To emphasize how the right-hand-side of the above inequality depends on the scheduling decision, for example  $\Delta^*$ , we use the identity  $D_k(t) = D_k^{(\Delta^*)}(t)$ . Thus,

### 2.3. Stability Optimal Scheduling Algorithms

---

under  $\Delta^*$  the drift can be expressed as

$$Dr^{(\Delta^*)}(\mathbf{q}(t)) \leq E + \sum_{k=1}^N q_k(t)a_k - \mathbb{E} \left\{ \sum_{k=1}^N q_k(t)D_k^{(\Delta^*)}(t) \mid \mathbf{q}(t) \right\}. \quad (2.22)$$

Let us define  $\Delta$  as any alternative (possibly randomized) scheduling policy that can be made on time-slot  $t$ . Using the definition of  $\Delta^*$ , that is, selecting  $\mathcal{L}_i$  that maximizes  $\sum_{k \in \mathcal{L}_i} q_k(t)D_k(t)$ , we have

$$\mathbb{E} \left\{ \sum_{k=1}^N q_k(t)D_k^{(\Delta^*)}(t) \mid \mathbf{q}(t) \right\} \geq \mathbb{E} \left\{ \sum_{k=1}^N q_k(t)D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\}. \quad (2.23)$$

Plugging the above directly into (2.22) yields

$$Dr^{(\Delta^*)}(\mathbf{q}(t)) \leq E - \sum_{k=1}^N q_k(t) \left[ \mathbb{E} \left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\} - a_k \right]. \quad (2.24)$$

Now, suppose the arrival rate vector  $\mathbf{a}$  is interior to the stability region  $\Lambda$ , and consider a particular policy  $\Delta$  that depends only on the channels states and that can stabilize the system. Because these states are assumed to be i.i.d. over slots, the resulting service rates,  $D_1^{(\Delta)}(t), \dots, D_N^{(\Delta)}(t)$ , are independent of the current queue length vector,  $\mathbf{q}(t)$ . Using the above, we can write

$$\mathbb{E} \left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\} = \mathbb{E} \left\{ D_k^{(\Delta)}(t) \right\} \geq a_k + \epsilon_{\max}(\mathbf{a}), \quad \forall k \in \{1, \dots, N\}, \quad (2.25)$$

where  $\epsilon_{\max}(\mathbf{a})$  is a positive constant that depends on the mean arrival rate vector, and consequently

$$Dr^{(\Delta^*)}(\mathbf{q}(t)) \leq E - \epsilon_{\max}(\mathbf{a}) \sum_{k=1}^N q_k(t). \quad (2.26)$$

Based on the above expression and using the statement of Theorem 2, we have that the Markov chain  $\{\mathbf{q}(t)\}$  is positive recurrent, and the system  $\mathbf{q}(t)$  is strongly stable. We next show why this latter strong-stability statement holds. Taking an expectation of the expression in (2.26) over the randomness of the queue lengths yields the following

$$\mathbb{E} \left\{ Dr^{(\Delta^*)}(\mathbf{q}(t)) \right\} \leq E - \epsilon_{\max}(\mathbf{a}) \sum_{k=1}^N \mathbb{E} \{ q_k(t) \}. \quad (2.27)$$

Using the definition of  $Dr^{(\Delta^*)}(\mathbf{q}(t))$  with the law of iterated expectations gives

$$\begin{aligned} \mathbb{E} \left\{ Dr^{(\Delta^*)}(\mathbf{q}(t)) \right\} &= \mathbb{E} \{ \mathbb{E} \{ Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t) \} \} \\ &= \mathbb{E} \{ Ly(\mathbf{q}(t+1)) \} - \mathbb{E} \{ Ly(\mathbf{q}(t)) \}. \end{aligned} \quad (2.28)$$

The above holds for all  $t \in \{0, 1, 2, \dots\}$ . Summing over  $t \in \{0, 1, \dots, T-1\}$  for some integer  $T > 0$

## 2.4. Suboptimal or Approximate Scheduling

---

yields (by telescoping sums)

$$\mathbb{E}\{Ly(\mathbf{q}(T))\} - \mathbb{E}\{Ly(\mathbf{q}(0))\} \leq ET - \epsilon_{\max}(\mathbf{a}) \sum_{t=0}^{T-1} \sum_{k=1}^N \mathbb{E}\{q_k(t)\}. \quad (2.29)$$

Rearranging terms, dividing by  $\epsilon_{\max}(\mathbf{a})T$ , and using the fact that  $Ly(\mathbf{q}(0)) \geq 0$  yields

$$\frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^N \mathbb{E}\{q_k(t)\} \leq \frac{E}{\epsilon_{\max}(\mathbf{a})} + \frac{\mathbb{E}\{Ly(\mathbf{q}(0))\}}{\epsilon_{\max}(\mathbf{a})T}. \quad (2.30)$$

Assuming  $\mathbb{E}\{Ly(\mathbf{q}(0))\} < \infty$  and taking a lim sup we eventually get

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^N \mathbb{E}\{q_k(t)\} \leq \frac{E}{\epsilon_{\max}(\mathbf{a})}. \quad (2.31)$$

Hence, all the queues are strongly stable, and the total average queue length (summed over all the queues in the system) is less than or equal to the bound  $\frac{E}{\epsilon_{\max}(\mathbf{a})}$ . Therefore, the Max-Weight algorithm ensures that the system is strongly stable whenever the mean arrival rate vector  $\mathbf{a}$  is interior to the stability region  $\Lambda$ . From equation (2.31), it can be seen that the average queue congestion bound is inversely proportional to the distance the rate vector is away from the stability region boundary.  $\square$

### 2.3.3 Scheduling with Known Mean Arrival Rates

We first point out that the scheduling approach we present in this case is based on a sufficient condition for *rate stability*, where this condition consists in having the mean service rate greater than or equal to the mean arrival rate [16]. If we have a system where the mean arrival rates are known, the rate stability of the system can be achieved by a predefined time-sharing strategy. Indeed, a mean arrival rate vector  $\mathbf{a} \in \Lambda$  can be expressed as a convex combination of the points in  $\mathcal{V}$ , where  $\mathcal{V}$  denotes the set that contains the corner points (i.e. vertices) of  $\Lambda$ . More in detail, we have  $\mathbf{a} = \sum_n p_n \mathbf{r}_n$ , where  $\mathbf{r}_n$  represents the  $n$ -th element of  $\mathcal{V}$ ,  $p_n \geq 0$  and  $\sum_n p_n = 1$ ; it should be noted that each point  $\mathbf{r}_n$  represents a specific scheduling decision. Then, we can find at least one point  $\boldsymbol{\lambda}$  on the boundary of  $\Lambda$  such that  $\mathbf{a} \preceq \boldsymbol{\lambda}$ . Since  $\boldsymbol{\lambda} \in \Lambda$ , we can write  $\boldsymbol{\lambda} = \sum_n \delta_n \mathbf{r}_n$ , with  $\delta_n \geq 0$  and  $\sum_n \delta_n = 1$ . Then, to achieve queues stability, each point (i.e. decision)  $\mathbf{r}_n$  should be selected with probability  $\delta_n$ .

## 2.4 Suboptimal or Approximate Scheduling

We here provide and discuss the definition of an approximate policy [15, 16]. We keep adopting the same scheme defined in the previous section. Let  $\Delta^*$  and  $\Delta$  denote the optimal and approximate scheduling policies, respectively. Recall that  $\Delta^*$  is a throughput optimal policy, i.e. it achieves the system stability region  $\Lambda$ . As mentioned earlier, the amount of data that can be transmitted to each user, e.g.  $D_k(t)$ , depends on the scheduling policy. To represent this dependency, we use the identities  $D_k(t) = D_k^{(\Delta^*)}(t)$  and  $D_k(t) = D_k^{(\Delta)}(t)$  under  $\Delta^*$  and  $\Delta$ , respectively. We note that when e.g. user  $k$

## 2.4. Suboptimal or Approximate Scheduling

---

is not selected by the scheduling policy for transmission, we have  $D_k(t) = 0$ .

**Definition 10.** *If for some constants  $\beta$  and  $E_a$ , such that  $0 \leq \beta \leq 1$  and  $0 \leq E_a < \infty$ , we have*

$$\sum_k \mathbb{E} \left\{ q_k(t) D_k^{(\Delta)}(t) \mid q_k(t) \right\} \geq \beta \sum_k \mathbb{E} \left\{ q_k(t) D_k^{(\Delta^*)}(t) \mid q_k(t) \right\} - E_a, \quad (2.32)$$

*then  $\Delta$  is called  $(\beta, E_a)$ -approximate policy with respect to the optimal policy  $\Delta^*$ .*

Under this definition, we have that policy  $\Delta$  can provide stability whenever the arrival rates are interior to  $\beta\Lambda$ , which is a  $\beta$ -scaled version of the stability region  $\Lambda$ . Hence, if the scheduling policy is off from the optimum by no more than an additive constant  $E_a$  (i.e.  $\beta = 1$ ), then the (stability) throughput optimality is still possible; but, in this case, the average queue length increases by an additive constant proportional to  $E_a$ . However, if the scheduling policy deviates from optimality by a multiplicative constant, the achievable stability region may be a subset of the capacity region.

From the above observations, one can conclude that if the maximum per-time-slot length change in any queue is bounded, full system stability can still be achieved even by using out of date queue lengths information. In other words, queue updates can be arbitrarily infrequent without affecting stability of the system, although the average queue length may increase in proportion to the duration between updates, which means greater average delay because *Little's Theorem* tells us that *average queue length is proportional to average delay* [16].

## Chapter 3

# Queueing Stability and CSI Probing of a TDD Wireless Network with Interference Alignment

### 3.1 Overview

In this chapter, we study the queueing stability of a MIMO interference channel under TDD mode with limited backhaul capacity and taking into account the probing cost. The interference channel is an information theoretic concept that models wireless networks in which several transmitters simultaneously communicate data to their paired receivers. One of the key issues in such systems is the interference that is caused by multiple transmitter-receiver pairs communicating on the same channel, resulting into severe performance degradations unless treated properly. The traditional approach for communication in such channels was orthogonalization, where resources are split among the various pairs. The use of multiple antennas [2] has emerged as one of the enabling technologies to increase the performance of wireless systems. Under such an approach, two important decisions need to be taken: (i) which interference management technique we should adopt, i.e. how to precode and decode the signals in such a way as to efficiently manage the interference problem, and (ii) at every slot, which pairs should be scheduled.

Regarding the choice of the interference management technique, IA was introduced [7] as an efficient technique and is shown to result in higher throughputs compared to conventional interference-agnostic methods. Indeed, IA is a linear precoding technique that attempts to align interfering signals in time, frequency, or space. In MIMO networks, IA utilizes the spatial dimension offered by multiple antennas for alignment. By aligning interference at all receivers, IA reduces the dimension of interference, allowing receivers to suppress interference via linear techniques and decode their desired signals interference free. Allowing some coordination between the transmitters and receivers enables interference alignment. In this way, it is possible to design the transmit strategies in such a way as to

ensure that the interference aligns at each receiver. From a sum rate perspective, with  $N$  transmitter-receiver pairs, an ideal IA strategy achieves a sum throughput on the order of  $N/2$  interference free links [7]. Basically each pair can effectively get half the system capacity. Thus, unlike the conventional interference channel, using IA there is a net sum capacity increase with the number of (active) pairs. A major challenge of the IA scheme lies in the fact that the global CSI must be available at each transmitter. Such a challenge was the subject of extensive research over the last few years. For instance, in scenarios where the receivers quantize and send the CSI back to the transmitters, the IA scheme is explored over frequency selective channels for single-antenna users in [87] and for multiple-antenna users in [88]. Both references provide degree-of-freedom (DoF)-achieving quantization schemes and establish the required scaling of the number of feedback bits. For alignment using spatial dimensions, [89] provides the scaling of feedback bits to achieve IA in MIMO interference channel (IC). To overcome the problem of scaling codebook size, and relax the reliance on frequency selectivity for quantization, [90] proposed an analog feedback strategy for constant MIMO interference channels. From another point of view, [91] provides an analysis of the effect of imperfect CSI on the mutual information of the interference alignment scheme. On the other side, for TDD systems, every transmitter can estimate its downlink channels from the uplink transmission phase thanks to reciprocity. However, for the IA scheme, this local knowledge is not sufficient, and the transmitters need to share their channel estimates that can be carried out through backhaul links between transmitters. These links generally have a limited capacity, which should be exploited efficiently. In [92] a compression scheme for the cloud radio access networks is proposed. In [93] the Grassmannian Manifold quantization technique was adopted to reduce the information exchange over the backhaul. It is noteworthy to mention that the above cited works do not account for the dynamic traffic processes of the users, meaning that they assume users with infinite back-logged data.

We draw the attention to the fact that IA technique can be used with a number of transmitter-receiver pairs greater than or equal to two. On the other side, for the special case where we have a point-to-point MIMO system, i.e. only one pair, SVD technique can be applied and was shown to provide very good performance [6, 94]; note that in this case other techniques can be used such as Zero Forcing (ZF) and Matched Filtering (MF) [95].

It is of great interest to investigate the impact of MIMO in the higher layers [96], more specifically in the media access control (MAC) layer. The cross-layer design goal here is the achievement of the entire *stability region* of the system. In broad terms, the stability region of a network is the set of arrival rate vectors such that the entire network load can be served by some service policy without an infinite blow up of any queue. The special *scheduling policy* achieving the entire stability region, called the optimal policy, is hereby of particular interest. The concept of stability-optimal operation concept has been extensively investigated in the wireless framework under various traffic and network scenarios. For instance, in [97] the authors have presented a precoding strategy that achieves the system stability region, under the assumptions of perfect CSI and use of Gaussian codebooks. This strategy is based on Lyapunov drift minimization given the queue lengths and channel states every time-slot. In [43], the authors have considered the broadcast channel (BC) and proposed a technique based on ZF precoding, with a heuristic user scheduling scheme that selects users whose channel states are nearly orthogonal vectors and illustrate the stability region this policy achieves via simulations. In [98], it has been noticed that the policy resulting from the minimization of the drift of a quadratic

Lyapunov function is to solve a weighted sum rate maximization problem (with weights being the queue lengths) each time-slot, and an iterative water-filling algorithm has been proposed for this purpose. In addition, taking into account the probing cost, the authors in [68] have examined three different scheduling policies (centralized, decentralized and mixed policies) for MISO wireless downlink systems under ZF precoding technique.

In this chapter, we consider a MIMO network where multiple transmitter-receiver pairs operate in TDD mode under backhaul links of limited capacity. Each transmitter acquires its local CSI from its corresponding receiver by exploiting the channel reciprocity. We use (pre-assigned) orthogonal pilot sequences among the users, so the length of each one of these sequences should be proportional to the number of active users in the system; orthogonal sequences are produced e.g. by Walsh-Hadamard on pseudonoise sequences. It means that after acquiring the CSI of, for example,  $L$  users, the throughput is multiplied by  $1 - L\theta$ , where  $\theta$  is the fraction of time that takes the CSI acquisition of one user [99]. Depending on the number of scheduled pairs at a time-slot, we distinguish two cases: (i) if the number of scheduled pairs is greater than or equal to two, IA technique is applied, and each transmitter needs to send a quantized version of its local CSI to other transmitters using a fixed number of bits  $B$  per channel matrix, and (ii) if only one pair is scheduled, we apply SVD technique. It is important to focus on the trade-off between having a large number of active transmitter-receiver pairs (so having a high probing cost but many pairs can communicate simultaneously) and having much time of the slot dedicated to data transmission (which means getting a low probing cost but few pairs can communicate simultaneously) [68]. In order to choose the subset of active pairs at each slot, we adopt a centralized scheme where the decision of which pairs to schedule is made at a central scheduler (CS) based only on the statistics of the channels of the users and the state of their queue lengths at each slot [63]. Note that the centralized approach is used in current standards (e.g. LTE [9]), where the base-station explicitly requests some users for their CSI.

## Contributions

The *main contributions* and the *organization* of this chapter are summarized as follows.

- Section 3.2 presents the adopted system model and the interaction between the physical layer and the queueing performance.
- The average rate expressions for both the perfect (i.e. unlimited backhaul capacity) and imperfect (i.e. limited backhaul capacity) cases are derived in Section 3.3.
- In Section 3.4, we present a stability analysis for the *symmetric* system where all the path loss coefficients are equal to each other. Specifically, for this system:
  - ★ We provide a precise characterization of the system stability region and we propose an optimal scheduling policy to achieve this region in both the perfect and imperfect cases.
  - ★ Under both cases, we investigate the conditions under which the use of IA (i.e. applied if the number of active pairs  $\geq 2$ ) can yield a queueing stability gain compared to SVD (i.e. applied if there is only 1 active pair).
  - ★ We examine the maximum gap between the stability region under the imperfect case and the

## 3.2. System Model

---

stability region under the perfect case. We also investigate the impact of changing the number of bits on the system stability region.

- In Section 3.5, we present a stability analysis for the *general* system where the path loss coefficients are not necessarily equal to each other. In detail, for this system:
  - ★ We investigate the stability performance by characterizing the system stability region and providing an optimal scheduling policy under both the imperfect and perfect cases.
  - ★ Since the scheduling policy is of a high computational complexity, under the imperfect case we propose an approximate policy that has a reduced complexity but that achieves only a fraction of the stability region of the imperfect case; we point out that this policy relies on an average rate approximation (for the imperfect case) that we first calculate. A characterization of the fraction this policy achieves is also provided.
  - ★ Using the average rate approximation mentioned above, we examine the gap between the stability region under the imperfect case and the stability region under the perfect case.
- In Section 3.6, we provide a stability analysis for the symmetric case with multiple rate levels. The investigated points are similar to those in the symmetric case with single rate level.
- Section 3.7 is dedicated to numerical results and relevant discussions.
- Finally, Section 3.8 concludes the chapter.

## 3.2 System Model

We consider the MIMO interference channel with  $N$  transmitter-receiver pairs shown in Figure 3.1. For simplicity of exposition, we consider a homogeneous network where all transmitters are equipped with  $N_t$  antennas and all receivers (i.e. users) with  $N_r$  antennas. We assume that time is slotted. As will be explained later, only a subset  $\mathcal{L}(t)$ , of cardinality  $L(t)$ , of pairs is active at time-slot  $t$ , with  $L(t) \leq N$ . Transmitter  $k$  has  $d_k \leq \min(N_t, N_r)$  independent data streams to transmit to its intended user  $k$ . For the cases where  $L(t) \geq 2$ , while each transmitter communicates with its intended receiver, it also creates interference to other  $L(t) - 1$  unintended receivers.

Given this channel model, the received signal at active user  $k$  ( $\in \mathcal{L}(t)$ ) can be expressed as

$$\mathbf{y}_k = \sum_{i \in \mathcal{L}(t)} \sqrt{\frac{\zeta_{ki} P}{d_i}} \mathbf{H}_{ki} \sum_{j=1}^{d_i} \mathbf{v}_i^{(j)} x_i^{(j)} + \mathbf{z}_k, \quad (3.1)$$

where  $\mathbf{y}_k$  is the  $N_r \times 1$  received signal vector,  $\mathbf{z}_k$  is the additive white complex Gaussian noise with zero mean and covariance matrix  $\sigma^2 \mathbf{I}_{N_r}$ ,  $\mathbf{H}_{ki}$  is the  $N_r \times N_t$  channel matrix between transmitter  $i$  and receiver  $k$  with independent and identically distributed (i.i.d.) zero mean and unit variance complex Gaussian entries,  $\zeta_{ki}$  represents the path loss of channel  $\mathbf{H}_{ki}$ ,  $x_i^{(j)}$  represents the  $j$ -th data stream from transmitter  $i$ ,  $P$  is the total power at each transmitting node, which is equally allocated among its data streams, and  $\mathbf{v}_i^{(j)}$  is the corresponding  $N_t \times 1$  precoding vector of unit norm. For the rest of the chapter we denote by  $\alpha_{ki}$  the fraction  $\frac{\zeta_{ki} P}{d_i}$ , i.e.  $\alpha_{ki} = \frac{\zeta_{ki} P}{d_i}$ .



### 3.2. System Model

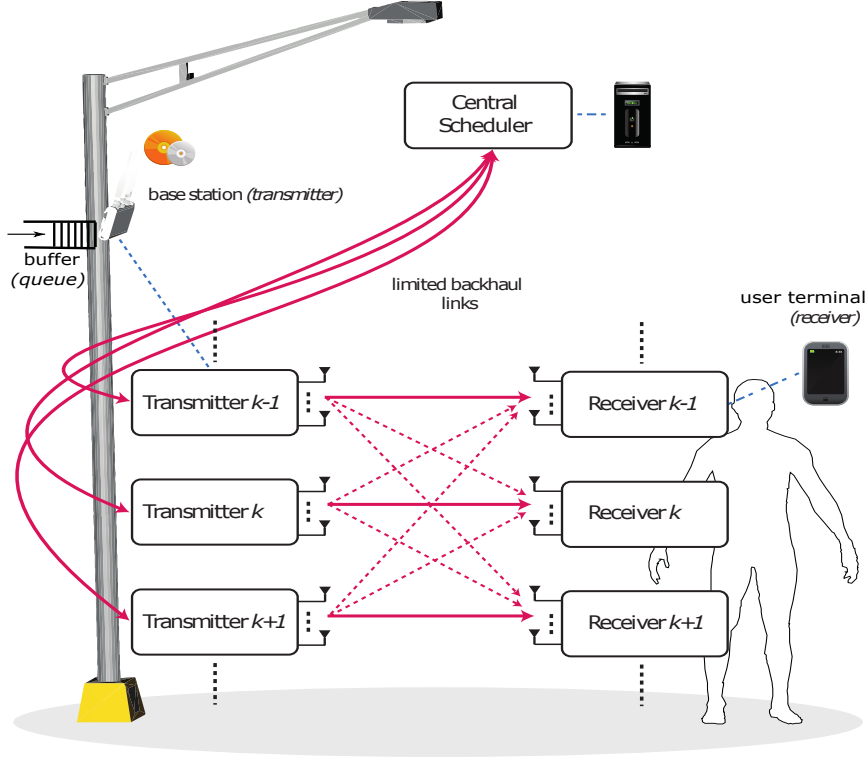


Figure 3.1: A sketch of an  $N$ -user MIMO interference channel with limited backhaul.

The scheduling rule to select the subset of active pairs at each time-slot will be discussed later in the chapter. Depending on the number of scheduled pairs  $L(t)$ , two cases are to consider:

- If  $L(t) = 1$ , i.e. only one pair is active at time-slot  $t$ : in this case we use singular value decomposition (SVD) technique, which was shown to provide very good performance for point-to-point MIMO systems [94]. Note that other techniques can be considered, such as Zero Forcing (ZF) and Matched Filtering (MF).
- If  $L(t) \geq 2$ , i.e. at least two pairs are active at time-slot  $t$ : in this case we perform Interference Alignment (IA) technique, which was shown to provide good performance for multipoint-to-multipoint MIMO systems [7].

For clarity of exposition, in this section we present the IA scheme and in the next Section (Section 3.3, in which we also derive the average rates) we present the SVD scheme.

#### 3.2.1 Interference Alignment Technique

IA is an efficient linear precoding technique that often achieves the full DoF supported by MIMO interference channels. In cases where the full DoF cannot be guaranteed, IA has been shown to provide significant gains in high SNR sum-rate. To investigate IA in our model, we start by examining the effective channels created after precoding and combining. For tractability, we restrict ourselves to

### 3.2. System Model

a per-stream zero-forcing receiver. Recall that in the high (but finite) SNR regime, in which IA is most useful, gains from more involved receiver designs are limited [100]. Under the adopted scheme, receiver  $k$  uses the  $N_r \times 1$  *combining* vector  $\mathbf{u}_k^m$  (of unit norm) to detect its  $m$ -th stream, such as

$$\begin{aligned} \hat{x}_k^{(m)} &= \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{y}_k \\ &= \overbrace{\sqrt{\alpha_{kk}} \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(m)} x_k^{(m)}}^{\text{desired signal}} + \overbrace{\sqrt{\alpha_{kk}} \sum_{j=1, j \neq m}^{d_k} \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(j)} x_k^{(j)}}^{\text{inter-stream interference (ISI)}} \\ &\quad + \underbrace{\sum_{i \in \mathcal{L}(t), i \neq k} \sqrt{\alpha_{ki}} \sum_{j=1}^{d_i} \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{ki} \mathbf{v}_i^{(j)} x_i^{(j)}}_{\text{inter-user interference (IUI)}} + \overbrace{\left( \mathbf{u}_k^{(m)} \right)^H \mathbf{z}_k}_{\text{noise}}, \end{aligned} \quad (3.2)$$

where the first term at the right-hand-side of this expression is the desired signal, the second one is the inter-stream interference (ISI) caused by the same transmitter, and the third one is the inter-user interference (IUI) resulting from the other transmitters. In order to mitigate these interferences and improve the system performance, IA is performed accordingly, that is designing the set of combining and precoding vectors such that

$$\left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{ki} \mathbf{v}_i^{(j)} = 0, \quad \forall (i, j) \neq (k, m), \text{ with } i, k \in \mathcal{L}(t). \quad (3.3)$$

In the following we state some additional assumptions on the design of these vectors [90]:

- Vector  $\mathbf{v}_i^{(j)}$  is a function of all the cross channels  $\mathbf{H}_{kl} \forall k, l, k \neq l$  only.
- Vector  $\mathbf{u}_k^{(m)}$  is a function of vectors  $\mathbf{H}_{ki} \mathbf{v}_i^{(j)} \forall i \neq k, \forall j$  and  $\mathbf{H}_{kk} \mathbf{v}_k^{(j)}, \forall j \neq m$ .
- Matrix  $\mathbf{V}_k = [\mathbf{v}_k^{(1)}, \dots, \mathbf{v}_k^{(d_k)}]$  is unitary.

From the above properties/assumptions we can deduce that  $\mathbf{u}_k^{(m)}$  is independent of  $\mathbf{H}_{kk} \mathbf{v}_k^{(m)}$ . Note that the conditions in (3.3) are those of a *perfect interference alignment*. In other words, suppose that all the transmitting nodes have perfect global CSI and each receiver obtains a perfect version of its corresponding combining vectors, ISI and IUI can be suppressed completely. However, obtaining perfect global CSI at the transmitters is not always practical due to the fact that backhaul links, which connect transmitters to each other, are of limited capacity. The CSI sharing mechanism is detailed in the next subsection.

Finally, some remarks are in order. We note that the approach to design the IA vectors is not a figure of interest in our contribution, however our analysis holds for every approach that produces IA vectors with simultaneously all the properties given before [90]; such an approach was the subject of investigation in several works, as for instance [7, 101]. In addition, we assume that each active receiver obtains a perfect version of its corresponding combining vectors. The cost of this latter process is not considered in our analysis. Moreover, we do not perform power control for our system. This is done to further simplify the transmission scheme.

### 3.2.2 CSIT Sharing Over Limited Capacity Backhaul Links

The process of CSI sharing is restricted to the scheduled pairs (represented by subset  $\mathcal{L}(t)$ ). Thus, here, even if we did not mention it, when we write “transmitter” (resp., “user”) we mean “active transmitter” (resp., “active user”). Three different scenarios regarding the CSI sharing problem can be considered:

1. Each transmitter receives all the required CSI and independently computes the IA vectors,
2. The IA processing node is a separate central node that computes and distributes the IA vectors to other transmitters,
3. One transmitter acts as the IA processing node.

For the last two scenarios, one node performs the computations and then distributes the IA vectors among transmitters. So, since the backhaul is limited in capacity, in addition to the quantization required for the CSI sharing process, another quantization is needed to distribute the IA vectors over the backhaul. This is not the case for the first scenario where only the first quantization process is needed. Thus, for simplicity of exposition and calculation, we focus on the first scenario, which we detail in the following paragraph.

We adopt a scenario where each transmitter receives all the required CSI and independently computes the IA vectors. As alluded earlier, global CSI is required at each transmitting node in order to design the IA vectors that satisfy (3.3). As shown in Figure 3.1, we suppose that all the transmitters are connected to a CS via their limited backhaul links, i.e. this CS serves as a way for connecting the transmitters to each other; as we will see later on, this scheduler decides which pairs to schedule at each time-slot. We assume a TDD transmission strategy, which enables the transmitters to estimate their channels toward different users by exploiting the reciprocity of the wireless channel. We consider throughout this chapter that there are no errors in the channel estimation.

Under the adopted strategy, the users send their training sequences in the uplink phase, allowing each transmitter to estimate (perfectly) its *local* CSI, meaning that the  $k$ -th transmitter estimates perfectly the channels  $\mathbf{H}_{ik}$ ,  $\forall i \in \mathcal{L}(t)$ . However, the local CSI of other transmitters, excluding the direct channels (i.e. for the  $k$ -th transmitter the direct channel is given by matrix  $\mathbf{H}_{kk}$ ), are obtained via backhaul links of limited capacity. In such limited backhaul conditions, a codebook-based quantization technique needs to be adopted to reduce the huge amount of information exchange used for CSI sharing, which we detail as follows. Let  $\mathbf{h}_{ik}$  denote the vectorization of the channel matrix  $\mathbf{H}_{ik}$ . Then, for each  $i \neq k$ , transmitter  $k$  selects the index  $n_o$  that corresponds to the optimal codeword in a predetermined codebook  $\mathcal{CB} = [\hat{\mathbf{h}}_{ik}^{(1)}, \dots, \hat{\mathbf{h}}_{ik}^{(2^B)}]$  according to

$$n_o = \arg \max_{1 \leq n \leq 2^B} \left| \left( \tilde{\mathbf{h}}_{ik} \right)^H \hat{\mathbf{h}}_{ik}^{(n)} \right|^2, \quad (3.4)$$

in which  $B$  is the number of bits used to quantize  $\mathbf{H}_{ik}$  and  $\tilde{\mathbf{h}}_{ik} = \mathbf{h}_{ik} \|\mathbf{h}_{ik}\|^{-1}$  is the channel direction vector. After quantizing all the matrices of its local CSI, we assume that transmitter  $k$  sends the corresponding optimal indexes to all other active transmitters, which share the same codebook, allowing these transmitters to reconstruct the quantized local knowledge of transmitter  $k$ .

We point out that each transmitter quantizes its local CSI excluding the direct channel since, as

### 3.2. System Model

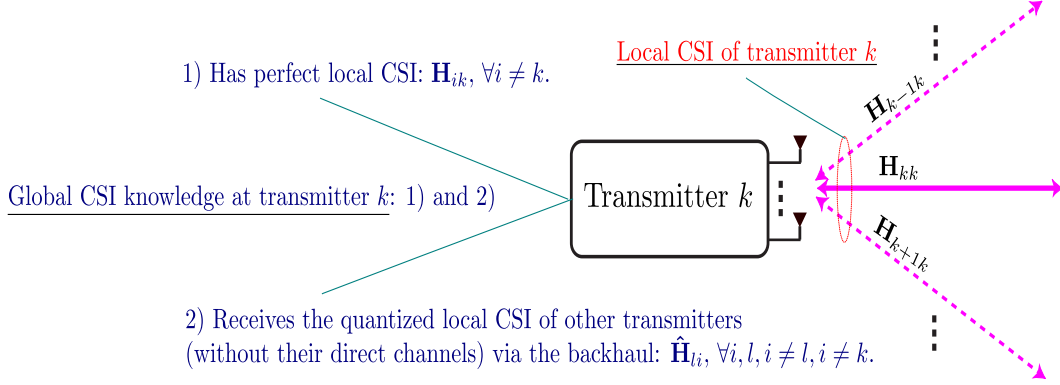


Figure 3.2: Global CSI knowledge at transmitter  $k$  under the imperfect case.

noted earlier, it is not required when computing the IA vectors of other transmitters. The global CSI knowledge at transmitter  $k$  is summarized in Figure 3.2. Let us now define the quantization error as

$$e_{ik} = 1 - \left( |\hat{\mathbf{h}}_{ik}^H \mathbf{h}_{ik}|^2 (\|\mathbf{h}_{ki}\|^2)^{-1} \right), \quad (3.5)$$

where  $\hat{\mathbf{h}}_{ik}$  is the quantization of  $\mathbf{h}_{ik}$ , and adopt a similar model to the one used in [102, Lemma 6], [63], which relies on the theory of quantization cell approximation. Let  $Q = N_t N_r - 1$ . The cumulative distribution function (CDF) of  $e_{ik}$  is then given by the following

$$\mathbb{P}\{e_{ik} \leq \varepsilon\} = \begin{cases} 2^B \varepsilon^Q, & \text{for } 0 \leq \varepsilon \leq 2^{-\frac{B}{Q}} \\ 1, & \text{for } \varepsilon > 2^{-\frac{B}{Q}} \end{cases} \quad (3.6)$$

As a final remark, we note that we assume that the number of bits  $B$  is fixed and does not change with the number of scheduled pairs  $L(t)$ . This assumption is made to simplify the analysis, because otherwise we should consider the relation between these two numbers, which would add a considerable amount of complexity to the analysis.

#### 3.2.3 Rate Model and Impact of Training

Before proceeding with the description, we define the *perfect case* as the case where the backhaul has an infinite capacity, which leads to perfect global CSI knowledge at the transmitters; so no quantization is needed. Further, we call *imperfect case* the model described previously, where a quantization is performed over the backhaul of limited capacity and a *fixed* number of bits  $B$  is used to quantize each channel matrix.

As explained in the previous subsection, for the perfect case the IA constraints null the ISI and the IUI, and no residual interference exists. On the other hand, for the imperfect case we recall that each (active) transmitter quantizes its local CSI, excluding the direct channel, and sends it to all other (active) transmitters. At the same time, this transmitter receives the quantized local CSI (except the direct channel) of each other transmitter. Let us denote  $\hat{\mathbf{H}}_{ki}$  as the quantization version of  $\mathbf{H}_{ki}$ , which

### 3.2. System Model

can be formed by reshaping vector  $\hat{\mathbf{h}}_{ki}$ . Then, e.g., transmitter  $k$  designs its IA vectors based on:

- ★  $\mathbf{H}_{kk}$ , i.e. perfect direct channel.
- ★  $\hat{\mathbf{H}}_{ik} \forall i \neq k$ , i.e. quantized version of the channels of its local CSI (without the direct channel).
- ★  $\hat{\mathbf{H}}_{li} \forall i, l, i \neq l, i \neq k$ , i.e. the imperfect knowledge it receives from other transmitters.

It can be noticed that although transmitter  $k$  has a perfect version of the  $\mathbf{H}_{ik} \forall i \neq k$ , this transmitter uses the quantized version of these channels in the design of its IA vectors. This is considered so that each transmitter can compute the (same) precoding vectors of other transmitters, where we recall that these vectors are used in the computation of the combining vector of this transmitter. Based on the above, under the imperfect case the IA technique is able to completely cancel the ISI but not the IUI. In other words, under this case we have

$$\begin{aligned} \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \hat{\mathbf{H}}_{ki} \hat{\mathbf{v}}_i^{(j)} &= 0, & \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{ki} \hat{\mathbf{v}}_i^{(j)} &\neq 0, & \forall i \neq k, \forall m, \forall j, \text{ and } k, i \in \mathcal{L}(t), \\ \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(j)} &= 0, & & & \forall j \neq m, \forall k, \text{ and } k \in \mathcal{L}(t), \end{aligned} \quad (3.7)$$

where  $\hat{\mathbf{v}}_k^{(m)}$  and  $\hat{\mathbf{u}}_k^{(m)}$ ,  $\forall k, m$ , are the precoding and combining vectors, respectively, designed under the imperfect case. Similarly to the perfect case, it can be shown that for the imperfect case the following properties hold [90]:

- ★ Vector  $\hat{\mathbf{v}}_i^{(j)}$  is a function of all the cross channels  $\hat{\mathbf{H}}_{kl} \forall k, l, k \neq l$  only.
- ★ Vector  $\hat{\mathbf{u}}_k^{(m)}$  is a function of vectors  $\hat{\mathbf{H}}_{ki} \hat{\mathbf{v}}_i^{(j)} \forall i \neq k, \forall j$  and  $\mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(j)}, \forall j \neq m$ .
- ★ Matrix  $\hat{\mathbf{V}}_k = [\hat{\mathbf{v}}_k^{(1)}, \dots, \hat{\mathbf{v}}_k^{(d_k)}]$  is unitary.

Thus, we can deduce that  $\hat{\mathbf{u}}_k^{(m)}$  is independent of  $\mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)}$ . Using the above, the SINR/SNR for stream  $m$  at active receiver  $k$  can be written as

$$\gamma_k^{(m)} = \begin{cases} \frac{\alpha_{kk} \left| \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)} \right|^2}{\sigma^2 + \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{ki} \hat{\mathbf{v}}_i^{(j)} \right|^2}, & \text{imperfect case} \\ \frac{\alpha_{kk} \left| \left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(m)} \right|^2}{\sigma^2}, & \text{perfect case} \end{cases} \quad (3.8)$$

where we recall that  $\mathbf{v}_k^{(m)}$  and  $\mathbf{u}_k^{(m)}$ ,  $\forall k, m$ , are designed under the perfect case (i.e. using perfect global CSI). As alluded earlier, only a subset  $\mathcal{L}(t)$  of pairs is scheduled at a time-slot, where we recall that  $L(t)$  denotes the number of scheduled pairs at time-slot  $t$ , i.e.  $L(t) = |\mathcal{L}(t)|$ . For notational convenience, we will use signal-to-interference-plus-noise ratio (SINR) as a general notation to denote SNR for the perfect case and SINR for the imperfect case, unless stated otherwise.

We now explain some useful points that are adopted in the rate model. At a given time-slot, the instantaneous rate  $R$  of stream  $m$  of pair  $k$  is transmitted successfully if the SINR  $\gamma_k^{(m)}$  of receiver  $k$  is higher than or equal to a given threshold  $\tau$ ; otherwise, the transmission is not successful and the instantaneous bit rate is 0. Of course, the choice of  $\tau$  depends on  $R$  and vice-versa. The relation (mainly point-to-point relation) between the SINR threshold and the bit rate has been studied widely

### 3.2. System Model

in the literature and therefore lies out of the scope of this chapter. In fact, our analysis is valid for any point-to-point relation between  $\tau$  and  $R$ . Our assumption is simply that a given rate  $R$  is transmitted if the SINR is above a given threshold  $\tau$ . Let us denote by  $\tilde{R}_k(t)$  the assigned rate (in units of bits/slot) for user  $k$  at time-slot  $t$ , thus  $\tilde{R}_k(t)$  is the sum of the assigned rates for all the streams of user  $k$  at time-slot  $t$ . In other words, we have

$$\tilde{R}_k(t) = \sum_{m=1}^{d_k} R \mathbb{1}_{(\gamma_k^{(m)}(t) \geq \tau)}, \quad (3.9)$$

where  $\mathbb{1}_{(\cdot)}$  is the indicator function. For this model, channel acquisition cost is not negligible and should be considered. As mentioned earlier, we consider a system under TDD mode where users send training sequences in the uplink so that the transmitters can estimate their channels. This scheme uses orthogonal sequences among the users, so their lengths are proportional to the number of active users in the system. We assume that acquiring the CSI of one user takes fraction  $\theta$  of the slot. Thus, since we have  $L(t)$  active users, the *actual* rate for transmission to active user  $k$  at time-slot  $t$  is  $(1 - L(t)\theta)\tilde{R}_k(t)$ . We denote this rate by  $D_k(t)$ , i.e.

$$\begin{aligned} D_k(t) &= (1 - L(t)\theta)\tilde{R}_k(t) \\ &= (1 - L(t)\theta) \sum_{m=1}^{d_k} R \mathbb{1}_{(\gamma_k^{(m)}(t) \geq \tau)}. \end{aligned} \quad (3.10)$$

Note that  $D_k(t)$  is set to 0 if pair  $k$  is not active at slot  $t$ .

Under the above setting, the *average rate* for active user  $k$  can be written in function of the transmission *success probability* (i.e. the probability that the corresponding SINR is greater than or equal to a certain threshold) conditioned on the subset of active pairs as

$$\mathbb{E}\{D_k(t) \mid \mathcal{L}(t)\} = (1 - L(t)\theta) \sum_{m=1}^{d_k} R \mathbb{P}\{\gamma_k^{(m)}(t) \geq \tau \mid \mathcal{L}(t)\}. \quad (3.11)$$

It can be noticed that the feedback overhead  $(1 - L(t)\theta)$  scales with the number of active pairs, meaning that when  $L(t)$  is large there will be little time left to transmit in the slot before the channels change again. We point out that if  $L(t) = 1$ , then  $\gamma_k^{(m)}(t)$  is the SINR obtained using the SVD scheme, which will be presented in Section 3.3.

#### 3.2.4 Queue Dynamics, Stability and Scheduling Policy

For each user, we assume that the incoming data is stored in a respective queue (buffer) until transmission, and we denote by  $\mathbf{q}(t) = (q_1(t), \dots, q_N(t))$  the queue length vector. We designate by  $\mathbf{A}(t) = (A_1(t), \dots, A_N(t))$  the vector of number of bits arriving in the buffers in time-slot  $t$ , which is an i.i.d. in time process, independent across users and with  $A_k(t) < A_{\max}$ . The mean arrival rate (in units of bits/slot) for user  $k$  is denoted by  $a_k = \mathbb{E}\{A_k(t)\}$ . We recall that, e.g., user  $k$  will get  $D_k(t)$  served bits per slot if it gets scheduled and zero otherwise. Note that  $D_k(t)$  is finite because  $R$  is finite, so we can define a finite positive constant  $D_{\max}$  such that  $D_k(t) < D_{\max}$ , for  $k = 1, \dots, N$ .

## 3.2. System Model

---

At each time-slot, the CS selects the pairs to schedule based on the queue lengths and average rates in the system. To this end, we suppose that:

- This scheduler has full knowledge of average rate values under different combinations of choosing active pairs, which can be provided offline since an average rate is time-independent,
- At each time-slot, each transmitter sends its queue length to the CS so that it can obtain all the queue dynamics of the system,
- The cost of providing such knowledge to the scheduler will not be taken into account.

After selecting the set of pairs to be scheduled (represented by  $\mathcal{L}(t)$ ), the CS broadcasts this information so that the selected transmitter-receiver pairs activate themselves, and then the active receivers send their pilots in the uplink so that the (active) transmitters can estimate the CSI. It is worth noting that, as alluded previously, if we select a large number of pairs ( $L(t)$ ) for transmission, many pairs can communicate but a high CSI acquisition cost is needed (i.e. this will leave a small fraction of time for transmission). On the other hand, a small  $L(t)$  requires a low acquisition cost, but, at the same time, it allows a few number of simultaneous transmissions. The decision of selecting active pairs is referred simply as the *scheduling policy*. At the  $t$ -th slot, this policy can be represented by an indicator vector  $\mathbf{s}(t) \in \mathcal{S} := \{0, 1\}^N$ , where the  $k$ -th component of  $\mathbf{s}(t)$ , denoted  $s_k(t)$ , is equal to 1 if the  $k$ -th queue (pair) is scheduled or otherwise equal to 0. It can be seen that the cardinality of set  $\mathcal{S}$  is equal to  $|\mathcal{S}| = 2^N$ . Remark that, in terms of notation,  $\mathbf{s}(t)$  and  $\mathcal{L}(t)$  are used to represent the same thing, that is the scheduled pairs at time-slot  $t$ , but they illustrate it differently. Specifically, using  $\mathbf{s}(t)$  the active pairs correspond to the non-zero coordinates (equal to 1), whereas  $\mathcal{L}(t)$  contains the indexes (i.e. positions) of these pairs. Let  $\mathfrak{L}$  be the set of all possible  $\mathcal{L}(t)$ .

Now, using the definition of  $D_k(t)$ , which was provided earlier, the queueing *dynamics* (i.e. how the queue lengths evolve over time) can be described by the following

$$q_k(t+1) = \max\{q_k(t) - D_k(t), 0\} + A_k(t), \quad \forall k \in \{1, \dots, N\}, \forall t \in \{0, 1, \dots\}, \quad (3.12)$$

where we note that  $D_k(t)$  depends on the scheduling policy.

### 3.2.4.1 Max-Weight Scheduling Policy

If the arrival rates are known, the queue stability can be achieved by pre-defined time-sharing between scheduling different subsets of queues [63]. However, in practice, the arrival rates are usually unknown. For this case, the queues can be stabilized using a policy that considers the knowledge of average rates and queue lengths. Such a policy is called Max-Weight [63, 103], i.e. it maximizes a weighted sum, and is described below.

$$\Delta^* : \mathbf{s}(t) = \arg \max_{\mathbf{s} \in \mathcal{S}} \{\mathbf{r}(\mathbf{s}) \cdot \mathbf{q}(t)\}, \quad (3.13)$$

where “ $\cdot$ ” is the scalar (dot) product, and  $\mathbf{r}(\mathbf{s})$  is constructed by replacing the non-zero coordinates of  $\mathbf{s}$ , which represent the selected pairs, with their corresponding average rate values. Recall that  $\mathcal{L}$  represents the positions (indexes) of the non-zero coordinates of  $\mathbf{s}$ ; for example, if  $\mathbf{s} = (0, 0, 1, 1, 1, 0)$ , then  $\mathcal{L} = \{3, 4, 5\}$ , meaning that pairs 3, 4 and 5 are active. Hence, vector  $\mathbf{r}(\mathbf{s})$  contains  $\mathbb{E}\{D_k \mid \mathcal{L}\}$  at

### 3.3. Success Probabilities and Average Rates

---

position  $k$  if the  $k$ -th coordinate of  $\mathbf{s}$  is '1' and 0 if this coordinate is '0'. The following lemma states the optimality of the above policy, where  $\Lambda$  denotes the system stability region.

**Lemma 1.** *Under the adopted system, the scheduling policy  $\Delta^*$  is throughput optimal, meaning that it can stabilize the system for every mean arrival rate vector in  $\Lambda$ .*

*Proof.* We show that policy  $\Delta^*$  stabilizes the system for all  $\mathbf{a} \in \Lambda$  by proving that the Markov chain of the corresponding system is positive recurrent. For this purpose, we use Foster's theorem. The proof is similar to the one provided in Section 2.3 in Chapter 2, we thus omit it for sake of brevity.  $\square$

It is worth noting that, in the above lemma, the stability region  $\Lambda$  is the result of the adopted system model. More in detail, this region depends mainly, but not only, on the following points:

- The use of IA, for  $L \geq 2$ , and SVD, for  $L = 1$ , as interference management techniques.
- Accounting for the probing cost.
- The scheduling is done in a centralized manner, i.e. in each slot the CS schedules the set of active pairs (before receiving any feedback) that must send their pilots.
- The quantization process over the backhaul (for the imperfect case).

#### 3.2.4.2 Computational Complexity of Max-Weight Policy $\Delta^*$

For such an optimal policy, an important factor to investigate is the computational complexity (CC), which we derive next. Because what  $\Delta^*$  looks for is the maximum over  $2^N$  possible values, due to  $2^N$  combinations, thus it takes  $\mathcal{O}(2^N)$  after computing all values  $\mathbf{r}(\mathbf{s}) \cdot \mathbf{q}(t)$  to find the maximum value. Note that for two fixed vectors we can compute this product in time  $\mathcal{O}(N)$ . Thus we would have  $\mathcal{O}(N2^N)$  ignoring computing  $\mathbf{r}(\mathbf{s})$ , which can be done offline. We can notice that this computational complexity increases considerably with the maximum number of pairs  $N$ .

## 3.3 Success Probabilities and Average Rates

In this section, we give the expression for the success probability and, subsequently, the expression for the average transmission rate for each of the imperfect and perfect cases. Then, we present the SVD scheme (used when  $L(t) = 1$ ) and derive its average rate expression.

### 3.3.1 Average Rate Expression for IA

For the calculation of the average rate, we next provide a proposition in which we calculate the success probabilities under the considered setting.



### 3.3. Success Probabilities and Average Rates

**Proposition 1.** *The probability that the received SINR corresponding to stream  $m$  of active user  $k$  exceeds a threshold  $\tau$  given that  $\mathcal{L}(t)$  is the set of scheduled pairs (including pair  $k$ ) can be given by*

$$\mathbb{P}\left\{\gamma_k^{(m)}(t) \geq \tau \mid \mathcal{L}(t)\right\} = \begin{cases} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \text{MGF}_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right), & \text{imperfect case} \\ e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}}, & \text{perfect case} \end{cases} \quad (3.14)$$

in which

$$RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \left( \hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{ki} \hat{\mathbf{v}}_i^{(j)} \right|^2 \quad (3.15)$$

is the residual interference, which appears in the denominator of  $\gamma_k^{(m)}$  in the imperfect case, and  $\text{MGF}_{RI_k^{(m)}}(\cdot)$  stands for the moment-generating function (MGF) of  $RI_k^{(m)}$ .

*Proof.* Please refer to Appendix 3.9.1 for the proof.  $\square$

In the above result, the success probability expression in the imperfect case is given in function of the MGF of the leakage interference  $RI_k^{(m)}$ . It is noteworthy to mention that the explicit expression of this MGF will be given afterwards during the average rate calculations. But first, let us focus on the expression  $RI_k^{(m)}$ . Indeed, we have

$$\begin{aligned} RI_k^{(m)} &= \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \left( \hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{ki} \hat{\mathbf{v}}_i^{(j)} \right|^2 \\ &= \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \mathbf{h}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 \\ &= \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \|\mathbf{h}_{ki}\|^2 \sum_{j=1}^{d_i} \left| \tilde{\mathbf{h}}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2, \end{aligned} \quad (3.16)$$

in which  $\mathbf{T}_{k,i}^{(m,j)} = \hat{\mathbf{v}}_i^{(j)} \otimes ((\hat{\mathbf{u}}_k^{(m)})^H)^T$  (where  $\otimes$  is the Kronecker product) and  $\tilde{\mathbf{h}}_{ki}$  is the normalized vector of channel  $\mathbf{h}_{ki}$ , i.e.  $\tilde{\mathbf{h}}_{ki} = \frac{\mathbf{h}_{ki}}{\|\mathbf{h}_{ki}\|}$ . Note that  $((\hat{\mathbf{u}}_k^{(m)})^H)^T$  is nothing but the conjugate of  $\hat{\mathbf{u}}_k^{(m)}$ . Following the model used in [104], the channel direction  $\tilde{\mathbf{h}}_{ki}$  can be written as follows

$$\tilde{\mathbf{h}}_{ki} = \sqrt{1 - e_{ki}} \hat{\mathbf{h}}_{ki} + \sqrt{e_{ki}} \mathbf{w}_{ki}, \quad (3.17)$$

where  $\hat{\mathbf{h}}_{ki}$  is the channel quantization vector of  $\mathbf{h}_{ki}$  and  $\mathbf{w}_{ki}$  is a unit norm vector isotropically distributed in the null space of  $\hat{\mathbf{h}}_{ki}$ , with  $\mathbf{w}_{ki}$  independent of  $e_{ki}$ . Since IA is performed based on the quantized CSI  $\hat{\mathbf{h}}_{ki}$ , we get

$$\begin{aligned} \left| \tilde{\mathbf{h}}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 &= \left| \sqrt{1 - e_{ki}} \hat{\mathbf{h}}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} + \sqrt{e_{ki}} \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 \\ &= e_{ki} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2. \end{aligned} \quad (3.18)$$

### 3.3. Success Probabilities and Average Rates

Therefore,  $RI_k^{(m)}$  can be rewritten as

$$RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \|\mathbf{h}_{ki}\|^2 e_{ki} \sum_{j=1}^{d_i} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2. \quad (3.19)$$

Based on the above results, we now have all the required materials to derive the average rate expressions for both the perfect and imperfect cases. We recall that if  $\mathcal{L}(t)$  is the subset of scheduled pairs, then the general formula of the average rate of active user  $k$  can be given as

$$\mathbb{E} \{D_k(t) \mid \mathcal{L}(t)\} = (1 - L(t)\theta) \sum_{m=1}^{d_k} R \mathbb{P} \left\{ \gamma_k^{(m)}(t) \geq \tau \mid \mathcal{L}(t) \right\}. \quad (3.20)$$

The explicit rate expressions are provided in the following theorem.

**Theorem 4.** *Given a subset of scheduled pairs,  $\mathcal{L}(t)$ , the average rate of user  $k$  ( $\in \mathcal{L}(t)$ ) is:*

- *For the imperfect case:*

$$(1 - L(t)\theta) d_k R e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} MGF_{RI_k^{(m)}} \left( -\frac{\tau}{\alpha_{kk}} \right), \quad (3.21)$$

in which the MGF can be written as

$$MGF_{RI_k^{(m)}} \left( -\frac{\tau}{\alpha_{kk}} \right) = \prod_{i \in \mathcal{L}(t), i \neq k} \left( \frac{\alpha_{ki} \tau d_i}{\alpha_{kk} 2^{\frac{B}{Q}}} + 1 \right)^{-Q} {}_2F_1(c_{2i}, Q; c_{1i} + c_{2i}; \frac{1}{\frac{\alpha_{kk} 2^{\frac{B}{Q}}}{\alpha_{ki} \tau d_i} + 1}). \quad (3.22)$$

In the above equation,  ${}_2F_1$  represents the Hypergeometric function,  $c_{1i} = (Q + 1)Q^{-1}d_i - Q^{-1}$  and  $c_{2i} = (Q - 1)c_{1i}$ , with  $Q = N_t N_r - 1$ .

- *For the perfect case:*

$$(1 - L(t)\theta) d_k R e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}}. \quad (3.23)$$

*Proof.* Please refer to Appendix 3.9.2 for the proof. □

### 3.3.2 SVD Scheme and its Average Rate Expression

#### 3.3.2.1 SVD Scheme

In the case where  $L(t) = 1$ , there is only one active pair (which we denote by index  $k$ ), and the system is reduced to a point-to-point MIMO system. We recall that receiver  $k$  sends its pilot sequence, then transmitter  $k$  estimates perfectly the channel matrix; clearly, here the probing cost is  $(1 - \theta)$ . With one active pair, the only source of interference is the ISI caused by the transmitter itself. To manage

### 3.3. Success Probabilities and Average Rates

this interference problem, we apply SVD technique. Specifically, by the singular value decomposition theorem we have

$$\mathbf{H}_{kk} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H, \quad (3.24)$$

where  $\mathbf{U}_k$  and  $\mathbf{V}_k$  are  $N_r \times N_r$  and  $N_t \times N_t$  unitary matrices, respectively.  $\mathbf{\Sigma}_k$  is an  $N_r \times N_t$  diagonal matrix with the singular values of  $\mathbf{H}_{kk}$  in diagonal. These singular values are denoted by  $\sqrt{\lambda_k^{(m)}}$ . Note that here we consider the same assumptions and parameters that we have used for IA. We assume that active transmitter has  $d_k$  ( $\leq \min(N_t, N_r)$ ) data streams to transmit to its receiver. We also assume that  $\mathbf{H}_{kk}$  is full rank, meaning that its rank is given by  $\min(N_t, N_r)$ ;  $d_k$  should be less than or equal to the rank of matrix  $\mathbf{H}_{kk}$ , which is satisfied under our setting. Further, we assume that the power  $P$  is equally allocated among the  $d_k$  data streams.

We define  $\mathbf{x}_k$  to be the following  $N_t \times 1$  vector:  $\mathbf{x}_k = (x_k^{(1)}, \dots, x_k^{(d_k)}, 0, \dots, 0)^T$ , where we recall that  $x_k^{(m)}$  represents stream  $m$  of pair  $k$ . Under SVD, the transmitter sends vector  $\mathbf{V}_k \mathbf{x}_k$  instead of  $\mathbf{x}_k$ , thus the received signal, which we denote by  $\mathbf{y}_k$ , can be written as

$$\mathbf{y}_k = \sqrt{\frac{\zeta_{kk}P}{d_k}} \mathbf{H}_{kk} \mathbf{V}_k \mathbf{x}_k + \mathbf{z}_k, \quad (3.25)$$

where we recall that  $\mathbf{H}_{kk}$  denotes the  $N_r \times N_t$  channel matrix with i.i.d. zero mean and unit variance complex Gaussian entries,  $\mathbf{z}_k$  is the additive white complex Gaussian noise vector with zero mean and covariance matrix  $\sigma^2 \mathbf{I}_{N_r}$ , and  $\zeta_{kk}$  stands for the path loss coefficient. Then, at the receiver we multiply the corresponding received signal by  $\mathbf{U}_k^H$  to detect the desired signal. Hence, after multiplying by  $\mathbf{U}_k^H$ , we eventually obtain

$$\begin{aligned} \mathbf{U}_k^H \mathbf{y}_k &= \sqrt{\frac{\zeta_{kk}P}{d_k}} \mathbf{U}_k^H \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H \mathbf{V}_k \mathbf{x}_k + \mathbf{U}_k^H \mathbf{z}_k \\ &= \sqrt{\frac{\zeta_{kk}P}{d_k}} \mathbf{\Sigma}_k \mathbf{x}_k + \mathbf{U}_k^H \mathbf{z}_k. \end{aligned} \quad (3.26)$$

Recall that  $\mathbf{U}_k$  and  $\mathbf{V}_k$  are unitary matrices, so  $\mathbf{U}_k^H \mathbf{z}_k$  and  $\mathbf{V}_k \mathbf{x}_k$  have the same distributions as  $\mathbf{z}_k$  and  $\mathbf{x}_k$ , respectively. Based on the above, it can be noticed that the ISI is completely canceled.

#### 3.3.2.2 Average Rate for SVD

The equivalent MIMO system can be seen as  $d_k$  uncoupled parallel sub-channels. The SNR for stream  $m$  can be written as the following

$$\gamma_k^{(m)} = \frac{\zeta_{kk}P}{d_k \sigma^2} \lambda_k^{(m)}, \quad 1 \leq m \leq d_k. \quad (3.27)$$

Let  $m_1 = \max(N_t, N_r)$  and  $m_2 = \min(N_t, N_r)$ . It was shown in [94] that the distribution of any one of the unordered eigenvalues, which we denote by  $\lambda$ , is given by

$$p(\lambda) = \frac{1}{m_2} \sum_{n=0}^{m_2-1} \frac{n!}{(n+m_1-m_2)!} [L_n^{m_1-m_2}(\lambda)]^2 \lambda^{m_1-m_2} e^{-\lambda}, \quad \lambda \geq 0, \quad (3.28)$$

### 3.4. Stability Analysis for the Symmetric Case

---

where  $L_n^{m_1-m_2}(x)$  is the associated Laguerre polynomial of degree (order)  $n$  and is given by

$$L_n^{m_1-m_2}(\lambda) = \sum_{l=0}^n (-1)^l \frac{(n+m_1-m_2)!}{(n-l)!(m_1-m_2+l)!} \frac{\lambda^l}{l!}. \quad (3.29)$$

Adopting the same rate model as for IA, the average rate of the active user can be written as

$$(1-\theta)d_k R \mathbb{P} \left\{ \gamma_k^{(m)} \geq \tau \right\}. \quad (3.30)$$

Based on the above, the average rate expression for SVD if pair  $k$  is the active pair, which we denote by  $r_{\text{svd},k}$ , is provided in the following statement.

**Proposition 2.** *Under SVD technique, the average rate for the active pair is given by*

$$r_{\text{svd},k} = (1-\theta)d_k R \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \Gamma(j+m_1-m_2+1, \frac{d_k \sigma^2 \tau}{\zeta_{kk} P}), \quad (3.31)$$

where  $\Gamma(\cdot, \cdot)$  stands for the upper incomplete Gamma function, and where

$$\Omega_n = n! (m_2(n+m_1-m_2)!)^{-1}, \quad (3.32)$$

$$\kappa_j = \sum_{i=0}^j \omega_i \omega_{j-i}, \quad (3.33)$$

$$\omega_l = (-1)^l (n+m_1-m_2)! ((n-l)!(m_1-m_2+l)!l!)^{-1}, \text{ with } \omega_l = 0 \text{ if } l > n. \quad (3.34)$$

*Proof.* Please refer to Appendix 3.9.3 for the proof. □

## 3.4 Stability Analysis for the Symmetric Case

In this section, we consider a symmetric system in which the path loss coefficients have the same value, namely  $\zeta = \zeta_{ki}$ ,  $\forall k, i$ , and all the pairs have equal number of data streams, namely  $d = d_k$ ,  $\forall k$ ; note that we still assume different average arrival rates. Under this system, the *feasibility condition* of IA, given in [105], becomes  $N_t + N_r \geq (L+1)d$ , which we assume is satisfied here. We recall that at each time-slot, for the selected pairs, rate  $R$  can be supported if the corresponding SINR is greater than or equal to a given threshold  $\tau$ ; otherwise, the packets are not received correctly and the instantaneous bit-rate is considered to be equal to 0. Let  $\alpha = \frac{P\zeta}{d}$ . For notational convenience, in the remainder of the chapter we drop the notation for dependence of  $\mathcal{L}$  and  $L$  on  $t$ .

### 3.4.1 Average Rate Expressions and their Variation

In the following, we provide the average rate expressions and study their variations under the above assumptions for the different cases in the system.

#### 3.4.1.1 If the Number of Scheduled Pairs $L \geq 2$

The IA technique is applied, and thus the SINR of stream  $m$  at user  $k$  can be given by

$$\gamma_k^{(m)} = \begin{cases} \frac{\alpha \left| \left( \hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)} \right|^2}{\sigma^2 + \sum_{i \in \mathcal{L}, i \neq k} \alpha \|\mathbf{h}_{ki}\|^2 e_{ki} \sum_{j=1}^d \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2}, & \text{imperfect case} \\ \frac{\alpha \left| \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(m)} \right|^2}{\sigma^2}, & \text{perfect case} \end{cases} \quad (3.35)$$

where the sum in the denominator of the first expression results from equation (3.19). As explained in the previous sections, if  $\mathcal{L}$  is the subset of scheduled pairs, the average transmission rate per active user is given by  $(1 - L\theta)dR \mathbb{P} \left\{ \gamma_k^{(m)} \geq \tau \mid \mathcal{L} \right\}$ . Relying on the average rate expressions in Theorem 4, we get the following results.

**a) Imperfect Case:** The average transmission rate for an active user  $k \in \mathcal{L}$  can be given by

$$(1 - L\theta)dR e^{-\frac{\sigma^2 \tau}{\alpha}} \left[ \left( d\tau 2^{-\frac{B}{Q}} + 1 \right)^{-Q} {}_2F_1 \left( c_2, Q; c_1 + c_2; \left( 2^{\frac{B}{Q}} (d\tau)^{-1} + 1 \right)^{-1} \right) \right]^{L-1}, \quad (3.36)$$

where  ${}_2F_1$  is the Hypergeometric function,  $c_1 = (Q + 1)Q^{-1}d - Q^{-1}$  and  $c_2 = (Q - 1)c_1$ . It can be noticed that this average rate is independent of the identity of active user  $k$  and the  $L - 1$  other active pairs, yet depends on the cardinality  $L$  of subset  $\mathcal{L}$ . By denoting this rate as  $r(L)$ , the expression in (3.36) can be re-written as

$$r(L) = (1 - L\theta)dR e^{-\frac{\sigma^2 \tau}{\alpha}} F^{L-1}, \quad (3.37)$$

in which

$$F = \left( d\tau 2^{-\frac{B}{Q}} + 1 \right)^{-Q} {}_2F_1 \left( c_2, Q; c_1 + c_2; \left( 2^{\frac{B}{Q}} (d\tau)^{-1} + 1 \right)^{-1} \right). \quad (3.38)$$

Consequently, the total average transmission rate of the system is given by

$$r_T(L) = L(1 - L\theta)dR e^{-\frac{\sigma^2 \tau}{\alpha}} F^{L-1}. \quad (3.39)$$

Studying the variation of these rate functions w.r.t. the number of active pairs  $L$  is essential for the stability analysis and is thus described by the following lemma.

**Lemma 2.** *Given a number of pairs to be scheduled,  $L$ , the average transmission rate is a decreasing function with  $L$ , whereas the total average transmission rate is increasing from 0 to  $L_0$  and decreasing*

### 3.4. Stability Analysis for the Symmetric Case

---

from  $L_0$  to  $\frac{1}{\theta}$ , meaning that  $r_T$  reaches its maximum at  $L_0$ , where  $L_0 < \frac{1}{2\theta}$  and is given by

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F}}}{2}. \quad (3.40)$$

*Proof.* The proof is provided in Appendix 3.9.4. □

From (3.40) we can notice that  $L_0$  is in general a real value. But, since it represents a number of users, we need to find the best and nearest integer to  $L_0$ , i.e. best in terms of maximizing the total average rate function. We denote this integer by  $L_I$  and we assume without loss of generality that  $2 \leq L_I \leq N$ . The following simple procedure can be used to compute  $L_I$ :

Let  $L_{01} = \lfloor L_0 \rfloor$  and  $L_{02} = \lceil L_0 \rceil$ , i.e. the largest previous and the smallest following integer of  $L_0$ , respectively. If  $r_T(L_{02}) \geq r_T(L_{01})$ , put  $L_I = L_{02}$ ; otherwise  $L_I = L_{01}$ .

**b) Perfect Case:** In this case no residual interference exists and the corresponding SINR expression is given in (3.35). Using Theorem 4, the *average* and *total average transmission rate* expressions can be given, respectively, by

$$\mu(L) = (1 - L\theta)dR e^{-\frac{\sigma^2\tau}{\alpha}}, \quad (3.41)$$

$$\mu_T(L) = L(1 - L\theta)dR e^{-\frac{\sigma^2\tau}{\alpha}}. \quad (3.42)$$

A similar observation to that given in the first case can be made here, that is, the rate functions depend only on the cardinality  $L$  of  $\mathcal{L}$  and not on the subset itself. Notice that  $\mu(L)$  is a decreasing function with  $L$ , while  $\mu_T(L)$  is concave at  $\frac{1}{2\theta}$ . Since  $\frac{1}{2\theta}$  represents a number of pairs, we can use a similar procedure to that proposed for the imperfect case to find the best and nearest integer to  $\frac{1}{2\theta}$ . For the remainder of this chapter, we denote this integer by  $L_P$ .

#### 3.4.1.2 If the Number of Scheduled Pairs $L = 1$

The SVD technique is applied, and thus the SINR of stream  $m$  at user  $k$  becomes

$$\gamma_k^{(m)} = \frac{\zeta P}{d\sigma^2} \lambda_k^{(m)}. \quad (3.43)$$

Based on Proposition 2, the average rate (which is also the total average rate) can be written as

$$r_{\text{svd},k} = (1 - \theta)dR \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \Gamma(j + m_1 - m_2 + 1, \frac{d\sigma^2\tau}{\zeta P}). \quad (3.44)$$

Note that the above expression is independent of the identity of the active pair.

For the rest of the chapter, we use  $r_{\text{svd}}$  to denote this expression. Obviously, here the average rate is independent of the case (i.e. perfect or imperfect) we adopt since anyway the backhaul is not used.

### 3.4.2 Stability Regions and Scheduling Policies

After presenting results on the average rate functions, we now provide a precise characterization of the stability region of the adopted system under both the imperfect and perfect cases.

#### 3.4.2.1 Imperfect Case

We first define subset  $S_L = \{\mathbf{s} \in \mathcal{S} : \|\mathbf{s}\|_1 = L\}$ , where we recall that  $\mathbf{s} \in \mathbb{Z}^N$  is the scheduling vector whose coordinates take values 0 or 1 (see Section 3.2); note that  $S_L$  is the subset of scheduling decision vectors for which the number of active pairs is equal to  $L$ . Given a number  $L \geq 2$ , the subset of average rate vectors is defined as  $I_L = \{r(L)\mathbf{s} : \mathbf{s} \in S_L\}$ . For  $L = 1$ , the subset of average rate vectors is defined as  $I_1 = \{r_{\text{svd}}\mathbf{s} : \mathbf{s} \in S_1\}$ . We define set  $\mathcal{I} = \{\mathbf{0}, I_1, I_2, \dots, I_{L_I}\}$ , i.e. it contains the origin point  $\mathbf{0}$  and the set of average rate vectors when the number of active pairs  $L$  is between 1 and  $L_I$ . We also define set  $\bar{\mathcal{I}} = \{I_{L_I+1}, \dots, I_N\}$ , i.e. it contains the set of average rate vectors for which  $L$  is between  $L_I + 1$  and  $N$ . Note that in terms of cardinality we have  $|\mathcal{I}| + |\bar{\mathcal{I}}| = |\mathcal{S}|$ . Using these definitions, we can state the following lemma which will be useful to characterize the stability region of the system.

**Lemma 3.** *Each point in the set  $\bar{\mathcal{I}}$  is inside the convex hull of  $\mathcal{I}$ . Consequently, this hull will also contain any point in the convex hull of  $\bar{\mathcal{I}}$ .*

*Proof.* We provide the following lemma that will help us in the proof of Lemma 3.

**Lemma 4.** *For any point  $\mathbf{s}_{i,L+1} \in S_{L+1}$ , there exists a point on the convex hull of  $S_L$  that is in the same direction toward the origin as  $\mathbf{s}_{i,L+1}$ . Furthermore,  $\mathbf{s}_{i,L+1}$  can be written as  $\frac{L+1}{L} \times$  its corresponding point on the convex hull of  $S_L$ .*

Please refer to Appendix 3.9.5 for the complete proof of Lemma 3, which includes the proof of Lemma 4.  $\square$

Lemma 3 means that by increasing the number of active pairs  $L$  beyond  $L_I$ , the set of achievable average rates will not increase. Based on the above, we are now able to characterize the stability region of the considered system. We recall that this region is defined as the set of all mean arrival rate vectors for which the system is stable. Here, this region is given by the following theorem.

**Theorem 5.** *The stability region of the system in the symmetric case with limited backhaul can be characterized as*

$$\Lambda_I = \mathcal{CH}\{\mathcal{I}\} = \mathcal{CH}\{\mathbf{0}, I_1, I_2, \dots, I_{L_I}\}, \quad (3.45)$$

where  $\mathcal{CH}$  represents the convex hull.

*Proof.* Please refer to Appendix 3.9.6 for the proof.  $\square$

Unlike classical results in which the stability region is given by the convex hull over all possible decisions, here the characterization is more precise and is defined by the decision subsets  $S_L$  for all

### 3.4. Stability Analysis for the Symmetric Case

---

$L \leq L_I$ . In addition, this theorem provides an exact specification of the corner points (i.e. vertices) of the stability region, meaning that this region is characterized by the set  $\mathcal{I}$  and not by the whole set  $\mathcal{I} \cup \bar{\mathcal{I}}$ . An additional point to note is that  $\Lambda_I$  is a convex polytope in the  $N$ -dimensional space  $\mathbb{R}_+^N$ .

In order to choose the active pairs at each time-slot, we use the Max-Weight scheduling policy defined earlier (see (3.13)). Under the symmetric and imperfect case, this policy, which we denote as  $\Delta_I^*$ , selects decision vector  $\mathbf{s}(t)$  that yields the following max

$$\Delta_I^* : \max \left\{ \max_{\mathbf{s} \in S_L, 2 \leq L \leq N} \{r(\|\mathbf{s}\|_1) \mathbf{s} \cdot \mathbf{q}(t)\}, \max_{\mathbf{s} \in S_1} \{r_{\text{svd}} \mathbf{s} \cdot \mathbf{q}(t)\} \right\}, \quad (3.46)$$

where  $\|\mathbf{s}\|_1$  gives the number of '1' coordinates in  $\mathbf{s}$  (or equivalently, the number of active pairs  $L$ ). Recall that these non-zero coordinates indicate which pairs to schedule. The above policy chooses the subset of pairs that should be active at time-slot  $t$ , which is represented by vector  $\mathbf{s}(t)$ . As mentioned earlier, if only one pair is selected to be active, then we use SVD technique, otherwise we use IA technique. For the proposed policy, the following proposition holds.

**Proposition 3.** *The scheduling policy  $\Delta_I^*$  is throughput optimal. In other words,  $\Delta_I^*$  stabilizes the system (under the imperfect case) for every arrival rate vector  $\mathbf{a} \in \Lambda_I$ .*

*Proof.* The proof can be done in the same way as the proof of Lemma 1 and is thus omitted to avoid repetition.  $\square$

Based on the analysis done at the end of Section 3.2, it was shown that applying policy  $\Delta^*$  will result in a computational complexity (CC) of  $\mathcal{O}(N2^N)$ . The same holds here for policy  $\Delta_I^*$ . Consequently, a moderately large  $N$  will lead to considerably high CC. Recall that this analysis is for the *classical* implementation of the Max-Weight algorithm, that is, finding the maximum over  $2^N$  products of two vectors. However, in our case the implementation of this algorithm does not require all this complexity. This is due to the fact that all the active users have the same average transmission rate. This structural property allows us to propose an equivalent reduced CC implementation of  $\Delta_I^*$ , which we provide in Algorithm 1.

---

**Algorithm 1 :** A Reduced Computational Complexity Implementation of  $\Delta_I^*$

---

- 1: Sort the queues in a descending order; break ties arbitrarily.
  - 2: Initialize  $L_s = 1$  and  $prod = 0$ .
  - 3: Set  $r(1) = r_{\text{svd}}$ . For  $L \geq 2$ , let  $l = L$  and  $r(l) = r(L)$ .
  - 4: **for**  $l = 1 : 1 : N$  **do**
  - 5:     Consider  $sum_l =$  sum of the lengths of the first  $l$  queues (i.e. sum of the  $l$  biggest queue lengths).
  - 6:     **if**  $r(l) sum_l > prod$  **then**
  - 7:         Put  $L_s = l$  and  $prod = r(L_s) sum_{L_s}$
  - 8:     **end if**
  - 9: **end for**
  - 10: Schedule pairs corresponding to the first  $L_s$  queues.
- 

The proposed implementation in Algorithm 1 depends essentially on two steps: the “sorting algorithm” and the “for loop”. We assume that we use a sorting algorithm of complexity  $\mathcal{O}(N^2)$ , such as



### 3.4. Stability Analysis for the Symmetric Case

---

the “bubble sort” algorithm. For the “for loop”, the (worst case) complexity is also  $\mathcal{O}(N^2)$  since this loop is executed  $N$  times (i.e. iterations) and every iteration has another dependency to  $N$ . Therefore, the computational complexity of the proposed implementation is  $\mathcal{O}(N^2 + N^2)$ , or equivalently  $\mathcal{O}(N^2)$ , which is very small compared to  $\mathcal{O}(N2^N)$ , especially for large  $N$ .

#### 3.4.2.2 Perfect Case

A similar study to that done for the imperfect case can be adopted here. To begin with, for  $L \geq 2$  we define  $P_L = \{\mu(L)\mathbf{s} : \mathbf{s} \in S_L\}$ ; i.e.  $P_L$  is the subset of average rate vectors for which the number of active pairs is  $L$ . Recall that  $S_L = \{\mathbf{s} \in \mathcal{S} : \|\mathbf{s}\|_1 = L\}$ . For  $L = 1$  we define  $P_1 = \{r_{\text{svd}}\mathbf{s} : \mathbf{s} \in S_1\}$ . As seen earlier, the total average rate for IA given in (3.42) reaches its maximum at  $\frac{1}{2\theta}$  for which the best and nearest integer is denoted by  $L_P$ , where we assume without loss of generality that  $2 \leq L_P \leq N$ . In addition, the average rate  $\mu(L)$  decreases with  $L$ .

Under the above observations, the stability region can be characterized as follows.

**Theorem 6.** *For the symmetric system with unlimited backhaul, the stability region can be defined as the following*

$$\Lambda_P = \mathcal{CH} \{\mathbf{0}, P_1, P_2, \dots, P_{L_P}\}. \quad (3.47)$$

*Proof.* The proof is very similar to that of Theorem 5, just consider the average rate functions  $\mu(L)$  and  $\mu_T(L)$  instead of  $r(L)$  and  $r_T(L)$ ; so we omit this proof to avoid redundancy.  $\square$

To achieve the stability region that is characterized in the above, we use the Max-Weight policy, which we denote here as  $\Delta_P^*$ . Specifically, this *optimal* policy (i.e. it can achieve  $\Lambda_P$ ) selects decision vector  $\mathbf{s}(t)$  that yields the following max

$$\Delta_P^* : \max \left\{ \max_{\mathbf{s} \in S_L, 2 \leq L \leq N} \{\mu(\|\mathbf{s}\|_1) \mathbf{s} \cdot \mathbf{q}(t)\}, \max_{\mathbf{s} \in S_1} \{r_{\text{svd}} \mathbf{s} \cdot \mathbf{q}(t)\} \right\}. \quad (3.48)$$

As observed in the imperfect case, applying this policy using its classical implementation will result in a CC of  $\mathcal{O}(N2^N)$ . Hence, to avoid a high complexity for large  $N$ , and since the structural properties of this policy and those of policy  $\Delta_I^*$  are similar, the equivalent implementation proposed for the imperfect case can be applied here but after replacing  $r(L)$  with  $\mu(L)$ . Consequently, we get a reduced complexity of  $\mathcal{O}(N^2)$ .

### 3.4.3 Conditions under which IA Provides a Gain in terms of Queueing Stability

In this subsection, we investigate the conditions under which the use of IA can provide a queueing stability gain. We provide this investigation under the imperfect case, while noting that a similar analysis can be done under the perfect case.

### 3.4. Stability Analysis for the Symmetric Case

Based on Theorem 5, we can notice that in the characterization of the stability region the vertices that correspond to IA are given by the subsets  $I_2, I_3, \dots, I_{L_I}$ . On the other hand, the vertices that correspond to SVD can be found in  $I_1$ . In order for IA to provide a queueing stability gain, at least one of its vertices should be outside the part of the stability region that is yielded by SVD, where this part is nothing but the convex hull  $\mathcal{CH}\{\mathbf{0}, I_1\}$ . In the following we provide a simple example for which we illustrate the different shapes of the stability region.

Example: Let  $N = 2$ . Here there is only one vertex of IA, given by point  $(r(2), r(2))$ . The vertices of SVD are  $(r_{\text{svd}}, 0)$  and  $(0, r_{\text{svd}})$ . Hence, the part of the stability region resulting from SVD is  $\mathcal{CH}\{\mathbf{0}, I_1\} = \mathcal{CH}\{\mathbf{0}, (r_{\text{svd}}, 0), (0, r_{\text{svd}})\}$ . Depending on whether vertex  $(r(2), r(2))$  is inside this convex hull, two stability region shapes are possible. These shapes are illustrated in Figure 3.3. We point out that the illustrations in Figure 3.3 are not the results of a specific setting and are only provided as general illustrations in order to help understanding the analysis.

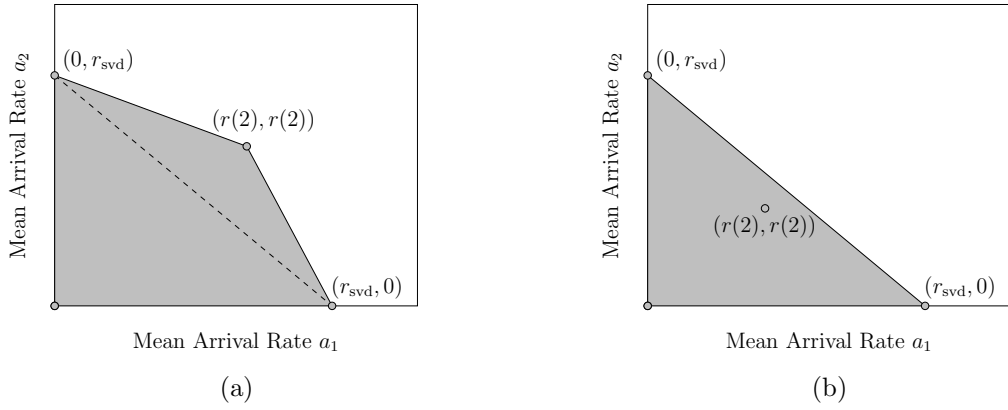


Figure 3.3: Stability region (in gray) for the symmetric system under the imperfect case, where  $N = 2$ . (a) IA provides a queueing stability gain, and (b) IA does not provide a queueing stability gain.

Based on the above observations, we now provide the conditions under which IA yields a queueing stability gain. These conditions are given (for the perfect and imperfect cases) as follows.

**Proposition 4.** *For the symmetric system under the imperfect case, IA provides a queueing stability gain if and only if there exists a number  $L$  such that  $Lr(L) > r_{\text{svd}}$ , with  $2 \leq L \leq L_I$ . For the same system but under the perfect case, we get a similar result, that is to say, IA yields a queueing stability gain if and only if there exists a number  $L$  such that  $L\mu(L) > r_{\text{svd}}$ , with  $2 \leq L \leq L_P$ .*

*Proof.* The proof is provided in Appendix 3.9.7. □

Based on the above, it can be noticed that if there exists an  $L$  that satisfies  $Lr(L) > r_{\text{svd}}$  and  $2 \leq L \leq L_I$ , the characterization of the stability region  $\Lambda_I$  can be more precise since the points in  $I_2, \dots, I_{L-1}$  are inside the stability region part resulting from SVD, and the only vertices of IA that are outside this part are given by the subsets  $I_L, \dots, I_{L_I}$ . In addition, if we cannot find an  $L$  that satisfies the above conditions, then all the vertices of IA will be inside the region part corresponding to SVD. These observations lead us to the following remark; recall that  $P_1 = I_1$ .

### 3.4. Stability Analysis for the Symmetric Case

---

**Remark 1.** For the imperfect case, if there exists an  $L$  such that  $Lr(L) > r_{\text{svd}}$ , with  $2 \leq L \leq L_I$ , a more precise characterization of  $\Lambda_I$  can be given as

$$\Lambda_I = \mathcal{CH}\{\mathbf{0}, I_1, I_L, \dots, I_{L_I}\}. \quad (3.49)$$

If such an  $L$  does not exist, then the characterization reduces to:  $\Lambda_I = \mathcal{CH}\{\mathbf{0}, I_1\}$ .

For the perfect case, if we can find an  $L$  such that  $L\mu(L) > r_{\text{svd}}$ , with  $2 \leq L \leq L_P$ , we get

$$\Lambda_P = \mathcal{CH}\{\mathbf{0}, P_1, P_L, \dots, P_{L_P}\}. \quad (3.50)$$

Otherwise, the characterization of  $\Lambda_P$  reduces to  $\Lambda_P = \mathcal{CH}\{\mathbf{0}, P_1\}$ .

#### 3.4.4 Comparison between the Imperfect and Perfect Cases in terms of Queueing Stability

After having characterized the stability region for both the perfect and imperfect cases, we now investigate the maximum *gap* between these two regions. Clearly, here the investigation is reserved for the scenario where *IA provides queueing stability gains* in the perfect case; because otherwise *IA* will also not provide queueing stability gains in the imperfect case, and thus the stability region for the imperfect case will be the same as that for the perfect case, and can be given by  $\mathcal{CH}\{\mathbf{0}, I_1\} = \mathcal{CH}\{\mathbf{0}, P_1\}$ , which is the stability region of SVD.

The gap here can be interpreted as the impact of having limited backhaul, and thus quantization, on the stability region. It is straightforward to notice that the quantization process will result in shrinking the stability region compared with that of the perfect case. To capture this shrinkage, we find that the imperfect case achieves a (guaranteed) fraction of the stability region achieved in the perfect case, which we refer to as *minimum fraction* in the sequel; i.e. the term “minimum fraction” is justified by the fact that the stability region in the imperfect case achieves at least this fraction of the stability region in the perfect case. To be more precise, if we denote this fraction by  $\beta$  ( $\leq 1$ ), then under the imperfect case the queues are stable for any mean arrival rate lying inside  $\beta\Lambda_P$  and it may be possible to achieve stability for some mean arrival rates outside  $\beta\Lambda_P$ . We highlight that this fraction represents the maximum gap between the two regions.

To begin with, we first draw the attention to the fact that, in addition to having  $\mu(L) > r(L)$ , we have  $L_P \geq L_I$ . In order to provide some insights into how we will derive the minimum achievable fraction, in the following we give a simple example for which the stability region shapes are illustrated.

Example: Let  $N = 2$ . Depending on whether *IA* provides queueing stability gains under the imperfect case, two scenarios are to consider. In Figure 3.4 we depict the general shapes of the two (i.e. perfect and imperfect cases) stability regions under these two scenarios; note that these depicted regions are not the result of a specific setting and are only given to help understanding the analysis. From this figure, we can observe that we have different gaps over different directions. To find the minimum fraction (i.e. maximum gap), we adopt the following approach. We take any point from subset  $P_{L_P}$ , and then we try to see how far is this point from the convex hull  $\Lambda_I$  in the direction

### 3.4. Stability Analysis for the Symmetric Case

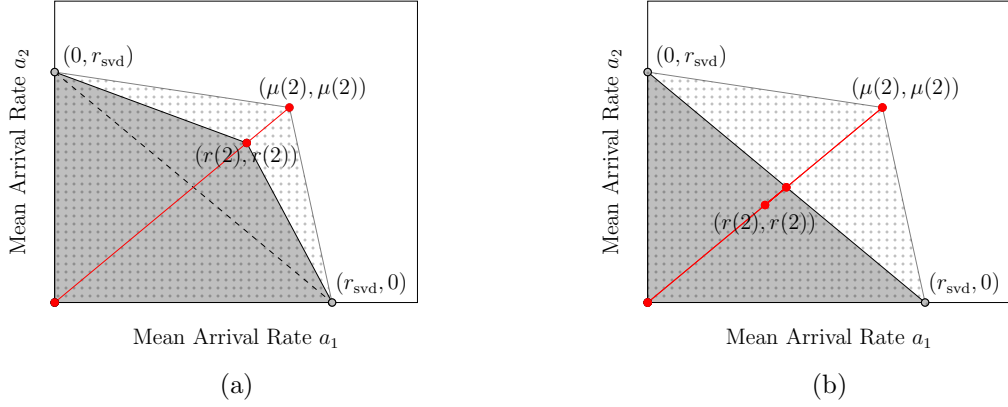


Figure 3.4: Stability regions of the perfect (dotted region) and imperfect (gray region) cases for the symmetric system, where  $N = 2$ . (a) IA provides a queueing stability gain, and (b) IA does not provide a queueing stability gain.

toward the origin. We adopt such an approach mainly for three reasons:

1.  $I_{L_I}$  is a subset of the vertices that characterize the convex hull of the imperfect case ( $\Lambda_I$ ).
2.  $P_{L_P}$  is the subset that contains points (vertices) on the convex hull of the perfect case.
3. The points in  $P_{L_P}$  are the farthest from  $\Lambda_I$ .

Using the above approach, we can state the following theorem that characterizes the minimum achievable fraction.

**Theorem 7.** *For the symmetric system, in general the stability region in the imperfect case achieves at least a fraction  $\frac{r_T(L_I)}{\mu_T(L_P)} = \frac{L_I r(L_I)}{L_P \mu(L_P)} (< 1)$  of the stability region achieved in the perfect case. In other words, the region  $\Lambda_I$  can be bounded as*

$$\frac{r_T(L_I)}{\mu_T(L_P)} \Lambda_P \subseteq \Lambda_I \subseteq \Lambda_P. \quad (3.51)$$

*In the special case where IA delivers no gain under the imperfect case, the above fraction becomes  $\frac{r_{\text{svd}}}{\mu_T(L_P)} = \frac{r_{\text{svd}}}{L_P \mu(L_P)}$ .*

*Proof.* We provide the following result that will help us in the proof of this theorem.

**Lemma 5.** *Each point in  $S_{L+1}$  can be written as  $\frac{L+1}{L+1-n} \times$  some point on the convex hull of  $S_{L+1-n}$ , for  $1 \leq n \leq L$ .*

Please refer to Appendix 3.9.8 for the complete proof of Theorem 7, which includes the proof of Lemma 5.  $\square$

It should be noted that the fraction given in the above theorem represents the maximum impact of limited backhaul on the stability region.

### 3.4.5 Impact of the Number of Bits $B$ on the Stability Region

Here the analysis is restricted for the imperfect case of IA, where the backhaul is of limited capacity. We recall that under the adopted system the number of bits used for the quantization of each channel matrix is  $B$ . Since the stability analysis depends essentially on this number, it is important to study the impact of changing this parameter on the stability performance of the system. But before conducting such a study, we note that an increasing from  $B'$  to  $B$  can be seen as a decreasing from  $B$  to  $B'$ . That is to say, it suffices to study one of these two ways of changing the number of bits. We next investigate the impact of reducing  $B$  to  $B'$ . In this investigation, we consider the scenario where IA always yields a queueing stability gain, as this is the most likely scenario.

To begin with, let  $\Delta_B^*$  and  $\Delta_{B'}^*$  denote the same algorithm, that is the Max-Weight policy, but the first one considers the case where the number of bits is equal to  $B$  and for the second one this number is  $B'$  (with  $B' < B$ ). Further, let  $\mathcal{L}_B$  and  $\mathcal{L}_{B'}$  denote the subsets of pairs selected by  $\Delta_B^*$  and  $\Delta_{B'}^*$ , respectively. Also, we denote by  $\Lambda_B$  and  $\Lambda_{B'}$  the stability regions achieved by  $\Delta_B^*$  and  $\Delta_{B'}^*$ , respectively. In addition, we define  $r(L, B)$  as the average rate  $r(L)$  with a number of bits  $B$ . Equivalently,  $r(L, B')$  is the average rate function  $r(L)$  in which we replace  $B$  by  $B'$ . Using the above definitions, the following theorem can be stated.

**Theorem 8.** *For the same (symmetric) system, where the maximum number of pairs is  $N$ , if we decrease the number of bits from  $B$  to  $B'$ , the stability region  $\Lambda_{B'}$  can be bounded as*

$$\frac{r(N, B')}{r(N, B)} \Lambda_B \subseteq \Lambda_{B'} \subseteq \Lambda_B. \quad (3.52)$$

*Proof.* Please refer to Appendix 3.9.9 for the proof. □

The above result implies that the stability region of the system with  $B'$  bits is at least a fraction  $\frac{r(N, B')}{r(N, B)}$  from the stability region of the system with  $B$  bits.

## 3.5 Algorithmic Design and Performance Analysis for the General Case

We now consider a more general model where, unlike the symmetric case, the path loss coefficients are not necessarily equal to each other. However, for the sake of simplicity, and without loss of generality, we keep the same assumption on the number of streams, that is, all the pairs have equal number of data streams, namely  $d$ . Also, as in the symmetric case, we suppose at each time-slot that rate  $R$  can be supported if the corresponding SINR is greater than or equal to  $\tau$ ; otherwise, the packets are not received correctly and the instantaneous bit-rate is considered to be equal to 0. We recall that  $\mathcal{L}$  stands for the subset of active pairs, with  $|\mathcal{L}| = L$ . We also recall that  $\alpha_{ki} = \frac{P\zeta_{ki}}{d}$  and  $\alpha_{kk} = \frac{P\zeta_{kk}}{d}$ .

### 3.5.1 Average Rate Expressions

#### 3.5.1.1 If the Number of Scheduled Pairs $L \geq 2$

The IA technique is applied in this case, and thus the SINR of stream  $m$  at active user  $k$  can be expressed as the following (see (3.8))

$$\gamma_k^{(m)} = \begin{cases} \frac{\alpha_{kk} \left| \left( \hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)} \right|^2}{\sigma^2 + \sum_{i \in \mathcal{L}, i \neq k} \alpha_{ki} \|\mathbf{h}_{ki}\|^2 e_{ki} \sum_{j=1}^d \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2}, & \text{imperfect case} \\ \frac{\alpha_{kk} \left| \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(m)} \right|^2}{\sigma^2}, & \text{perfect case} \end{cases} \quad (3.53)$$

Based on the above equation, we next derive the average transmission rate expression for each of the perfect and imperfect cases.

**a) Imperfect Case:** Using Theorem 4 and the fact that  $\frac{\alpha_{ki}}{\alpha_{kk}} = \frac{\zeta_{ki}}{\zeta_{kk}}$ , the average transmission rate of user  $k$  ( $\in \mathcal{L}$ ) can be written as

$$r_k = (1 - L\theta) d R e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} \left( \zeta_{ki} \tau d (\zeta_{kk} 2^{\frac{B}{Q}})^{-1} + 1 \right)^{-Q} {}_2F_1(c_2, Q; c_1 + c_2; (\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki} \tau d)^{-1} + 1)^{-1}), \quad (3.54)$$

where we recall that  $c_1 = (Q + 1)Q^{-1}d - Q^{-1}$  and  $c_2 = (Q - 1)c_1$ .

**b) Perfect Case:** Based on Theorem 4, the average transmission rate of active user  $k$  can be expressed as follows

$$\mu_k = (1 - L\theta) d R e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}}. \quad (3.55)$$

#### 3.5.1.2 If the Number of Scheduled Pairs $L = 1$

The SVD technique is applied in this case, and the SINR of stream  $m$  of the active user, which is denoted by index  $k$ , can be given by (see (3.27))

$$\gamma_k^{(m)} = \frac{\zeta_{kk} P}{d \sigma^2} \lambda_k^{(m)}. \quad (3.56)$$

Using (3.31), the average rate for the active user (i.e. user  $k$ ) is

$$r_{\text{svd},k} = (1 - \theta) d R \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \Gamma(j + m_1 - m_2 + 1, \frac{d \sigma^2 \tau}{\zeta_{kk} P}). \quad (3.57)$$

### 3.5.2 Stability Regions and Scheduling Policies

We here provide a characterization of the system stability region for each of the perfect and imperfect cases. Further, we present scheduling policies that can achieve these regions.

#### 3.5.2.1 Imperfect case

Let  $\mathbf{r}$  be a vector that contains at position  $k$  the average rate of user  $k$  if pair  $k$  is active and 0 otherwise; for ex, let  $N = 2$ : if only pair 1 is active then  $\mathbf{r} = (r_{\text{svd},1}, 0)$ , whereas if both of pairs 1 and 2 are active, then  $\mathbf{r} = (r_1, r_2)$ . As mentioned earlier (see Section 3.2, for instance),  $\mathbf{s}$  and  $\mathcal{L}$  are two different representations for the (same) set of active pairs, so we will use  $\mathbf{r}(\mathbf{s})$  to represent the fact that  $\mathbf{r}$  results from decision vector  $\mathbf{s}$ . Notice that, in contrast to the symmetric case, here the average rate expression depends on the identity of the active pairs; consider the same example as before: if  $\mathbf{s} = (1, 1)$ , meaning that pairs 1 and 2 are scheduled for transmission, then  $\mathbf{r}(\mathbf{s}) = (r_1, r_2)$  where  $r_1$  is different from  $r_2$  (in general).

This lack of symmetry will make us incapable of finding the exact set of vertices of the corresponding stability region. However, we can still provide a characterization of this stability region, denoted  $\Lambda_{\text{GI}}$ , by considering all the possible decisions of scheduling the pairs, as follows

$$\Lambda_{\text{GI}} = \mathcal{CH} \{ \mathbf{0}, GI_1, GI_2, \dots, GI_N \}, \quad (3.58)$$

where  $GI_L = \{\mathbf{r}(\mathbf{s}) : \mathbf{s} \in S_L\}$ , i.e.  $GI_L$  is the set of average rate vectors when the number of active pairs is  $L$ . To achieve this stability region we can apply the Max-Weight rule, which is an optimal scheduling policy, denoted by  $\Delta_{\text{GI}}^*$ , such as

$$\Delta_{\text{GI}}^* : \mathbf{s}(t) = \arg \max_{\mathbf{s} \in \mathcal{S}} \{ \mathbf{r}(\mathbf{s}) \cdot \mathbf{q}(t) \}. \quad (3.59)$$

where we recall that  $\mathcal{S}$  is the set of all possible decision vectors  $\mathbf{s}$ .

#### 3.5.2.2 Perfect Case

For this case, we denote by  $\boldsymbol{\mu}$  the average rate vector that contains at position  $k$  the average rate of pair  $k$  if this pair is scheduled and 0 otherwise. Also, let  $\boldsymbol{\mu}(\mathbf{s})$  be the rate vector under decision vector  $\mathbf{s}$ ; as a simple example, let  $N = 2$ : if  $\mathbf{s} = (0, 1)$ , then  $\boldsymbol{\mu}(\mathbf{s}) = (0, r_{\text{svd},2})$ , whereas if  $\mathbf{s} = (1, 1)$ , then  $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1, \mu_2)$ . Let  $\Lambda_{\text{GP}}$  be the system stability region under this case. This region can be represented as the following

$$\Lambda_{\text{GP}} = \mathcal{CH} \{ \mathbf{0}, GP_1, GP_2, \dots, GP_N \}, \quad (3.60)$$

where  $GP_L = \{\boldsymbol{\mu}(\mathbf{s}) : \mathbf{s} \in S_L\}$ . The (optimal) policy that schedules the pairs and achieves this above region is denoted by  $\Delta_{\text{GP}}^*$  and can be given by

$$\Delta_{\text{GP}}^* : \mathbf{s}(t) = \arg \max_{\mathbf{s} \in \mathcal{S}} \{ \boldsymbol{\mu}(\mathbf{s}) \cdot \mathbf{q}(t) \}. \quad (3.61)$$

### 3.5.3 $\beta_A$ -Approximate Scheduling Policy

As detailed earlier, the classical implementation of the Max-Weight policy, in both the perfect and imperfect cases, has a computational complexity of  $\mathcal{O}(N2^N)$ . Whereas for the symmetric case some structural properties allowed us to find a low computational complexity implementation of this policy, here (in the general case) no such properties exist.

To deal with this problem, we propose an alternative policy that has a reduced computational complexity, so that we can apply it instead of the optimal policy, and that achieves an important fraction of the system stability region. Specifically, here we are interested in finding this alternative policy under the imperfect case, which can be considered as the hardest case to analyze since the average rate expressions are very complicated compared with those under the perfect case. The alternative policy in this case is denoted by  $\Delta^A$  and termed as  $\beta_A$ -approximate policy, where this latter expression is justified by the fact that this policy approximates  $\Delta_{GI}^*$  to a fraction of  $\beta_A$  (with  $\beta_A \leq 1$ ). More specifically, for every queue length vector  $\mathbf{q}$ , the following holds ([99])

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq \beta_A (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)}, \quad (3.62)$$

where we recall that  $\mathbf{q}$  is the queue lengths vector, and where  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)}$  (resp.,  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)}$ ) implies that  $\mathbf{r}$  results from the scheduling decisions of policy  $\Delta^A$  (resp.,  $\Delta_{GI}^*$ ).

For the rest of the chapter, for notational conciseness, we will use the term “approximate policy” instead of “ $\beta_A$ -approximate policy” unless stated otherwise. A key step in the investigation is to determine a specific approximation of the average rate expression  $r_k$ , more specifically an approximation that possesses a set of structural features that let us define the approximate policy. Indeed, we will derive such an approximation and prove that it is very accurate if the fractions  $\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki} \tau d)^{-1}$  (or equivalently,  $\alpha_{kk} 2^{\frac{B}{Q}} (\alpha_{ki} \tau d)^{-1}$ ) are sufficiently high (i.e.  $\geq 10$ ), for  $i \neq k$ . For fixed  $\tau$  and  $d$ , these conditions correspond to a scenario where the number of quantization bits is high and/or the cross channels have small path loss coefficients in comparison with the direct channels (i.e. low interference scenario). It should be noted that here the approximation is only provided for the average rates  $r_k$ , meaning that rate  $r_{\text{svd},k}$  is not approximated; recall that if  $L \geq 2$ ,  $r_k$  stands for the average rate for active user  $k$ , whereas if  $L = 1$ , i.e. one active pair,  $r_{\text{svd},k}$  is the corresponding average rate.

Under the aforementioned assumptions/conditions, the approximation of average rate  $r_k$  is given by the following result.

**Rate Approximation 1.** *Under the general system and the imperfect case, and given a subset of active pairs,  $\mathcal{L}$ , with cardinality  $L \geq 2$ , if we have a relatively low interference scenario, the average rate of active user  $k$  ( $\in \mathcal{L}$ ) can be accurately approximated as*

$$r_k \approx (1 - L\theta) d R e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}), \quad (3.63)$$

where  $g_{ki} = \left( \alpha_{kk} 2^{\frac{B}{Q}} (\alpha_{ki} \tau d)^{-1} + 1 \right)^{-1} = \left( \zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki} \tau d)^{-1} + 1 \right)^{-1}$ .

*Proof.* The derivation is provided in Appendix 3.9.10. □



### 3.5. Algorithmic Design and Performance Analysis for the General Case

---

It can be noted that the low-interference condition, i.e.  $\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki} \tau d)^{-1}$  is sufficiently high, implies that all the  $g_{ki}$ ,  $\forall i \neq k$ , are sufficiently small. To proceed further with the analysis, let  $\bar{g}_k$  be the average value of all the  $g_{ki}$ ,  $\forall i \neq k$ , for the same number of active pairs  $L$ . More specifically, for a fixed cardinality  $L$  we take all the possible subsets (i.e. scheduling decisions) in which user  $k$  is active. For each of these subsets, there are  $L - 1$  values of  $g_{ki}$ . Hence,  $\bar{g}_k$  is the average of these  $g_{ki}$  values over all the considered decisions. Using the average value  $\bar{g}_k$  and the approximate expression of  $r_k$  (given in (3.63)), we define function  $\phi_k(L)$  as the following

$$\phi_k(L) = \begin{cases} (1 - L\theta) dR e^{-\frac{\sigma^2 \tau}{\alpha_{kk}} (1 - \bar{g}_k)^{L-1}}, & \text{if } L \geq 2 \\ r_{\text{svd},k}, & \text{if } L = 1 \end{cases} \quad (3.64)$$

where the expression of  $r_{\text{svd},k}$  is given in (3.57).

Let  $\phi$  be the vector containing  $\phi_k(L)$  at position  $k$  if pair  $k$  is scheduled (with  $L - 1$  other pairs); otherwise, we put 0 at this position. Under this setting, we define the approximate policy  $\Delta^A$  as

$$\Delta^A : \mathbf{s}(t) = \arg \max_{\mathbf{s} \in \mathcal{S}} \{ \phi(\mathbf{s}) \cdot \mathbf{q}(t) \},$$

where  $\phi(\mathbf{s})$  results from decision vector  $\mathbf{s}$ ; as a simple example, set  $N = 2$ : if  $\mathbf{s} = (1, 0)$ , then  $\phi(\mathbf{s}) = (r_{\text{svd},1}, 0)$ , whereas if  $\mathbf{s} = (1, 1)$ , then  $\phi(\mathbf{s}) = (\phi_1(2), \phi_2(2))$ . It is noteworthy to mention that, for  $L \geq 2$ , although we use  $\phi_k(L)$  to make the scheduling decision under  $\Delta^A$ , the *actual* average rate of user  $k$  is still  $r_k$ . Also, remark that  $\Delta^A$  follows the Max-Weight rule, thus, as was shown earlier, implementing  $\Delta^A$  as a classical maximization problem over all the possible decisions  $\mathbf{s}$  needs a CC of  $\mathcal{O}(N2^N)$ . However, in contrast to  $\Delta_{\text{GI}}^*$ , policy  $\Delta^A$  has a structural property that will allow us to propose an equivalent reduced CC implementation instead of the classical one. This property results from the fact that  $\phi_k(L)$  is independent of the  $L - 1$  other active users, and only depends on pair  $k$  and the cardinality  $L$ . The proposed implementation of policy  $\Delta^A$  is given by Algorithm 2.

---

**Algorithm 2 :** A Reduced Computational Complexity Implementation of  $\Delta^A$

---

- 1: Initialize  $L_g = 0$  and  $ws_{L_g} = 0$ .
  - 2: **for**  $l = 1 : 1 : N$  **do**
  - 3:   Sort the pairs in a descending order with respect to the product  $pro_k = \phi_k(l) q_k$ ; break ties arbitrarily.
  - 4:   Let  $ws_l =$  sum of the first  $l$  biggest  $pro_k$  values.
  - 5:   Save  $sq_l$  as the subset of  $l$  pairs that yields  $ws_l$ .
  - 6:   **if**  $ws_l > ws_{L_g}$  **then**
  - 7:     Put  $ws_{L_g} = ws_l$  and  $L_g = l$ .
  - 8:   **end if**
  - 9: **end for**
  - 10: Schedule the pairs given by  $sq_{L_g}$ .
- 

*Computational Complexity of the proposed implementation of  $\Delta^A$ :* To compare with the classical implementation, we now focus on the computational complexity of the proposed implementation, which depends essentially on a “for loop” of  $N$  iterations, each of which contains: (i) a “bubble sorting algorithm”, which needs  $\mathcal{O}(N^2)$ , (ii) a sum of  $l$  terms in iteration  $l$ , and (iii) other steps of small CC

### 3.5. Algorithmic Design and Performance Analysis for the General Case

compared with those mentioned before. Thus, by neglecting the CC of the steps in (iii) and noticing that the summing steps (in (ii)) over all the iterations need  $\mathcal{O}(\frac{N(N+1)}{2}) = \mathcal{O}(N^2)$ , the CC of the proposed implementation is roughly  $\mathcal{O}(N^2N + N^2) = \mathcal{O}(N^3)$ , which is very small compared with  $\mathcal{O}(N2^N)$  for large  $N$ .

In general, an approximate policy comes with the disadvantage of reducing the achievable stability region compared with the optimal policy. Indeed, we will show that policy  $\Delta^A$  guarantees to achieve only a fraction of the stability region achieved by policy  $\Delta_{\text{GI}}^*$ . Let us now provide some remarks and definitions that will be useful for the analysis. We recall that  $g_{ki} = \left(\zeta_{kk}2^{\frac{B}{Q}}(\zeta_{ki}\tau d)^{-1} + 1\right)^{-1}$ . In addition, we define  $\mathcal{L}_A$  as the subset of pairs chosen by  $\Delta^A$ , and we let the cardinality of this subset be  $L_A$ . As for  $\Delta_{\text{GI}}^*$  we keep the original notation, i.e.  $\mathcal{L}$  is the scheduled subset, with  $L = |\mathcal{L}|$ . Recall that  $\mathfrak{L}$  stands for the set of all possible decision subsets, so  $\mathcal{L}_A$  and  $\mathcal{L}$  are elements of  $\mathfrak{L}$ . In the following we provide a result that is essential for the characterization of the fraction that  $\Delta^A$  can achieve.

**Rate Approximation 2.** *If the values of  $g_{ki}$ ,  $\forall i \neq k$ , are relatively close to  $\bar{g}_k$ , the approximation of  $r_k$  in (3.63) can in its turn be accurately approximated as*

$$\begin{aligned} r_k &\approx (1 - L\theta) dR e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}) \\ &\approx (1 - L\theta) dR e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \left[ (1 - \bar{g}_k)^{L-1} - (1 - \bar{g}_k)^{L-2} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right]. \end{aligned} \quad (3.65)$$

*Proof.* The derivation is provided in Appendix 3.9.11. □

Let us define  $\Lambda_{\text{GI}}$  as the stability region of the general system under the imperfect case, which can be achieved by  $\Delta_{\text{GI}}^*$ . We also define  $\Lambda_A$  to be the stability region that the approximate policy  $\Delta^A$  can achieve. Concerning the fraction that policy  $\Delta^A$  can achieve w.r.t. the stability region  $\Lambda_{\text{GI}}$ , and under all the above-mentioned conditions, we have the following important result.

**Theorem 9.** *The approximate policy  $\Delta^A$  achieves at least a fraction  $\beta_A$  ( $\leq 1$ ) of the stability region achieved by the optimal policy  $\Delta_{\text{GI}}^*$ , meaning that  $\Lambda_A$  can be bounded as*

$$\beta_A \Lambda_{\text{GI}} \subseteq \Lambda_A \subseteq \Lambda_{\text{GI}}, \quad (3.66)$$

where  $\beta_A$  is given as follows

$$\beta_A = \frac{1 + \min_{\mathcal{L}_A \in \mathfrak{L}} \left\{ \min_{k \in \mathcal{L}_A} \left\{ -(1 - \bar{g}_k)^{-1} \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right\} \right\}}{1 + \max_{\mathcal{L} \in \mathfrak{L}} \left\{ \max_{k \in \mathcal{L}} \left\{ -(1 - \bar{g}_k)^{-1} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right\} \right\}}. \quad (3.67)$$

*Proof.* Please refer to Appendix 3.9.12 for the proof. □

We point out that in the above theorem we suppose, without loss of generality, that the min term in the numerator takes values in the interval  $] -1, 0[$ , whereas the max term in the denominator is greater than 0. More details about these assumptions can be found in the proof of the theorem.

### 3.5.4 Compare the Imperfect and Perfect Cases in terms of Queueing Stability

At the very beginning of this section, we showed that policy  $\Delta_{\text{GP}}^*$  achieves the (general) system stability region under the perfect case. Let us denote by  $\mathcal{L}_{\text{P}}$  the subset of scheduled pairs using  $\Delta_{\text{GP}}^*$  and by  $L_{\text{P}}$  the cardinality of this subset. Also, we define  $\Lambda_{\text{GP}}$  to be the stability region of the perfect case. On the other hand, for the imperfect case we adopt the same notation as before, i.e. the subset of scheduled users and its cardinality are represented by  $\mathcal{L}$  and  $L$ , respectively, and the stability region is denoted by  $\Lambda_{\text{GI}}$ . Under our system, an essential parameter to investigate is the fraction the stability region the imperfect case achieves compared with the stability region of the perfect case. This fraction is captured in the following theorem.

**Theorem 10.** *The stability region of the imperfect case reaches at least a fraction  $\beta_{\text{P}}$  of the stability region achieved in the perfect case, meaning that  $\Lambda_{\text{GI}}$  can be bounded as*

$$\beta_{\text{P}} \Lambda_{\text{GP}} \subseteq \Lambda_{\text{GI}} \subseteq \Lambda_{\text{GP}}, \quad (3.68)$$

in which  $\beta_{\text{P}}$  is given as

$$\beta_{\text{P}} = \min_{\mathcal{L}_{\text{P}} \in \mathfrak{L}} \left\{ \min_{k \in \mathcal{L}_{\text{P}}} \left\{ \prod_{i \in \mathcal{L}_{\text{P}}, i \neq k} (1 - g_{ki}) \right\} \right\}, \quad (3.69)$$

where  $g_{ki} = \left( \zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki} \tau d)^{-1} + 1 \right)^{-1}$ .

*Proof.* Please refer to Appendix 3.9.13 for the proof.  $\square$

We draw the attention to the fact that the proof of the above result relies on the (first) approximation of  $r_k$  given in Rate Approximation 1, meaning that the above theorem holds under the low-interference condition (i.e. all the  $g_{ki}$ , with  $i \neq k$ , are sufficiently small).

#### 3.5.4.1 Relation between $\beta_{\text{P}}$ and $B$

An important factor on which fraction  $\beta_{\text{P}}$  depends is the number of quantization bits  $B$ , so it is essential to compute the number of bits that can guarantee this fraction. Finding the explicit relation that gives the number of bits in function of  $\beta_{\text{P}}$  is a difficult task, however we can obtain the required result numerically. In detail, using the expression of  $\beta_{\text{P}}$  given in the above theorem, we start from a small value of  $B$  for which we calculate the corresponding fraction, then we keep increasing  $B$  until the desired value of  $\beta_{\text{P}}$  is obtained.

### 3.6. Stability Analysis with Multiple Rate Levels

---

Although computing a closed form relation of  $B$  in function of  $\beta_P$  is hard to achieve, we can still find a relation that gives a rough idea of the required number of bits. Specifically, we know that  $1 - g_{ki} = (1 + 2^{-\frac{B}{Q}} c_{ki})^{-1}$ , where  $c_{ki} = \zeta_{ki} \tau d \zeta_{kk}^{-1}$ , then, after selecting the subset  $\mathcal{L}_P$  (of cardinality  $L_P$ ) and  $k \in \mathcal{L}_P$  that yield  $\beta_P$ , we find a parameter  $c$  such that

$$c = \min_{i \in \mathcal{L}_P, i \neq k} c_{ki}. \quad (3.70)$$

We thus obtain the following relation

$$\beta_P \leq \left(1 + 2^{-\frac{B}{Q}} c\right)^{-(L_P-1)}, \quad (3.71)$$

or equivalently we get

$$B \geq Q \log_2 \left( \frac{c}{\beta_P^{-(L_P-1)} - 1} \right). \quad (3.72)$$

Therefore, it suffices to use a number of quantization bits equals to the lower bound in the above inequality to guarantee the fraction  $\beta_P$ . Note that the exact number of bits, given by the numerical method, is less than the calculated lower bound.

## 3.6 Stability Analysis with Multiple Rate Levels

In this section, we consider the case of multiple modulations (instead of one), meaning that we have multiple achievable rate levels. Under this case, we provide a stability analysis for the symmetric system, while noting that such an analysis under the general system is not provided since it is of considerable complexity. We first give the average rate expressions and we study their behaviors. After that, we provide the stability analysis for the IA technique. Finally, after presenting some results for SVD technique, we investigate the conditions under which IA yields a stability gain.

We consider a setting where the set of possible rates is given by  $\{R_1, R_2, \dots, R_M\}$ , in which rate  $R_j$  is per slot and is assigned for stream  $m$  of active user  $k$  if the corresponding SINR (i.e.  $\gamma_k^{(m)}$ ) is higher than or equal to threshold  $\tau_j$ . Without loss of generality, we assume that  $\tau_1 < \tau_2 < \dots < \tau_M$ , which implies  $R_1 < R_2 < \dots < R_M$ ; note that all these rates are finite, i.e.  $R_M < \infty$ . Further, under the symmetry assumption, all the pairs have the same number of streams, namely  $d$ , and the path loss coefficients are all equal to the same value, which is denoted by  $\zeta$ . Under the above setting and the fact that the set of active pairs is represented by  $\mathcal{L}$  of cardinality  $L$ , the *average rate* expression for active user  $k$  can be written in function of the transmission success probability conditioned on the subset of active pairs as

$$(1 - L\theta) \sum_{m=1}^d \left[ R_M \mathbb{P} \left\{ \gamma_k^{(m)} \geq \tau_M \mid \mathcal{L}(t) \right\} + \sum_{j=1}^{M-1} R_j \mathbb{P} \left\{ \tau_j \leq \gamma_k^{(m)} < \tau_{j+1} \mid \mathcal{L}(t) \right\} \right]. \quad (3.73)$$

### 3.6. Stability Analysis with Multiple Rate Levels

Using the following equality

$$\mathbb{P}\left\{\tau_j \leq \gamma_k^{(m)} < \tau_{j+1} \mid \mathcal{L}(t)\right\} = \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_j \mid \mathcal{L}(t)\right\} - \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_{j+1} \mid \mathcal{L}(t)\right\}, \quad (3.74)$$

we can rewrite the expression in (3.73) as

$$(1 - L\theta) \sum_{m=1}^d \left[ R_M \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_M \mid \mathcal{L}\right\} + \sum_{j=1}^{M-1} R_j \left( \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_j \mid \mathcal{L}(t)\right\} - \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_{j+1} \mid \mathcal{L}\right\} \right) \right]. \quad (3.75)$$

Rearranging its terms, the above equation can be re-expressed as

$$(1 - L\theta) \sum_{m=1}^d \left[ R_1 \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_1 \mid \mathcal{L}\right\} + \sum_{j=1}^{M-1} (R_{j+1} - R_j) \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_{j+1} \mid \mathcal{L}\right\} \right]. \quad (3.76)$$

We point out that here the SINR expressions for the cases  $L \geq 2$  and  $L = 1$  are exactly the same as for the single rate level scenario and are provided in (3.35) and (3.43), respectively.

We next provide a more explicit expression of the average rate functions for both the perfect and imperfect cases, and we study the behavior (w.r.t.  $L$ ) of each of these functions.

#### 3.6.1 Average Rate Expressions and their Variation

##### 3.6.1.1 If the Number of Scheduled Pairs $L \geq 2$

**a) Imperfect Case:** Recalling that  $\mathbb{P}\left\{\gamma_k^{(m)} \geq \tau \mid \mathcal{L}\right\} = e^{-\frac{\sigma^2 \tau}{\alpha}} (F(\tau))^{L-1}$  (see the single rate level scenario) and denoting the average rate for active user  $k \in \mathcal{L}$  as  $r_{kM}$  in this case, we get

$$r_{kM} = (1 - L\theta)d \left[ R_1 e^{-\frac{\sigma^2 \tau_1}{\alpha}} (F(\tau_1))^{L-1} + \sum_{j=1}^{M-1} (R_{j+1} - R_j) e^{-\frac{\sigma^2 \tau_{j+1}}{\alpha}} (F(\tau_{j+1}))^{L-1} \right], \quad (3.77)$$

where  $F(\tau_j)$  stands for function  $F$ , for which the expression can be found in the single rate case, calculated by replacing  $\tau$  with  $\tau_j$ . We can observe that this rate function depends only on the cardinality  $L$  of the subset of active pairs and not on the identity of these pairs, hence all the active users have the same average rate; we define this latter one as  $r_M(L) = r_{kM}$ , for all  $k \in \mathcal{L}$ . Consequently, the total average rate, which we denote by  $r_{MT}(L)$ , is given by

$$r_{MT}(L) = L(1 - L\theta)d \left[ R_1 e^{-\frac{\sigma^2 \tau_1}{\alpha}} (F(\tau_1))^{L-1} + \sum_{j=1}^{M-1} (R_{j+1} - R_j) e^{-\frac{\sigma^2 \tau_{j+1}}{\alpha}} (F(\tau_{j+1}))^{L-1} \right]. \quad (3.78)$$

For the variation of functions  $r_M(L)$  and  $r_{MT}(L)$ , we state the following lemma.

**Lemma 6.** *The average rate function  $r_M(L)$  decreases w.r.t. the number of pairs  $L$ . On the other hand, the total average rate function  $r_{MT}(L)$  has the following behavior over the interval  $[0, \frac{1}{\theta}]$ : For  $L = 0$  or  $L = \frac{1}{\theta}$ ,  $r_{MT}(L) = 0$  and between 0 and  $\frac{1}{\theta}$ , the function is positive and might have several*

### 3.6. Stability Analysis with Multiple Rate Levels

*maxima, one of which is a global maximum.*

*Proof.* Please refer to Appendix 3.9.14 for the proof.  $\square$

One final thing to mention here is that providing an explicit expression for the number at which  $r_{\text{MT}}(L)$  reaches its *global* maximum is a difficult task, however we can easily find this point numerically. In general, the resulting value is not an integer, so we can adopt the procedure used for the single rate level case to determine the best and nearest integer to this value. For the remainder of the manuscript, we denote this integer by  $L_{\text{MT}}$ .

**b) Perfect Case:** For this case, we know that  $\mathbb{P}\left\{\gamma_k^{(m)} \geq \tau \mid \mathcal{L}\right\} = e^{-\frac{\sigma^2 \tau}{\alpha}}$ , which is independent of the set of active pairs. In a similar way as for the imperfect case, we can notice that the average rate will be the same for all active users. Letting this rate and the total one in this case be  $\mu_{\text{M}}(L)$  and  $\mu_{\text{MT}}(L)$ , respectively, we can write

$$\mu_{\text{M}}(L) = (1 - L\theta)d \left[ R_1 e^{-\frac{\sigma^2 \tau_1}{\alpha}} + \sum_{j=1}^{M-1} (R_{j+1} - R_j) e^{-\frac{\sigma^2 \tau_{j+1}}{\alpha}} \right], \quad (3.79)$$

$$\mu_{\text{MT}}(L) = L(1 - L\theta)d \left[ R_1 e^{-\frac{\sigma^2 \tau_1}{\alpha}} + \sum_{j=1}^{M-1} (R_{j+1} - R_j) e^{-\frac{\sigma^2 \tau_{j+1}}{\alpha}} \right]. \quad (3.80)$$

It is obvious that  $\mu_{\text{M}}(L)$  is a decreasing function whereas  $\mu_{\text{MT}}(L)$  is concave at  $\frac{1}{2\theta}$ . We define  $L_{\text{MP}}$  as the best and nearest integer to  $\frac{1}{2\theta}$ , which can be computed in a similar way as for the single rate case.

#### 3.6.1.2 If the Number of Scheduled Pairs $L = 1$

The SVD technique is applied in this case, and the SINR is the same as for the case with single rate level. Let us define  $MF(\tau)$  as

$$MF(\tau) = \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \Gamma(j + m_1 - m_2 + 1, \frac{d\sigma^2 \tau}{\zeta P}). \quad (3.81)$$

Hence, the average rate in this case, which we denoted by  $r_{\text{Msvd}}$ , can be expressed as

$$\begin{aligned} r_{\text{Msvd}} &= (1 - \theta)d \left[ R_1 \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_1\right\} + \sum_{j=1}^{M-1} (R_{j+1} - R_j) \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau_{j+1}\right\} \right] \\ &= (1 - \theta)d \left[ R_1 MF(\tau_1) + \sum_{j=1}^{M-1} (R_{j+1} - R_j) MF(\tau_{j+1}) \right]. \end{aligned} \quad (3.82)$$

It can be seen that the above average rate is independent of whether we are under the perfect case or imperfect case since the backhaul is not used when there is only one active pair. We also note that the average rate here is independent of the identity of the scheduled pair.

## 3.6.2 Stability Regions and Scheduling Policies

### 3.6.2.1 Imperfect Case

We define the subset of average rate vectors if  $L \geq 2$  pairs are active as  $MI_L = \{r_M(L)\mathbf{s} : \mathbf{s} \in S_L\}$ . If  $L = 1$ , we have  $MI_1 = \{r_{\text{Msvd}}\mathbf{s} : \mathbf{s} \in S_1\}$ . Similarly to the condition on  $L_I$  for the single rate level scenario, we assume without loss of generality that  $2 \leq L_{\text{MI}} \leq N$ . Based on the above, we can state the following result on the characterization of the stability region.

**Theorem 11.** *For the symmetric system with multiple rate levels, the stability region in the imperfect case can be characterized as*

$$\Lambda_{\text{MI}} = \mathcal{CH}\{\mathbf{0}, MI_1, MI_2, \dots, MI_{L_{\text{MI}}}\}. \quad (3.83)$$

*Proof.* The proof of this theorem can be done in the same way as the proof of Theorem 5, we thus omit it to avoid repetition.  $\square$

In order to select the active pairs at each time-slot, we use the Max-Weight rule, which we denote by  $\Delta_{\text{MI}}^*$ . Specifically, this rule selects the decision vector  $\mathbf{s}(t)$  that yields the following max

$$\Delta_{\text{MI}}^* : \max \left\{ \max_{\mathbf{s} \in S_L, 2 \leq L \leq N} \{r_M(\|\mathbf{s}\|_1) \mathbf{s} \cdot \mathbf{q}(t)\}, \max_{\mathbf{s} \in S_1} \{r_{\text{Msvd}} \mathbf{s} \cdot \mathbf{q}(t)\} \right\}. \quad (3.84)$$

It is noteworthy to mention that the above policy is throughput optimal, meaning that it can achieve the entire stability region  $\Lambda_{\text{MI}}$ . We also point out that the classical implementation of this policy has the same computational complexity as that of  $\Delta_I^*$ , namely  $\mathcal{O}(N2^N)$ . To reduce this complexity to  $\mathcal{O}(N^2)$ , we can use the equivalent implementation proposed in Algorithm 1 but after replacing  $r(L)$  and  $r_{\text{svd}}$  with  $r_M(L)$  and  $r_{\text{Msvd}}$ , respectively.

### 3.6.2.2 Perfect Case

We define the subset of average rate vectors if  $L \geq 2$  pairs are active as  $MP_L = \{\mu_M(L)\mathbf{s} : \mathbf{s} \in S_L\}$ . For  $L = 1$ , we define  $MP_1 = \{r_{\text{Msvd}}\mathbf{s} : \mathbf{s} \in S_1\}$ . We suppose that  $L_{\text{MP}} \leq N$ . Hence, the stability region under the perfect case is given by the following theorem.

**Theorem 12.** *For the symmetric system with multiple modulations, the stability region in the perfect case can be represented as*

$$\Lambda_{\text{MP}} = \mathcal{CH}\{\mathbf{0}, MP_1, MP_2, \dots, MP_{L_{\text{MP}}}\}. \quad (3.85)$$

*Proof.* The proof of the theorem is along the lines of the other proofs in this chapter and is thus omitted to avoid redundancy.  $\square$

### 3.6. Stability Analysis with Multiple Rate Levels

To decide which pairs to schedule at each time-slot, we use the following optimal scheduling policy, denoted by  $\Delta_{\text{MP}}^*$ , which selects the decision vector  $\mathbf{s}(t)$  that yields the following max

$$\Delta_{\text{MP}}^* : \max \left\{ \max_{\mathbf{s} \in S_L, 2 \leq L \leq N} \{ \mu_{\text{M}}(\|\mathbf{s}\|_1) \mathbf{s} \cdot \mathbf{q}(t) \}, \max_{\mathbf{s} \in S_1} \{ r_{\text{Msvd}} \mathbf{s} \cdot \mathbf{q}(t) \} \right\}. \quad (3.86)$$

Note that the same remark about the computational complexity of  $\Delta_{\text{MI}}^*$  can be made for  $\Delta_{\text{MP}}^*$ , that is, the classical implementation of this latter policy results in a complexity of  $\mathcal{O}(N2^N)$ , which can be reduced using Algorithm 1 but after considering  $\mu_{\text{M}}(L)$  and  $r_{\text{Msvd}}$  instead of  $r(L)$  and  $r_{\text{svd}}$ , respectively.

#### 3.6.3 Conditions under which IA Provides a Gain in terms of Queueing Stability

In this subsection, and based on a similar investigation as that done in the single rate level case, we provide the conditions under which the use of IA can yield a queueing stability gain. These conditions are given for the perfect and imperfect cases as follows.

**Proposition 5.** *For the symmetric system with multiple rate levels, under the imperfect case IA yields a queueing stability gain if and only if there exists a number  $L$  such that  $Lr_{\text{M}}(L) > r_{\text{Msvd}}$ , with  $2 \leq L \leq L_{\text{MI}}$ . For the same system but under the perfect case, we get a similar result, that is, IA yields a queueing stability gain iff there exists a number  $L$  such that  $L\mu_{\text{M}}(L) > r_{\text{Msvd}}$ , with  $2 \leq L \leq L_{\text{MP}}$ .*

*Proof.* The proof can be done in the same way as the proof of Proposition 4 and is thus omitted to avoid repetition.  $\square$

Based on similar observations as in the single rate level case, the following remark can be made.

**Remark 2.** *For the imperfect case with multiple rate levels, if we can find an  $L$  such that  $Lr_{\text{M}}(L) > r_{\text{Msvd}}$ , with  $2 \leq L \leq L_{\text{MI}}$ , a more precise characterization of  $\Lambda_{\text{MI}}$  is as follows*

$$\Lambda_{\text{MI}} = \mathcal{CH} \{ \mathbf{0}, MI_1, MI_L, \dots, MI_{L_{\text{MI}}} \}. \quad (3.87)$$

*If such an  $L$  does not exist, then the region characterization becomes  $\Lambda_{\text{MI}} = \mathcal{CH} \{ \mathbf{0}, MI_1 \}$ .*

*For the perfect case with multiple rate levels, if there exists an  $L$  such that  $L\mu_{\text{M}}(L) > r_{\text{Msvd}}$  is satisfied, with  $2 \leq L \leq L_{\text{MP}}$ , we obtain*

$$\Lambda_{\text{MP}} = \mathcal{CH} \{ \mathbf{0}, MP_1, MP_L, \dots, MP_{L_{\text{MP}}} \}. \quad (3.88)$$

*Otherwise, the characterization of  $\Lambda_{\text{MP}}$  reduces to  $\Lambda_{\text{MP}} = \mathcal{CH} \{ \mathbf{0}, MP_1 \}$ .*



### 3.6.4 Compare the Imperfect and Perfect Cases in terms of Queueing Stability

We now want to examine the gap between the stability regions resulting from the perfect and imperfect cases. To this end, we compute the (minimum) fraction that the stability region in the imperfect case achieves from the stability region in the perfect case. Similarly to the approach used in the single rate scenario, this fraction corresponds to that between a point in  $P_{L_{MP}}$  and its corresponding point (i.e. in the same direction toward the origin) on the convex hull of the imperfect case (given by  $\Lambda_{MI}$ ). Under this approach, the following result holds.

**Theorem 13.** *For the symmetric system with multiple rate levels, the stability region in the imperfect case achieves at least a fraction  $\frac{r_{MT}(L_{MI})}{\mu_{MT}(L_{MP})} = \frac{L_{MI} r_M(L_{MI})}{L_{MP} \mu_M(L_{MP})}$  of the stability region achieved in the perfect case. In other words, the region  $\Lambda_{MI}$  can be bounded as the following*

$$\frac{r_{MT}(L_{MI})}{\mu_{MT}(L_{MP})} \Lambda_{MP} \subseteq \Lambda_{MI} \subseteq \Lambda_{MP}. \quad (3.89)$$

*In the special case where IA yields no gain under the imperfect case, the above fraction reduces to*

$$\frac{r_{Msvd}}{\mu_{MT}(L_{MP})} = \frac{r_{Msvd}}{L_{MP} \mu(L_{MP})}. \quad (3.90)$$

*Proof.* The proof of this theorem is essentially the same as the proof of Theorem 7, we thus omit it to avoid repetition.  $\square$

## 3.7 Validation of the Proposed Model

In this section we present the numerical results. We consider a system where the number of antennas  $N_t = N_r = 7$ ,  $P = 10$ ,  $\sigma = 1$ ,  $d = 2$ ,  $\theta = 0.01$ . We take  $N = 6$ , which satisfies the condition  $N_t + N_r \geq (N + 1)d$ . In addition, we assume that all the users have Poisson incoming traffic with the same average arrival rates as  $a_k = a$ . Also, we assume a scheme with a rate of  $\log_2(1 + \tau)$  bits per channel use if the SINR of a scheduled user exceeds  $\tau$ . We set the slot duration to be  $T_s = 1000$  channel uses. Even though in practice all the path loss coefficients are different, we consider in this section a very special case that simplifies the simulations and can still provide insights on the comparison between IA and time-division multiple access (TDMA)-SVD. In detail, we assume that all the direct links have a path loss coefficient of 1 and all the cross links have a path loss coefficient of  $\zeta_c$  (with  $\zeta_c \leq 1$ ). This setting allows us to examine, with respect to parameter  $\zeta_c$ , the impact of the cross channels (or equivalently, the impact of interference) on the system stability performance and it let us detect when IA technique can provide a queueing stability gain.

To show the stability performance of the system, we plot the *total average queue length* given by

$$\frac{1}{M_s} \sum_{t=0}^{M_s-1} \sum_{k=1}^N q_k(t) \quad (3.91)$$

### 3.7. Validation of the Proposed Model

Table 3.1: List of simulation parameters.

Parameter	Description	Value
$N$	Maximum number of pairs	6
$N_t$	Number of antennas at each transmitter	7
$N_r$	Number of antennas at each receiver	7
$d_k$	Number of data streams for pair $k$	2
$P$	Total power at each transmitter	10
$\sigma^2$	Noise variance	1
$\theta$	Fraction of slot duration to probe one user	0.01
$T_s$	Slot duration (in number of channel uses)	$10^3$
$M_s$	Number of slots per simulation	$10^4$

for different values of  $a$ , where each simulation lasts  $M_s$  time-slots. We set  $M_s = 10^4$ . We point out that the point where the total average queue length function increases very *steeply* is the point at which the system becomes unstable.

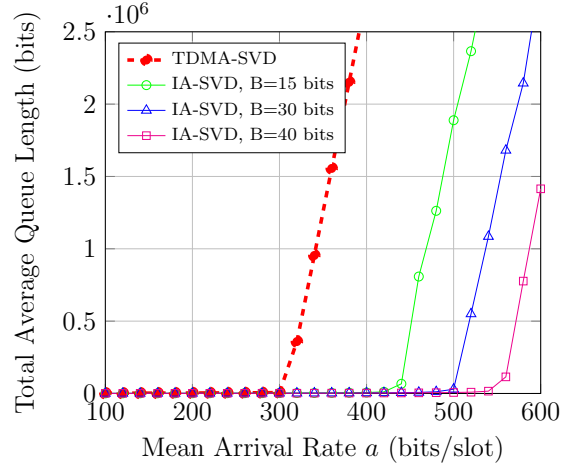


Figure 3.5: Total average queue length vs. mean arrival rate  $a$ . Here,  $\zeta_c = 0.2$  and  $\tau = 1$ .

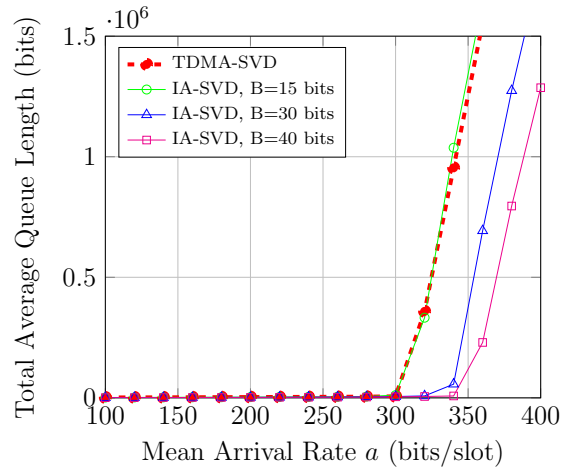


Figure 3.6: Total average queue length vs. mean arrival rate  $a$ . Here,  $\zeta_c = 0.4$  and  $\tau = 1$ .

### 3.7. Validation of the Proposed Model

Figures 3.5 and 3.6 depicts the variations of the total average queue length with respect to the mean arrival rate  $a$  for two systems: (i) our adopted system, i.e. IA and SVD are used, where we consider several values of the number of quantization bits  $B$ , and (ii) a TDMA-SVD system, i.e. there is always only one active pair and where SVD is applied. In Figure 3.5, where a relatively low-interference scenario ( $\zeta_c = 0.2$ ) is considered, we can see that IA can provide a queueing stability gain and this gain increases with the number of quantization bits. This is due to the fact that the more the quantization is precise, the more we achieve higher rates which implies better stability performance. On the other side, from Figure 3.6, where a relatively high-interference scenario ( $\zeta_c = 0.4$ ) is considered, we can notice that for small  $B$  (e.g.  $B = 15$  bits) IA does not provide any additional stability gain to that of SVD. However, when we increase the number of bits (e.g.  $B = 30$  and 40 bits), IA becomes capable of yielding a stability gain. This results from the fact that in high interference scenarios, IA needs better CSI knowledge in order to maintain a good alignment of interference, and this can be provided by using a sufficiently large number of bits in the quantization process. It is worth noting that, although we considered the impact of  $B$  and  $\zeta_c$ , there exist other parameters that may affect the system performance, such as the number of antennas, the threshold  $\tau$ , the number of data streams, etc. Furthermore, we point out that for low values of  $a$  the average queue lengths appear to be 0 in Figure 3.5 and 3.6, however these length values are not 0 but just very small compared to the maximum length value (of order  $10^5$ ).

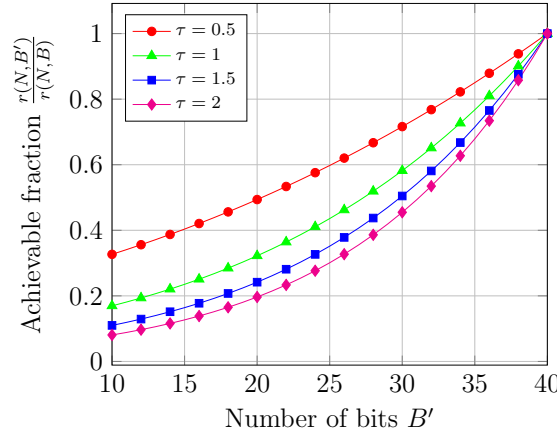


Figure 3.7: Achievable fraction  $\frac{r(N,B')}{r(N,B)}$  vs. number of bits  $B'$ . Here,  $\zeta_c = 1$  and  $B = 40$ .

Figure 3.7 depicts the variation of the fraction  $\frac{r(N,B')}{r(N,B)}$  with the number of bits  $B'$ , for different values of  $\tau$  and for a fixed reference number of bits  $B = 40$  bits; here, we set  $\zeta_c = 1$  since  $\frac{r(N,B')}{r(N,B)}$  is defined for the symmetric system. It is clear from this figure that increasing the number of quantization bits or decreasing the threshold  $\tau$  result in higher achievable fractions. Also, one can notice that changing the number of quantization bits has a higher impact on the achievable fraction for greater values of  $\tau$ , meaning that the more the threshold is high, the more the fraction  $\frac{r(N,B')}{r(N,B)}$  is sensitive to the variation of the number of bits.

In Figure 3.8 we illustrate the variation of fraction  $\beta_P$  with the number of bits  $B$ , for different values of  $\tau$  and  $\zeta_c$ . The plots in this figure confirm the expectation that the stability region in the imperfect case gets bigger, i.e. the fraction this stability region achieves with respect to the stability

### 3.8. Closing Remarks

---

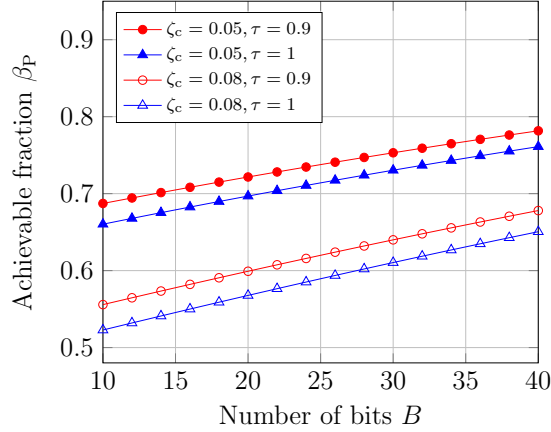


Figure 3.8: Achievable fraction  $\beta_P$  vs. number of bits  $B$ .

region in the perfect case is bigger, for greater  $B$  and/or lower  $\zeta_c$ .

## 3.8 Closing Remarks

In this chapter, we have considered a MIMO interference network under TDD mode with limited backhaul capacity and taking into account the probing cost, and where we adopt a centralized scheduling scheme to select the active pairs in each time-slot. For the case where only one pair is active we apply the SVD technique, whereas if this number is greater than or equal to two we apply the IA technique. Under this setting, we have characterized the stability region and proposed a scheduling policy to achieve this region for the perfect case (i.e. unlimited backhaul) as well as for the imperfect case. Then, we have captured the maximum gap between these two resulting stability regions. These stability regions, scheduling rules and maximum gaps are provided for the symmetric system (i.e. equal path loss coefficients) as well as for the general system. In addition, for the symmetric system we have provided the conditions under which IA can deliver a queueing stability gain compared to SVD. Moreover, under the general system, since the scheduling policy is of a high computational complexity, we have proposed an approximate policy that has a reduced complexity but that achieves only a fraction of the system stability region. Also, a characterization of this achievable fraction is provided. Finally, in the same vein as the symmetric case with single rate level, we have presented a stability analysis for the symmetric system under a multiple rate levels scheme.

## 3.9 Appendix

### 3.9.1 Proof of Proposition 1

It was shown in [90, Appendix A] that both  $\left| \left( \hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)} \right|^2$  and  $\left| \left( \mathbf{u}_k^{(m)} \right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(m)} \right|^2$  have an exponential distribution with parameter 1, thus the proof for the perfect case follows directly. However, for the imperfect case, the proof is not straightforward and needs some investigations.

By defining  $Num = \left| \left( \hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)} \right|^2$ , we can write

$$\begin{aligned} \mathbb{P} \left\{ \gamma_k^{(m)} \geq \tau \mid \mathcal{L}(t) \right\} &= \mathbb{P} \left\{ \frac{Num}{RI_k^{(m)} + \sigma^2} \geq \frac{\tau}{\alpha_{kk}} \right\} \\ &= \mathbb{P} \left\{ Num \geq \frac{RI_k^{(m)} \tau}{\alpha_{kk}} + \frac{\sigma^2 \tau}{\alpha_{kk}} \right\} \\ &= \int_0^\infty CCDF_{Num} \left( \frac{RI_k^{(m)} \tau}{\alpha_{kk}} + \frac{\sigma^2 \tau}{\alpha_{kk}} \right) PDF(RI_k^{(m)}) dRI_k^{(m)} \\ &= \int_0^\infty e^{-\frac{RI_k^{(m)} \tau}{\alpha_{kk}} - \frac{\sigma^2 \tau}{\alpha_{kk}}} PDF(RI_k^{(m)}) dRI_k^{(m)}, \end{aligned} \quad (3.92)$$

where the last equality holds since  $Num$  is exponentially distributed with parameter 1 and thus its complementary cumulative distribution function can be given by  $CCDF_{Num}(x) = e^{-x}$ . Note that  $PDF(RI_k^{(m)})$  is the probability distribution function (PDF) of  $RI_k^{(m)}$ . Hence, we get

$$\begin{aligned} \mathbb{P} \left\{ \gamma_k^{(m)} \geq \tau \mid \mathcal{L}(t) \right\} &= \int_0^\infty e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} e^{-\frac{RI_k^{(m)} \tau}{\alpha_{kk}}} PDF(RI_k^{(m)}) dRI_k^{(m)} \\ &= e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} MGF_{RI_k^{(m)}} \left( -\frac{\tau}{\alpha_{kk}} \right), \end{aligned} \quad (3.93)$$

in which  $MGF_{RI_k^{(m)}}(\cdot)$  is the MGF of  $RI_k^{(m)}$ . This concludes the proof.

### 3.9.2 Proof of Theorem 4

For the perfect case, the statement follows directly from Proposition 1. Using this proposition, it can be seen that we need to calculate  $MGF_{RI_k^{(m)}} \left( -\frac{\tau}{\alpha_{kk}} \right)$  in order to prove the statement for the imperfect case. To this end, we first recall that, using (3.19), we can write

$$RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \|\mathbf{h}_{ki}\|^2 e_{ki} \sum_{j=1}^{d_i} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2. \quad (3.94)$$

Since  $\mathbf{w}_{ki}$  and  $\mathbf{T}_{k,i}^{(m,j)}$  are independent and identically distributed (i.i.d.) isotropic vectors in the null space of  $\hat{\mathbf{h}}_{ki}$ ,  $\left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2$  is i.i.d. Beta(1,  $Q - 1$ ) distributed for all  $i$ , where  $Q = N_t N_r - 1$ . Hence,

### 3.9. Appendix

$\sum_{j=1}^{d_i} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2$  is the sum of  $d_i$  i.i.d. Beta variables, which can be approximated to another Beta distribution [106]. Specifically, we have

$$\sum_{j=1}^{d_i} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 \sim d_i \text{Beta}(c_{1i}, c_{2i}), \quad (3.95)$$

where  $c_{1i} = \frac{(Q+1)d_i}{Q} - \frac{1}{Q}$  and  $c_{2i} = (Q-1)c_{1i}$ .

According to [107],  $e_{ki} \|\mathbf{h}_{ki}\|^2$  is  $\text{Gamma}(Q, 2^{\frac{B}{Q}})$  distributed, where  $Q$  and  $2^{\frac{B}{Q}}$  are the shape and rate parameters, respectively. Let us define  $\delta$  such as  $\delta = 2^{\frac{B}{Q}}$ . It follows that

$$RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \rho_{ki} X_i Y_i, \quad (3.96)$$

with  $\rho_{ki} = \alpha_{ki} d_i$ ,  $X_i \sim \text{Gamma}(Q, \delta)$  and  $Y_i \sim \text{Beta}(c_{1i}, c_{2i})$ . It is clear that  $X_i Y_i$  is nothing but the product of a Gamma and Beta random variables, thus the PDF of  $P_i = X_i Y_i$  can be given as follows (refer to [108] for more details)

$$f_{P_i}(p_i) = \frac{\delta^Q \Gamma(c_{2i})}{\Gamma(Q) \text{Bet}(c_{1i}, c_{2i})} p_i^{Q-1} e^{-\delta p_i} \Psi(c_{2i}, 1 + Q - c_{1i}; \delta p_i), \quad (3.97)$$

where  $\Psi$  is the Kummer function defined as

$$\Psi(a, b; x) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt} t^{a-1} (1+t)^{b-a-1} dt, \quad (3.98)$$

and where  $\Gamma(\cdot)$  is the Gamma function and  $\text{Bet}(\cdot, \cdot)$  is the Beta function.

Therefore, the MGF of random variable  $P_i$  can be written as

$$\begin{aligned} MGF_{P_i}(-t) &= \int_{-\infty}^{+\infty} e^{-tp_i} f_{P_i}(p_i) dp_i \\ &= \kappa \int_0^{+\infty} p_i^{Q-1} e^{-tp_i - \delta p_i} \Psi(c_{2i}, 1 + Q - c_{1i}; \delta p_i) dp_i \\ &\stackrel{(a)}{=} \kappa \frac{\Gamma(Q) \Gamma(c_{1i})}{\delta^Q \Gamma(c_{1i} + c_{2i})} \left( \frac{t}{\delta} + 1 \right)^{-Q} {}_2F_1(c_{2i}, Q; c_{1i} + c_{2i}; \frac{1}{1 + \frac{\delta}{t}}) \\ &\stackrel{(b)}{=} \left( \frac{t}{\delta} + 1 \right)^{-Q} {}_2F_1(c_{2i}, Q; c_{1i} + c_{2i}; \frac{1}{1 + \frac{\delta}{t}}), \end{aligned} \quad (3.99)$$

where  $\kappa$  is defined as follows

$$\kappa = \frac{\delta^Q \Gamma(c_{2i})}{\Gamma(Q) B(c_{1i}, c_{2i})}. \quad (3.100)$$

### 3.9. Appendix

Equality (a) is obtained using [109], whereas equality (b) holds due to the definition of the Beta function

$$\text{Bet}(c_1, c_2) = \frac{\Gamma(c_1)\Gamma(c_2)}{\Gamma(c_1 + c_2)}. \quad (3.101)$$

It is clear that we can write  $RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \rho_{ki} P_i$ , which is the sum of weighed (independent) random variables ( $P_i$ ) with  $\rho_{ki}$  as weights. The MGF of  $RI_k^{(m)}$  at  $-t$  is then given by

$$\begin{aligned} MGF_{RI_k^{(m)}}(-t) &= \prod_{i \in \mathcal{L}(t), i \neq k} MGF_{P_i}(-t\rho_{ki}) \\ &= \prod_{i \in \mathcal{L}(t), i \neq k} \left( \frac{\alpha_{ki} d_i t}{\delta} + 1 \right)^{-Q} {}_2F_1(c_{2i}, Q; c_{1i} + c_{2i}; \frac{1}{\frac{\delta}{\alpha_{ki} d_i t} + 1}). \end{aligned} \quad (3.102)$$

This results from the fact that the moment-generating function of a sum of independent random variables is the product of the moment-generating functions of these variables. Hence, by taking  $t = -\frac{\tau}{\alpha_{kk}}$  and recalling that  $\delta = 2^{\frac{B}{Q}}$ , we eventually get

$$MGF_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right) = \prod_{i \in \mathcal{L}(t), i \neq k} \left( \frac{\alpha_{ki} \tau d_i}{\alpha_{kk} 2^{\frac{B}{Q}}} + 1 \right)^{-Q} {}_2F_1(c_{2i}, Q; c_{1i} + c_{2i}; \frac{1}{\frac{\alpha_{kk} 2^{\frac{B}{Q}}}{\alpha_{ki} \tau d_i} + 1}). \quad (3.103)$$

Notice that the above MGF expression is independent of the identity of the data stream, so we have

$$\sum_{m=1}^{d_k} MGF_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right) = d_k MGF_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right). \quad (3.104)$$

Hence, the desired result follows.

#### 3.9.3 Proof of Proposition 2

To begin with, we define two variables  $w_l$  and  $\Omega_n$  as the following

$$\omega_l = (-1)^l (n + m_1 - m_2)! ((n - l)! (m_1 - m_2 + l)! l!)^{-1}, \quad (3.105)$$

$$\Omega_n = n! (m_2(n + m_1 - m_2)!)^{-1}. \quad (3.106)$$

Based on these definitions, we can write

$$L_n^{m_1 - m_2}(\lambda) = \sum_{l=0}^n \omega_l \lambda_m^l, \quad (3.107)$$

### 3.9. Appendix

---

$$p(\lambda_m) = \sum_{n=0}^{m_2-1} \Omega_n [L_n^{m_1-m_2}(\lambda_m)]^2 \lambda_m^{m_1-m_2} e^{-\lambda_m}. \quad (3.108)$$

For the Laguerre polynomial, we have

$$[L_n^{m_1-m_2}(\lambda_m)]^2 = \sum_{j=0}^{2n} \kappa_j \lambda_m^j, \quad (3.109)$$

where  $\kappa_j = \sum_{i=0}^j \omega_i \omega_{j-i}$ , with  $\omega_s = 0$  if  $s > n$ . Since  $\gamma_k^{(m)} = \frac{\zeta P}{d\sigma^2} \lambda_m$ , the probability that the corresponding SINR exceeds a certain threshold  $\tau$  can be written as

$$\begin{aligned} \mathbb{P}\left\{\gamma_k^{(m)} \geq \tau\right\} &= \mathbb{P}\left\{\lambda_m \geq \frac{d\sigma^2\tau}{\zeta P}\right\} \\ &= \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \int_{\frac{d\sigma^2\tau}{\zeta P}}^{\infty} \lambda_m^{j+m_1-m_2} e^{-\lambda_m} d\lambda_m \\ &= \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \Gamma(j+m_1-m_2+1, \frac{d\sigma^2\tau}{\zeta P}), \end{aligned} \quad (3.110)$$

where  $\Gamma(\cdot, \cdot)$  stands for the upper incomplete Gamma function. Hence, the desired result follows.

#### 3.9.4 Proof of Lemma 2

We start the proof by first showing that  $r(L)$  decreases with  $L$ . The first derivative of this rate function is given by

$$\frac{dr}{dL} = dRe^{-\frac{\sigma^2\tau}{\alpha}} (-\theta + (1 - L\theta) \log F) F^{L-1}. \quad (3.111)$$

Since we have  $L < \frac{1}{\theta}$  and  $\log F < 0$ , the first derivative is negative and so  $r$  decreases with  $L$ .

To study the variation of  $r_T(L)$  (w.r.t.  $L$ ) we need to first compute its first derivative, which will help us determine the optimal number of pairs. The first derivative can be written as

$$\frac{dr_T}{dL} = dRe^{-\frac{\sigma^2\tau}{\alpha}} (-L^2\theta \log F + L(-2\theta + \log F) + 1) F^{L-1}. \quad (3.112)$$

To study the sign of this derivative w.r.t.  $L$ , we first calculate its zeros and investigate if they are feasible (i.e. they satisfy  $L\theta < 1$ ). Setting  $\frac{dr_T}{dL} = 0$  yields

$$-L^2\theta \log F + L(-2\theta + \log F) + 1 = 0, \quad (3.113)$$

or equivalently

$$L^2 + L\left(\frac{2}{\log F} - \frac{1}{\theta}\right) - \frac{1}{\theta \log F} = 0. \quad (3.114)$$



### 3.9. Appendix

We can easily show that the only zeros of  $\frac{dr_T}{dL}$  are at

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F}}}{2}, \quad (3.115)$$

$$L_1 = \frac{\frac{1}{\theta} - \frac{2}{\log F} + \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F}}}{2}. \quad (3.116)$$

Note that  $\log F < 0$  and  $\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F} = \frac{1}{\theta^2} + \frac{4}{(\log F)^2}$ . Let us now examine the feasibility of  $L_0$  and  $L_1$ . Indeed, under our setting a number  $L$  is feasible if it satisfies  $0 < L < \frac{1}{\theta}$ , since  $L\theta$  should be  $< 1$ . For  $L_0$  we have

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}}{2} < \frac{\frac{1}{\theta} - \frac{2}{\log F} - \frac{2}{|\log F|}}{2} = \frac{1}{2\theta}, \quad (3.117)$$

where the inequality results from the fact that  $\frac{2}{|\log F|} < \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}$ . We can also observe that

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}}{2} > \frac{\frac{1}{\theta} - \frac{2}{\log F} - \frac{1}{\theta} - \frac{2}{|\log F|}}{2} = 0. \quad (3.118)$$

Thus,  $L_0$  is a feasible solution since  $0 < L_0 < \frac{1}{\theta}$ . On the other hand, for  $L_1$  we can notice that

$$L_1 = \frac{\frac{1}{\theta} - \frac{2}{\log F} + \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}}{2} > \frac{\frac{1}{\theta} + \sqrt{\frac{1}{\theta^2}}}{2} = \frac{1}{\theta}. \quad (3.119)$$

Hence,  $L_1$  is not a feasible solution because  $L_1 > \frac{1}{\theta}$ . To complete the proof it suffices to show that  $r_T(L)$  reaches its maximum at  $L_0$ . To this end, we note that  $r_T(0) = 0$ ,  $r_T(\frac{1}{\theta}) = 0$  and  $\frac{dr_T}{dL}|_{L=\frac{1}{2\theta}} < 0$ , and we recall that  $0 < L_0 < \frac{1}{2\theta} < \frac{1}{\theta}$ . In addition, one can easily notice that  $r_T$  and its first derivative ( $\frac{dr_T}{dL}$ ) are continuous over  $[0, \frac{1}{\theta}]$ . Based on these observations, the variation of  $r_T$  over  $[0, \frac{1}{\theta}]$  can be described as follows:  $r_T$  is increasing from 0 to  $L_0$  and decreasing from  $L_0$  to  $\frac{1}{\theta}$ .

#### 3.9.5 Proof of Lemma 3

We first provide the proof of Lemma 4 that will help us in the proof of Lemma 3.

##### • Step 1 (Proof of Lemma 4):

*Proof.* We start the proof by first defining  $\mathcal{E}_{i,L}$  as the set containing the points (vectors) that only have  $L$  '1' (the other coordinate values are '0') and where the positions (indexes) of these '1' are the same as those of  $L$  '1' coordinates of  $\mathbf{s}_{i,L+1}$ . Note that the points in  $\mathcal{E}_{i,L}$  are all different from each other. The cardinality of  $\mathcal{E}_{i,L}$ , which is denoted by  $|\mathcal{E}_{i,L}|$ , is nothing but the result of the combination

### 3.9. Appendix

of  $L + 1$  elements taken  $L$  at a time without repetition, and it can be computed as the following

$$|\mathcal{E}_{i,L}| = \binom{L+1}{L} = \frac{(L+1)!}{L!(L+1-L)!} = L+1. \quad (3.120)$$

Thus, we have  $L + 1$  elements from  $S_L$  that if we take them in a specific convex combination, we get a point on the same line (from the origin) as that of  $\mathbf{s}_{i,L+1}$ . This can be represented by

$$\sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L} \equiv \mathbf{s}_{i,L+1}, \quad (3.121)$$

where  $\equiv$  is a notation used to represent the fact that these two points are on the same line from the origin, and where  $\sum_{j \in \mathcal{E}_{i,L}} \delta_j = 1$  and  $\delta_j \geq 0$ . Let us suppose that all the coefficients  $\delta_j = \frac{1}{L+1}$ . This assumption satisfies the above constraints, namely  $\sum_{j \in \mathcal{E}_{i,L}} \delta_j = 1$  and  $\delta_j \geq 0$ . By replacing these coefficients in the term at the left-hand-side of (3.121), we get

$$\sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L} = \frac{1}{L+1} \sum_{j \in \mathcal{E}_{i,L}} \mathbf{s}_{j,L} = \frac{L}{L+1} \mathbf{s}_{i,L+1}, \quad (3.122)$$

where the second equality holds since we have  $L + 1$  elements to sum (due to the fact that  $|\mathcal{E}_{i,L}| = L + 1$ ), each of which contains  $L$  '1' at the same positions as  $L$  '1' coordinates of  $\mathbf{s}_{i,L+1}$ , and (these elements) differ from each other in the position of one '1' (and consequently of one '0'); for instance, suppose that  $N = 5$ ,  $L = 2$  and  $\mathbf{s}_{i,L+1} = (1, 1, 1, 0, 0)$ , then the points  $\mathbf{s}_{j,L}$  are given by the subset  $\mathcal{E}_{i,L} = \{(1, 1, 0, 0, 0); (1, 0, 1, 0, 0); (0, 1, 1, 0, 0)\}$ . The sum corresponding to each coordinate is then equal to  $L$ . To complete the proof, it remains to show that  $\frac{1}{L+1} \sum_{j \in \mathcal{E}_{i,L}} \mathbf{s}_{j,L}$  is on the convex hull of  $S_L$ . To this end, note that all the points in  $S_L$  are on the same hyperplane (in  $\mathbb{R}_+^N$ ), which is described by the equation  $\sum_{k=1}^N \nu_k - L = 0$ ;  $\nu_k$  represents the  $k$ -th coordinate. Hence, a point on the convex hull of  $S_L$  is also on this hyperplane. If we compute  $\sum_{i=k}^N \nu_k$  for point  $\frac{1}{L+1} \sum_{j \in \mathcal{E}_{i,L}} \mathbf{s}_{j,L}$ , it yields  $\frac{(L+1)L}{L+1} = L$  due to the definition of  $\mathcal{E}_{i,L}$ , thus this point is on the defined hyperplane and consequently on the convex hull of  $S_L$ .

Example: In order to better understand the result of this lemma, we provide a simple example for which the geometric illustration is in Figure 3.9. In this example, we take  $N = 2$ ,  $S_1 = \{(1, 0); (0, 1)\}$  and  $S_2 = \{(1, 1)\}$ . In addition, we define points  $P_2 = (1, 1)$  and  $P_1 = (\frac{1}{2}, \frac{1}{2})$ . Note that  $P_2 \in S_2$  and  $P_1$  is on the convex hull of  $S_1$ . We can express  $P_2$  as  $P_2 = \frac{2}{1}[\frac{1}{2}(1, 0) + \frac{1}{2}(0, 1)] = 2(\frac{1}{2}, \frac{1}{2}) = \frac{2}{1}P_1$ . Thus,  $P_2$  equals  $\frac{2}{1} \times$  its corresponding point ( $P_1$ ) on the convex hull of  $S_1$ .

This completes the proof of Lemma 4.  $\square$

• **Step 2:** Now, using the above lemma, a point  $\mathbf{s}_{i,L+1}$  in  $S_{L+1}$  can be expressed in function of  $L + 1$  specific points in  $S_L$  as  $\mathbf{s}_{i,L+1} = \frac{L+1}{L} \sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L}$ , which implies that

$$r(L+1)\mathbf{s}_{i,L+1} = r(L+1)\frac{L+1}{L} \sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L}, \quad (3.123)$$

where the definition of  $\mathcal{E}_{i,L}$  can be found in the proof of Lemma 4. By Lemma 2, we have

$$(L+1)r(L+1) < Lr(L), \text{ for } L \geq L_1. \quad (3.124)$$

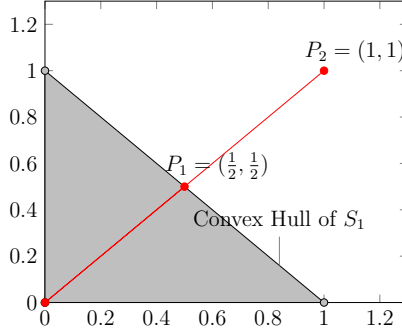


Figure 3.9: Example that illustrates the result of Lemma 4.

We thus get

$$r(L+1) \frac{L+1}{L} \sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L} \prec r(L) \frac{L}{L} \sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L} = r(L) \sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L}. \quad (3.125)$$

Note that the inequality operator  $\prec$  in (3.125) is component-wise. Therefore, each point in  $I_{L+1}$  is in the convex hull of  $I_L$ , for  $L \geq L_1$ , since  $r(L+1)\mathbf{s}_{i,L+1} \in I_{L+1}$  and  $r(L) \sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L}$  is in the convex hull of  $I_L$ . Consequently, all the points in  $I_{L+1}$  for  $L \geq L_1$  (i.e. these points form  $\bar{\mathcal{I}}$ ) are in the convex hull of  $I_{L_1}$ , which is a subset of  $\mathcal{I}$ . Therefore, the desired result holds.

### 3.9.6 Proof of Theorem 5

The proof consists of two steps. First we will prove that the region in the statement of the Theorem is indeed achievable. We then have to prove the converse, that is, if there exists a centralized policy that stabilizes the system for a mean arrival rate vector  $\mathbf{a}$ , then  $\mathbf{a} \in \Lambda_I$ .

- **Step 1:** Using the fact that a queue is stable if the arrival rate is strictly lower than the departure rate, it is sufficient to show that for each point in the stability region there exists a scheduling policy that achieves this point. Indeed, a point  $\mathbf{r}_I$  in  $\Lambda_I$  can be written as the convex combination of the points in  $\mathcal{I}$  as  $\mathbf{r}_I = \sum_{i=1}^{|\mathcal{I}|} p_i \mathbf{r}_i$ , where  $\mathbf{r}_i$  represents a point in  $\mathcal{I}$ ,  $p_i \geq 0$  and  $\sum_{i=1}^{|\mathcal{I}|} p_i = 1$ . Note that each point  $\mathbf{r}_i$  represents a different scheduled subset of pairs and that the probability of choosing point  $\mathbf{0}$  is equal to 0. To achieve  $\mathbf{r}_I$  it suffices to use a randomized policy that at the beginning of each time-slot selects (decision)  $\mathbf{r}_i$  with probability  $p_i$ . Since  $\mathbf{r}_I$  is an arbitrary point in  $\Lambda_I$ , we can claim that this region is achievable.

- **Step 2:** Assume the system is stable for a mean arrival rate vector  $\mathbf{a}$ . As explained earlier, the scheduling decision (i.e. subset  $\mathcal{L}$ ) under the centralized policy depends on the queues only, so we show this dependency by  $\mathcal{L}(\mathbf{q})$ . Let us denote by  $\mathbf{r}_s$  the mean service rate vector. In addition, we denote by  $\mathbf{r}(\mathcal{L}(\mathbf{q}))$  the average rate vector if the queues state is  $\mathbf{q}$  and the selected subset of pairs is  $\mathcal{L}(\mathbf{q})$ . It is obvious that the set of all possible values of  $\mathbf{r}(\mathcal{L}(\mathbf{q}))$  is nothing but  $\mathcal{I} \cup \bar{\mathcal{I}}$ . Under the adopted model, the system can be described as a discrete time Markov chain on a countable state space with a single communicating class (i.e. irreducible) [27]. Since we assume strong stability, then we have

### 3.9. Appendix

that this chain is positive recurrent and the mean service rates greater than the mean arrival rates [27]. In addition, we can deduce that the Markov chain has a unique stationary distribution (because the chain is irreducible and positive recurrent), which we denote by  $\pi(\mathbf{q})$ . Thus, the following holds for the mean service rate vector

$$\mathbf{r}_s = \sum_{\mathbf{q} \in \mathbb{Z}_+^N} \pi(\mathbf{q}) \mathbf{r}(\mathcal{L}(\mathbf{q})) = \sum_{\mathcal{L} \in \mathfrak{L}} \mathbf{r}(\mathcal{L}) \sum_{\mathbf{q} \in \mathbb{Z}_+^N: \mathcal{L}(\mathbf{q}) = \mathcal{L}} \pi(\mathbf{q}) \succ \mathbf{a}, \quad (3.126)$$

where the operator  $\succ$  is component-wise. By setting  $p(\mathcal{L}) = \sum_{\mathbf{q} \in \mathbb{Z}_+^N: \mathcal{L}(\mathbf{q}) = \mathcal{L}} \pi(\mathbf{q})$  and noticing that the set of all possible values of  $\mathbf{r}(\mathcal{L})$  is the same as  $\mathbf{r}(\mathcal{L}(\mathbf{q}))$ , that is,  $\mathcal{I} \cup \bar{\mathcal{I}}$ , the mean service rate can be re-written as the following

$$\mathbf{r}_s = \sum_{j=1}^{|\mathcal{I} \cup \bar{\mathcal{I}}|} p_j \mathbf{r}_j, \quad (3.127)$$

in which  $j$  is used to denote decision  $\mathcal{L}$ , meaning that  $p_j = p(\mathcal{L})$ , and  $\mathbf{r}_j$  represents a point in set  $\mathcal{I} \cup \bar{\mathcal{I}}$ , and where  $|\mathcal{I} \cup \bar{\mathcal{I}}|$  represents the cardinality of this set. Hence, we can state that  $\mathbf{r}_s$  is in the convex hull of  $\mathcal{I} \cup \bar{\mathcal{I}}$ . But, since we have demonstrated that  $\bar{\mathcal{I}}$  is in the convex hull of  $\mathcal{I}$  (see Lemma 3), we have  $\mathbf{r}_s \in \Lambda_I$  and consequently  $\mathbf{a} \in \Lambda_I$ . This completes the proof.

#### 3.9.7 Proof of Proposition 4

The proof consists of two steps. In the first step we prove that if there exists an  $L$  such that  $Lr(L) > r_{\text{svd}}$ , then IA provides a queueing stability gain compared to SVD. In the second step, we show that if for all the  $L$  we have  $Lr(L) \leq r_{\text{svd}}$ , then IA does not provide any gain.

• **Step 1:** Here we assume that there exists an  $L$  such that  $Lr(L) > r_{\text{svd}}$ , with  $2 \leq L \leq L_I$ , where we recall that  $r(L)$  is the average rate (per user) when  $L$  pairs are active. To prove that here IA can provide a queueing stability gain, we show that IA combined with SVD is capable of achieving points that are outside the stability region of SVD. This is proven as follows.

It is well known that any point in a stability region characterized by its vertices can be written as a convex combination of these vertices. Let  $\mathbf{p}_d$  represent any point in the stability region resulting from SVD, thus this point can be expressed as a convex combination of the vertices of this region, and we can write

$$\mathbf{p}_d = \delta_1(r_{\text{svd}}, 0, \dots, 0) + \delta_2(0, r_{\text{svd}}, 0, \dots, 0) + \dots + \delta_N(0, \dots, 0, r_{\text{svd}}) = (\delta_1 r_{\text{svd}}, \dots, \delta_N r_{\text{svd}}), \quad (3.128)$$

where  $\delta_i \geq 0$  and  $\sum_i \delta_i = 1$ . Let us define a specific scheduling policy that at each time-slot schedules a subset of  $l$  pairs and where pair  $i$  is selected with a probability  $\pi_i^{(l)}$ . Here we assume that  $l$  can be 1 or  $L$ . Clearly, for  $l = 1$  we use SVD, whereas for  $l = L$  we use IA. For  $l = 1$  we choose  $\pi_i^{(1)}$  such that  $\pi_i^{(1)} \leq \delta_i$ ,  $\forall i$ , while for  $l = L$  we set  $\pi_i^{(L)} = (\delta_i - \pi_i^{(1)})L$ ; it can be noticed that  $\sum_i \pi_i^{(1)} + L^{-1} \sum_i \pi_i^{(L)} = \sum_i \delta_i = 1$ . Then, the points in the stability region achieved by combining

### 3.9. Appendix

IA and SVD can be written as

$$\begin{aligned}\mathbf{p}_{\text{ad}} &= (\pi_1^{(1)} r_{\text{svd}} + \pi_1^{(L)} r(L), \dots, \pi_N^{(1)} r_{\text{svd}} + \pi_N^{(L)} r(L)) \\ &= (\pi_1^{(1)} r_{\text{svd}} + (\delta_1 - \pi_1^{(1)}) Lr(L), \dots, \pi_N^{(1)} r_{\text{svd}} + (\delta_N - \pi_N^{(1)}) Lr(L)).\end{aligned}\quad (3.129)$$

Since here we have  $Lr(L) > r_{\text{svd}}$ , then it can be deduced that

$$\begin{aligned}\mathbf{p}_{\text{ad}} &= (\pi_1^{(1)} r_{\text{svd}} + (\delta_1 - \pi_1^{(1)}) Lr(L), \dots, \pi_N^{(1)} r_{\text{svd}} + (\delta_N - \pi_N^{(1)}) Lr(L)) \\ &\succ (\pi_1^{(1)} r_{\text{svd}} + (\delta_1 - \pi_1^{(1)}) r_{\text{svd}}, \dots, \pi_N^{(1)} r_{\text{svd}} + (\delta_N - \pi_N^{(1)}) r_{\text{svd}}) = (\delta_1 r_{\text{svd}}, \dots, \delta_N r_{\text{svd}}) = \mathbf{p}_{\text{d}},\end{aligned}\quad (3.130)$$

where the operator  $\succ$  is component-wise. Hence, we can claim that the proposed policy achieves points that are outside the stability region of SVD.

• **Step 2:** In this step, we assume that the condition  $Lr(L) \leq r_{\text{svd}}$  holds true for any  $L$  such that  $2 \leq L \leq L_I$ . To prove that here IA cannot yield a queueing stability gain, we show that SVD is capable of achieving any point in the stability region of IA combined with SVD. This is detailed as follows. Any point in the stability region resulting from combining IA and SVD can be achieved by a scheduling policy that at each time-slot schedules a subset of pairs where pair  $i$  is selected with a probability  $\pi_i^{(L)}$ , such that  $\sum_i \sum_L L^{-1} \pi_i^{(L)} = 1$ . So this point, denoted by  $\mathbf{p}_{\text{ad}}$ , can be written as the following

$$\mathbf{p}_{\text{ad}} = (\pi_1^{(1)} r_{\text{svd}} + \sum_{L:L \geq 2} \pi_1^{(L)} r(L), \dots, \pi_N^{(1)} r_{\text{svd}} + \sum_{L:L \geq 2} \pi_N^{(L)} r(L)).\quad (3.131)$$

Let us define a scheduling policy under SVD (i.e. one pair is active) that at each time-slot selects pair  $i$  with a probability  $\delta_i = \sum_L L^{-1} \pi_i^{(L)}$ , where it is clear that  $\sum_i \delta_i = \sum_i \sum_L L^{-1} \pi_i^{(L)} = 1$ . Hence, this policy can achieve a point, denoted by  $\mathbf{p}_{\text{d}}$ , such as

$$\mathbf{p}_{\text{d}} = (\delta_1 r_{\text{svd}}, \dots, \delta_N r_{\text{svd}}) = (\sum_L L^{-1} \pi_1^{(L)} r_{\text{svd}}, \dots, \sum_L L^{-1} \pi_N^{(L)} r_{\text{svd}}).\quad (3.132)$$

The condition  $Lr(L) \leq r_{\text{svd}}$  yields

$$\begin{aligned}\mathbf{p}_{\text{d}} &= (\sum_L L^{-1} \pi_1^{(L)} r_{\text{svd}}, \dots, \sum_L L^{-1} \pi_N^{(L)} r_{\text{svd}}) \\ &= (\pi_1^{(1)} r_{\text{svd}} + \sum_{L:L \geq 2} L^{-1} \pi_1^{(L)} r_{\text{svd}}, \dots, \pi_N^{(1)} r_{\text{svd}} + \sum_{L:L \geq 2} L^{-1} \pi_N^{(L)} r_{\text{svd}}) \\ &\succeq (\pi_1^{(1)} r_{\text{svd}} + \sum_{L:L \geq 2} L^{-1} \pi_1^{(L)} Lr(L), \dots, \pi_N^{(1)} r_{\text{svd}} + \sum_{L:L \geq 2} L^{-1} \pi_N^{(L)} Lr(L)) = \mathbf{p}_{\text{ad}},\end{aligned}\quad (3.133)$$

where the equality is achieved for some judicious choice of the  $\pi_i^{(L)}$ . This shows that the policy using SVD can achieve any point in the stability region of IA combined with SVD.

A similar proof can be done for the perfect case by replacing  $r(L)$  by  $\mu(L)$ , we thus omit this proof to avoid repetition. Therefore, the desired result holds.

### 3.9.8 Proof of Theorem 7

We first provide the proof of Lemma 5 that will help us in the proof of the theorem.

• **Step 1 (Proof of Lemma 5):**

*Proof.* From Lemma 4, a point in  $S_{L+1}$  can be expressed in function of some subset of points, represented by  $\mathcal{E}_{i,L}$ , in  $S_L$  as  $\mathbf{s}_{i,L+1} = \frac{L+1}{L} \sum_{i_1 \in \mathcal{E}_{i,L}} \delta_{i_1,L} \mathbf{s}_{i_1,L}$ . More specifically, we found that the coefficients  $\delta_{i_1,L}$  are all equal to  $\frac{1}{L+1}$ , and thus  $\mathbf{s}_{i,L+1} = \frac{L+1}{L} \sum_{i_1 \in \mathcal{E}_{i,L}} \frac{1}{L+1} \mathbf{s}_{i_1,L} = \frac{1}{L} \sum_{i_1 \in \mathcal{E}_{i,L}} \mathbf{s}_{i_1,L}$ . Similarly, each point  $\mathbf{s}_{i_1,L}$  ( $\in S_L$ ) can be written in function of some specific subset of points, denoted  $\mathcal{E}_{i_1,L-1}$ , in  $S_{L-1}$  as  $\mathbf{s}_{i_1,L} = \frac{1}{L-1} \sum_{i_2 \in \mathcal{E}_{i_1,L-1}} \mathbf{s}_{i_2,L-1}$ . Following this reasoning until index  $L - (n - 1)$ , we can express  $\mathbf{s}_{i,L+1}$  as

$$\mathbf{s}_{i,L+1} = \frac{1}{L(L-1) \dots (L-(n-1))} \sum_{i_1 \in \mathcal{E}_{i,L}} \sum_{i_2 \in \mathcal{E}_{i_1,L-1}} \dots \sum_{i_n \in \mathcal{E}_{i_{n-1},L-(n-1)}} \mathbf{s}_{i_n,L-(n-1)}. \quad (3.134)$$

We denote by  $\mathcal{E}_{n-1,L-(n-1)}$  the subset containing the points (vectors) that only have  $L - n$  '1' (the other coordinate values are '0') and where the positions (indexes) of these '1' coordinates are the same as the positions of  $L - n$  '1' coordinates of  $\mathbf{s}_{i,L+1}$ . It is important to point out that the elements in  $\mathcal{E}_{n-1,L-(n-1)}$  are all different from each other. It can be easily noticed that each  $\mathcal{E}_{i_{n-1},L-(n-1)}$  (in (3.134)) is a subset of  $\mathcal{E}_{n-1,L-(n-1)}$ . The cardinality of this latter set is the result of the combination of  $L + 1$  elements taken  $L - (n - 1)$  at a time without repetition, thus we get the following

$$|\mathcal{E}_{n-1,L-(n-1)}| = \binom{L+1}{L-(n-1)} = \frac{(L+1)!}{(L-(n-1))!n!}. \quad (3.135)$$

Let us now examine the nested summation in the expression in (3.134). We can remark that the summands are the elements of  $\mathcal{E}_{n-1,L-(n-1)}$ . Hence, the result of this nested summation is nothing but a simple sum of the vectors in  $\mathcal{E}_{n-1,L-(n-1)}$ , each of which multiplied by the number of times it appears in the summation. For each vector, this number is the result of the number of possible orders in which we can remove  $(L + 1 - (L - (n - 1)))$  particular '1' coordinates from  $\mathbf{s}_{i,L+1}$ . It follows that the required numbers are all equal to each other and given by  $(L + 1 - (L - (n - 1)))! = n!$ . From the above and the fact that  $(L(L-1) \dots (L-(n-1)))^{-1} = (L-n)!(L!)^{-1}$ , the expression in (3.134) can be rewritten as

$$\begin{aligned} \mathbf{s}_{i,L+1} &= \frac{(L-n)!}{L!} \sum_{j \in \mathcal{E}_{n-1,L-(n-1)}} n! \mathbf{s}_{j,L-(n-1)} \\ &= \frac{(L-n)!n!}{L!} \frac{(L+1)!}{(L-(n-1))!n!} \sum_{j \in \mathcal{E}_{n-1,L-(n-1)}} \frac{(L-(n-1))!n!}{(L+1)!} \mathbf{s}_{j,L-(n-1)}, \end{aligned} \quad (3.136)$$

where the second equality is due to multiplying and dividing by  $|\mathcal{E}_{n-1,L-(n-1)}|$ . By noticing that the factor that multiplies the summation is equal to  $\frac{L+1}{L-(n-1)}$ , (3.136) can be re-expressed as

$$\mathbf{s}_{i,L+1} = \frac{L+1}{L-(n-1)} \sum_{j \in \mathcal{E}_{n-1,L-(n-1)}} \frac{(L-(n-1))!n!}{(L+1)!} \mathbf{s}_{j,L-(n-1)}. \quad (3.137)$$

### 3.9. Appendix

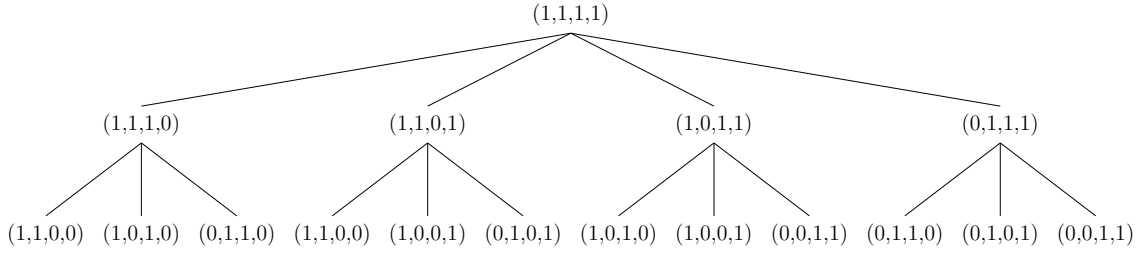


Figure 3.10: A tree that shows the vectors in  $S_2$  that yield  $\mathbf{s}_{i,4} = (1, 1, 1, 1)$ . Here,  $N = 4$  and  $n = 2$ .

From the proof of Lemma 4, we can claim that the point formed by the convex combination

$$\sum_{j \in \mathcal{E}_{n-1, L-(n-1)}} |\mathcal{E}_{n-1, L-(n-1)}|^{-1} \mathbf{s}_{j, L-(n-1)} \quad (3.138)$$

is on the convex hull of  $S_{L-(n-1)}$  and in the same direction from the origin as  $\mathbf{s}_{i, L+1}$ ; this combination is convex since we have  $|\mathcal{E}_{n-1, L-(n-1)}|^{-1} > 0$  and  $\sum_{j \in \mathcal{E}_{n-1, L-(n-1)}} |\mathcal{E}_{n-1, L-(n-1)}|^{-1} = 1$ , meaning that the coefficients of this combination are non-negative and sum to 1.

Example: In order to clarify the result of this lemma, we provide a simple example in which we set  $N = 4$  and  $L + 1 = 4$ . Under this example, we know that  $S_4$  will contain one point, namely  $\mathbf{s}_{i,4} = (1, 1, 1, 1)$ . For this point, we want to find its corresponding point on the convex hull of  $S_2$ ; this implies that  $n = 2$ . From the tree in Figure 3.10, it can be seen that

$$\begin{aligned} \mathbf{s}_{i,4} &= \frac{1}{3}((1, 1, 1, 0) + (1, 1, 0, 1) + (1, 0, 1, 1) + (0, 1, 1, 1)) \\ &= \frac{1}{3}((1, 1, 0, 0) + (1, 0, 1, 0) + (0, 1, 1, 0) + (1, 0, 0, 1) + (0, 1, 0, 1) + (0, 0, 1, 1)). \end{aligned} \quad (3.139)$$

Remark that the 6 different vectors in the second equality form the set  $\mathcal{E}_{n-1, L-(n-1)} = \mathcal{E}_{1,2}$ , thus  $|\mathcal{E}_{1,2}| = 6$ . Using  $\mathcal{E}_{1,2}$ , the point that corresponds to  $\mathbf{s}_{i,4}$  and that lies on the convex hull of  $S_2$  is given by the following

$$\frac{1}{6}(1, 1, 0, 0) + \frac{1}{6}(1, 0, 1, 0) + \frac{1}{6}(0, 1, 1, 0) + \frac{1}{6}(1, 0, 0, 1) + \frac{1}{6}(0, 1, 0, 1) + \frac{1}{6}(0, 0, 1, 1). \quad (3.140)$$

We can obtain  $\mathbf{s}_{i,4}$  by just multiplying this convex combination by a factor of 2, which verifies the general formula provided in (3.137).

This completes the proof of Lemma 5.  $\square$

• **Step 2:** To find the minimum achievable fraction between the stability region of the imperfect case ( $\Lambda_I$ ) and the stability region of the perfect case ( $\Lambda_P$ ), we examine the gap between each vertex that contributes in the characterization of  $\Lambda_P$ , where the set of these vertices is given by  $\{P_1, \dots, P_{L_P}\}$ , and the convex hull of  $\Lambda_I$ . To begin with, using the above lemma, we recall that a point in  $S_{L_P}$  can be written in function of some point that lies on the convex hull of  $S_{L_I}$ , where these two points are in the same direction toward the origin. Furthermore, the gap between these two points can be captured using the fraction  $\frac{L_I}{L_P}$ . Since any point in  $P_{L_P}$  can be written as  $\mu(L_P)$  times its corresponding point in  $S_{L_P}$  and a point on the convex hull of  $I_{L_I}$  is  $r(L_I)$  times its corresponding point on the convex hull of  $S_{L_I}$ , we can claim that the fraction between any point in  $P_{L_P}$  and its corresponding point on the

### 3.9. Appendix

convex hull of  $I_{L_I}$ , and thus on the convex hull of  $\Lambda_I$ , can be given by the following

$$\frac{L_I r(L_I)}{L_P \mu(L_P)} = \frac{r_T(L_I)}{\mu_T(L_P)}. \quad (3.141)$$

More generally, using the above approach, we can show that the fraction between any point (vertex) in  $P_L$ , for  $L_I \leq L \leq L_P$ , and the convex hull of  $\Lambda_I$  is equal to  $\frac{L_I r(L_I)}{L \mu(L)} = \frac{r_T(L_I)}{\mu_T(L)}$ . For these fractions, since  $\mu_T(L)$  increases for  $L \leq L_P$ , the following holds

$$\frac{r_T(L_I)}{\mu_T(L_P)} < \frac{r_T(L_I)}{\mu_T(L_P - 1)} < \dots < \frac{r_T(L_I)}{\mu_T(L_I)}. \quad (3.142)$$

On the other side, the fraction between any vertex in  $P_L$ , for  $2 \leq L \leq L_I$ , and its corresponding point on the convex hull of  $\Lambda_I$  is given by  $\frac{L r(L)}{L \mu(L)} = \frac{r_T(L)}{\mu_T(L)}$ . This is due to the fact that the point on the convex hull of  $\Lambda_I$  and that corresponds to a specific vertex in  $P_L$ , for  $2 \leq L \leq L_I$ , is nothing but a vertex in  $I_L$ . For the fractions in this case, it is obvious that

$$\frac{r_T(L_I)}{\mu_T(L_I)} < \frac{r_T(L_I - 1)}{\mu_T(L_I - 1)} < \dots < \frac{r_T(2)}{\mu_T(2)}. \quad (3.143)$$

Hence, using the inequalities in (3.142) and (3.143), the minimum achievable fraction is given by  $\frac{r_T(L_I)}{\mu_T(L_P)}$ . Therefore, the desired result holds.

#### 3.9.9 Proof of Theorem 8

The proof consists of three steps. We first show that  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \frac{r(L_B, B')}{r(L_B, B)} (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$ . We then find the minimum fraction  $\frac{r(L_B, B')}{r(L_B, B)}$  under the condition that the number of active pairs,  $L_B$ , can be less than or equal to  $N$ ; we get  $\frac{r(N, B')}{r(N, B)}$  as a minimum fraction. Finally, we show that the stability region  $\Lambda_{B'}$  achieves at least a fraction  $\frac{r(N, B')}{r(N, B)}$  of the stability region  $\Lambda_B$  and we conclude that  $\Lambda_{B'}$  can be bounded as given in (3.52).

• **Step 1:** Recall that under the symmetric case all the active pairs have the same average rate, which we denote here by  $r(L_B, B)$ . Thus, we can write

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)} = r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k, \quad (3.144)$$

where  $L_B = |\mathcal{L}_B|$ . Similarly, we get

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} = r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k, \quad (3.145)$$

with  $L_{B'} = |\mathcal{L}_{B'}|$ . Note that  $r(L_B, B') \leq r(L_B, B)$ , or equivalently  $\frac{r(L_B, B')}{r(L_B, B)} \leq 1$ , if  $B' \leq B$ .

Depending on the subset of scheduled pairs under each of  $\Delta_{B'}^*$  and  $\Delta_B^*$ , four cases are to consider:

1) *If the Number of Scheduled Pairs is 1 under both of  $\Delta_{B'}^*$  and  $\Delta_B^*$ :* Let us denote the active pair by  $i$ . It is clear that in this case the average rate expression is independent of the number of bits and



### 3.9. Appendix

---

thus we have  $r(1, B) = r(1, B') = r_{\text{svd}}$ . Based on this, it can be seen that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)} = (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} = r_{\text{svd}} q_i. \quad (3.146)$$

Hence, for any positive fraction  $\beta \leq 1$ , we have  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$ . We can therefore write

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \frac{r(L_B, B')}{r(L_B, B)} (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}. \quad (3.147)$$

2) *If the Number of Scheduled Pairs is 1 under  $\Delta_B^*$  and strictly greater than 1 under  $\Delta_{B'}^*$ :* It can be observed that this case cannot take place. We prove this claim using the concept of proof by contradiction. Let  $i$  denote the index of the scheduled pair under  $\Delta_B^*$ . Here we have  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)} = r_{\text{svd}} q_i$  and  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} = r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k$ . Since  $\Delta_{B'}^*$  maximizes the product  $(\mathbf{r} \cdot \mathbf{q})$  for the case where  $B'$  is the number of bits, we get

$$r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k \geq r_{\text{svd}} q_i. \quad (3.148)$$

On the other side, based on the definition of  $\Delta_B^*$  and the fact that  $r(L_B, B') < r(L_B, B)$ , we get

$$r_{\text{svd}} q_i \geq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k > r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k. \quad (3.149)$$

The above inequality holds for any  $\mathcal{L}_B$  and in particular for  $\mathcal{L}_B = \mathcal{L}_{B'}$ , we thus have

$$r_{\text{svd}} q_i \geq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k > r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k. \quad (3.150)$$

Based on equations (3.149) and (3.150), we obtain  $r_{\text{svd}} q_i > r_{\text{svd}} q_i$ , which is incorrect.

3) *If the Number of Scheduled Pairs is 1 under  $\Delta_{B'}^*$  and strictly greater than 1 under  $\Delta_B^*$ :* If we denote by  $i$  the index of the scheduled pair under  $\Delta_{B'}^*$ , we can write  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} = r_{\text{svd}} q_i$ . In addition, we have  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)} = r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k$ . Since  $\Delta_{B'}^*$  maximizes the product  $(\mathbf{r} \cdot \mathbf{q})$  for the case where the number of bits is  $B'$ , the following holds

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r_{\text{svd}} q_i. \quad (3.151)$$

In addition, using the definition of  $\Delta_B^*$  and the fact that  $r(L_B, B') \leq r(L_B, B)$ , we can write

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \quad (3.152)$$

In order to obtain

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \geq \beta r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k, \quad (3.153)$$

### 3.9. Appendix

---

for some  $\beta \leq 1$ , it suffices to take  $\beta \leq \frac{r(L_B, B')}{r(L_B, B)}$ . Setting  $\beta = \frac{r(L_B, B')}{r(L_B, B)}$ , we get

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k = \beta r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \quad (3.154)$$

Combining this equality with the inequality in (3.151) yields

$$r_{\text{svd}} q_i \geq \frac{r(L_B, B')}{r(L_B, B)} r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \quad (3.155)$$

Hence, we can deduce that  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$ , i.e.

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \frac{r(L_B, B')}{r(L_B, B)} (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}. \quad (3.156)$$

4) *If the Number of Scheduled Pairs is strictly greater than 1 under both of  $\Delta_{B'}^*$  and  $\Delta_B^*$ :* Here, the analysis is similar to the third case. In detail, since  $\Delta_{B'}^*$  maximizes the product  $(\mathbf{r} \cdot \mathbf{q})$  for the case where the number of bits is  $B'$ , it follows that

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k. \quad (3.157)$$

Also, using the definition of  $\Delta_B^*$  and the fact that  $r(L_B, B') \leq r(L_B, B)$ , we can write

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \quad (3.158)$$

In order to get

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \geq \beta r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k, \quad (3.159)$$

for some  $\beta \leq 1$ , it suffices to take  $\beta \leq \frac{r(L_B, B')}{r(L_B, B)}$ . Setting  $\beta = \frac{r(L_B, B')}{r(L_B, B)}$ , we get

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k = \beta r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \quad (3.160)$$

Combining this equality with the inequality in (3.157) yields

$$r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k \geq \frac{r(L_B, B')}{r(L_B, B)} r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \quad (3.161)$$

Therefore, we can deduce that  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \frac{r(L_B, B')}{r(L_B, B)} (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$ .

• **Step 2:** In Step 1 we have proven that, in the four considered cases, the following holds:

$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$ , with  $\beta = \frac{r(L_B, B')}{r(L_B, B)}$ .

### 3.9. Appendix

We now want to find the minimum fraction  $\frac{r(L_B, B')}{r(L_B, B)}$  w.r.t.  $L_B$ , such as

$$\underset{L_B}{\text{minimize}} \quad \frac{r(L_B, B')}{r(L_B, B)} \quad (3.162)$$

$$\text{subject to} \quad L_B \leq N \quad (3.163)$$

To solve this problem, we show that the objective function to minimize in (3.162) is a decreasing function w.r.t.  $L_B$ . Indeed, using (3.37), we have

$$\begin{aligned} \frac{r(L_B, B')}{r(L_B, B)} &= \frac{(1 - L_B \theta) dRe^{-\frac{\sigma^2 \tau}{\alpha}} (F(B'))^{L_B - 1}}{(1 - L_B \theta) dRe^{-\frac{\sigma^2 \tau}{\alpha}} (F(B))^{L_B - 1}} \\ &= \left( \frac{F(B')}{F(B)} \right)^{L_B - 1}, \end{aligned} \quad (3.164)$$

in which function  $F$  was already defined for equation (3.37). It is clear that  $F(B') < F(B)$  because  $B' < B$ , which implies that  $\left( \frac{F(B')}{F(B)} \right)^{L_B - 1}$  decreases with  $L_B$ . Since  $L_B \leq N$ , the optimization problem reaches its minimum at  $L_B = N$ . Based on the above, the minimum fraction can be given by  $\frac{r(N, B')}{r(N, B)}$ . For the rest of the proof, we define  $\beta_m = \frac{r(N, B')}{r(N, B)}$ .

• **Step 3:** Using the minimum fraction derived before, we now want to examine the stability region achieved by  $\Delta_{B'}^*$ . To this end, we define the quadratic *Lyapunov function* as

$$Ly(\mathbf{q}(t)) \triangleq \frac{1}{2} (\mathbf{q}(t) \cdot \mathbf{q}(t)) = \frac{1}{2} \sum_{k=1}^N q_k(t)^2. \quad (3.165)$$

From the evolution equation for the queue lengths (see (3.12)) we have

$$\begin{aligned} Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) &= \frac{1}{2} \sum_{k=1}^N [q_k(t+1)^2 - q_k(t)^2] \\ &= \frac{1}{2} \sum_{k=1}^N [(\max\{q_k(t) - D_k(t), 0\} + A_k(t))^2 - q_k(t)^2] \\ &\leq \sum_{k=1}^N \frac{[A_k(t)^2 + D_k(t)^2]}{2} + \sum_{k=1}^N q_k(t) [A_k(t) - D_k(t)], \end{aligned} \quad (3.166)$$

where in the final inequality we have used the fact that for any  $q \geq 0$ ,  $A \geq 0$ ,  $D \geq 0$ , we have

$$(\max\{q - D, 0\} + A)^2 \leq q^2 + A^2 + D^2 + 2q(A - D).$$

Now define  $Dr(\mathbf{q}(t))$  as the *conditional Lyapunov drift* for time-slot  $t$

$$Dr(\mathbf{q}(t)) \triangleq \mathbb{E} \{ Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t) \}. \quad (3.167)$$

### 3.9. Appendix

From (3.166), we have that  $Dr(\mathbf{q}(t))$  for a general scheduling policy satisfies

$$Dr(\mathbf{q}(t)) \leq \mathbb{E} \left\{ \sum_{k=1}^N \frac{A_k(t)^2 + D_k(t)^2}{2} \mid \mathbf{q}(t) \right\} + \sum_{k=1}^N q_k(t) a_k - \mathbb{E} \left\{ \sum_{k=1}^N q_k(t) D_k(t) \mid \mathbf{q}(t) \right\}, \quad (3.168)$$

where we have used the fact that arrivals are i.i.d. over slots and hence independent of current queue backlogs, so that  $\mathbb{E} \{A_k(t) \mid \mathbf{q}(t)\} = \mathbb{E} \{A_k(t)\} = a_k$ . Now define  $E$  as a finite positive constant that bounds the first term on the right-hand-side of the above drift inequality, so that for all  $t$ , all possible  $q_k(t)$ , and all possible control decisions that can be taken, we have

$$\mathbb{E} \left\{ \sum_{k=1}^N \frac{A_k(t)^2 + D_k(t)^2}{2} \mid \mathbf{q}(t) \right\} \leq E. \quad (3.169)$$

Note that  $E$  exists since  $A_k(t) < A_{\max}$  and  $D_k(t) < D_{\max}$ . Using the expression in (3.168) yields

$$Dr(\mathbf{q}(t)) \leq E + \sum_{k=1}^N q_k(t) a_k - \mathbb{E} \left\{ \sum_{k=1}^N q_k(t) D_k(t) \mid \mathbf{q}(t) \right\}. \quad (3.170)$$

The conditional expectation at the right-hand-side of the above inequality is with respect to the randomly observed channel states. Thus, the drift under  $\Delta_{B'}^*$  can be expressed as

$$Dr^{(\Delta_{B'}^*)}(\mathbf{q}(t)) \leq E - \sum_{k=1}^N q_k(t) \left[ \mathbb{E} \left\{ D_k^{(\Delta_{B'}^*)}(t) \mid \mathbf{q}(t) \right\} - a_k \right], \quad (3.171)$$

Note that here we have  $\mathbb{E} \left\{ D_k^{(\Delta_{B'}^*)}(t) \mid \mathbf{q}(t) \right\} = r(L_{B'}, B')$ , where the expectation at the left-hand-side of this latter equality is over the randomly observed channel state. Similarly, we have

$$\mathbb{E} \left\{ D_k^{(\Delta_B^*)}(t) \mid \mathbf{q}(t) \right\} = r(L_B, B). \quad (3.172)$$

Hence, using (3.161) and the fact that the minimum fraction is  $\beta_m = \frac{r(N, B')}{r(N, B)}$ , we can write

$$\sum_{k=1}^N q_k(t) \mathbb{E} \left\{ D_k^{(\Delta_{B'}^*)}(t) \mid \mathbf{q}(t) \right\} \geq \sum_{k=1}^N q_k(t) \beta_m \mathbb{E} \left\{ D_k^{(\Delta_B^*)}(t) \mid \mathbf{q}(t) \right\}. \quad (3.173)$$

Plugging this directly into (3.171) yields

$$Dr^{(\Delta_{B'}^*)}(\mathbf{q}(t)) \leq E - \sum_{k=1}^N q_k(t) \left[ \beta_m \mathbb{E} \left\{ D_k^{(\Delta_B^*)}(t) \mid \mathbf{q}(t) \right\} - a_k \right]. \quad (3.174)$$

The above expression can be re-expressed as

$$Dr^{(\Delta_{B'}^*)}(\mathbf{q}(t)) \leq E - \beta_m \sum_{k=1}^N q_k(t) \left[ \mathbb{E} \left\{ D_k^{(\Delta_B^*)}(t) \mid \mathbf{q}(t) \right\} - \beta_m^{-1} a_k \right]. \quad (3.175)$$

Because  $\Delta_B^*$  maximizes the weighted sum  $\sum_{k=1}^N q_k(t) \mathbb{E} \{D_k(t) \mid \mathbf{q}(t)\}$  over all alternative decisions,

### 3.9. Appendix

---

the following holds

$$\sum_{k=1}^N q_k(t) \mathbb{E} \left\{ D_k^{(\Delta_B^*)}(t) \mid \mathbf{q}(t) \right\} \geq \sum_{k=1}^N q_k(t) \mathbb{E} \left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\}. \quad (3.176)$$

where  $\Delta$  represents any alternative (possibly randomized) scheduling decision that can stabilize the system. Plugging the above directly into (3.175) yields

$$Dr^{(\Delta_B^*)}(\mathbf{q}(t)) \leq E - \beta_m \sum_{k=1}^N q_k(t) \left[ \mathbb{E} \left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\} - \beta_m^{-1} a_k \right]. \quad (3.177)$$

Let us suppose that the mean arrival rate vector  $\mathbf{a}$  is interior to fraction  $\beta_m$  of the stability region  $\Lambda_B$ . Thus, there exists an  $\epsilon_{\max}$  such that

$$(a_1 + \epsilon_{\max}, \dots, a_N + \epsilon_{\max}) \in \beta_m \Lambda_B, \quad (3.178)$$

or equivalently we have

$$(\beta_m^{-1} a_1 + \beta_m^{-1} \epsilon_{\max}, \dots, \beta_m^{-1} a_N + \beta_m^{-1} \epsilon_{\max}) \in \Lambda_B. \quad (3.179)$$

Based on the above and considering a particular policy  $\Delta$  that depends only on the states of the channels, we can write

$$\mathbb{E} \left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\} = \mathbb{E} \left\{ D_k^{(\Delta)}(t) \right\} \geq \beta_m^{-1} a_k + \beta_m^{-1} \epsilon_{\max}, \quad \forall k \in \{1, \dots, N\}. \quad (3.180)$$

Plugging the above in (3.177) yields

$$Dr^{(\Delta_{B'}^*)}(\mathbf{q}(t)) \leq E - \epsilon_{\max} \sum_{k=1}^N q_k(t). \quad (3.181)$$

Taking an expectation of  $Dr^{(\Delta_{B'}^*)}$  over the randomness of the queue lengths and summing over  $t \in \{0, 1, \dots, T-1\}$  for some integer  $T > 0$  we get

$$\mathbb{E} \{ Ly(\mathbf{q}(T)) \} - \mathbb{E} \{ Ly(\mathbf{q}(0)) \} \leq ET - \epsilon_{\max} \sum_{t=0}^{T-1} \sum_{k=1}^N \mathbb{E} \{ q_k(t) \}. \quad (3.182)$$

Rearranging terms, dividing by  $\epsilon_{\max} T$ , and taking a limsup we eventually obtain

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^N \mathbb{E} \{ q_k(t) \} \leq \frac{E}{\epsilon_{\max}}. \quad (3.183)$$

Based on the above inequality and the definition of strong stability (see Chapter 2), it follows that  $\Delta_{B'}^*$  stabilizes the system for any arrivals such that the mean arrival rate vector is interior to fraction  $\beta_m$  of the stability region of  $\Delta_{B'}^*$ , meaning that  $\Delta_{B'}^*$  achieves up to  $\Lambda_{B'} = \frac{r(N, B')}{r(N, B)} \Lambda_B$ . Note that this achievable region corresponds to the worst case, that is, when the fraction is  $\frac{r(N, B')}{r(N, B)}$ . Hence, since the

### 3.9. Appendix

fraction is greater than or equal to  $\frac{r(N, B')}{r(N, B)}$ , we get

$$\frac{r(N, B')}{r(N, B)} \Lambda_B \subseteq \Lambda_{B'} \subseteq \Lambda_B, \quad (3.184)$$

meaning that  $\Lambda_{B'}$  achieves *at least* a fraction  $\frac{r(N, B')}{r(N, B)}$  of  $\Lambda_B$ . This completes the proof.

#### 3.9.10 Derivation of Rate Approximation 1

To begin with, we note that the expression of  $r_k$  given in (3.54) can be re-expressed as

$$r_k = (1 - L\theta) d Re^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki})^Q {}_2F_1(c_2, Q; c_1 + c_2; g_{ki}), \quad (3.185)$$

which follows since  $1 - g_{ki} = \left( \zeta_{ki} \tau d (\zeta_{kk} 2^{\frac{B}{Q}})^{-1} + 1 \right)^{-1}$ . We recall that we work under the assumption/condition that the  $g_{ki}$  are sufficiently small.

We focus on the term  $(1 - g_{ki})^Q {}_2F_1(c_2, Q; c_1 + c_2; g_{ki})$ . Using linear transformations (of variable) properties for the Hypergeometric function, we have the relation [110, Page 559]

$$(1 - g_{ki})^Q {}_2F_1(c_2, Q; c_1 + c_2; g_{ki}) = {}_2F_1(c_1, Q; c_1 + c_2; \frac{g_{ki}}{g_{ki} - 1}). \quad (3.186)$$

For sufficiently small  $g_{ki}$  values, we have the approximation:  $\frac{g_{ki}}{g_{ki} - 1} \approx -g_{ki}$ . We can (numerically) verify that for  $g_{ki} < 0.1$  the following accurate approximation holds

$${}_2F_1(c_1, Q; c_1 + c_2; \frac{g_{ki}}{g_{ki} - 1}) \approx {}_2F_1(c_1, Q; c_1 + c_2; -g_{ki}). \quad (3.187)$$

We recall that  $c_1 = (Q + 1)Q^{-1}d - Q^{-1}$ ,  $c_2 = (Q - 1)c_1$  and  $Q = N_t N_r - 1$ . For sufficiently large  $Q$ , and since  $d \leq \min(N_t, N_r)$  (this implies that  $Q$  is sufficiently larger than  $d$ ), we can easily see that  $c_1 \approx d$ ,  $c_2 \approx Qd - d$  and  $c_1 + c_2 \approx Qd$ . Now, using the Maclaurin expansion to the second order we can write

$${}_2F_1(c_1, Q; c_1 + c_2; -g_{ki}) \approx 1 - \frac{c_1 Q}{c_1 + c_2} g_{ki} + \frac{1}{2} \frac{c_1 Q}{c_1 + c_2} \frac{(c_1 + 1)(Q + 1)}{c_1 + c_2 + 1} g_{ki}^2 + \mathcal{O}(g_{ki}^3) \approx 1 - g_{ki}. \quad (3.188)$$

In this approximation we have used the facts that  $\frac{1}{2} g_{ki}^2 \ll g_{ki}$ ,  $\frac{c_1 Q}{c_1 + c_2} = \frac{dQ}{Qd} = 1$ ,  $d$  is sufficiently high, and  $\frac{(c_1 + 1)(Q + 1)}{c_1 + c_2 + 1} = \frac{(d + 1)(Q + 1)}{Qd + 1} \approx 1$ . In addition,  $\mathcal{O}(g_{ki}^3)$  can be removed since it is negligible compared with  $1 - g_{ki}$ . This latter property follows from the fact that the Maclaurin expansion to higher orders (greater than two) will add terms in  $g_{ki}^3, g_{ki}^4, \dots$ , which are, as  $g_{ki}^2$ , very small with respect to 1 and to the term in  $g_{ki}$ ; this is due to the condition that  $g_{ki}$  is sufficiently small (i.e.  $g_{ki} < 0.1$ ). Hence, by replacing the above approximation in the expression of  $r_k$  given in (3.185), we obtain the following

$$r_k \approx (1 - L\theta) d Re^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}). \quad (3.189)$$

### 3.9.11 Derivation of Rate Approximation 2

Let us first recall that here we suppose that all the  $g_{ki}$  (with  $i \neq k$ ) are relatively close to  $\bar{g}_k$ . We focus on the product  $\prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki})$ . Our goal here is to show that this product can be accurately approximated by the following expression

$$(1 - \bar{g}_k)^{L-1} - (1 - \bar{g}_k)^{L-2} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k). \quad (3.190)$$

To prove this latter result, we start by a simple example and then we provide the general result. We consider a simple example where the product function is

$$f(g_{k1}, g_{k2}) = (1 - g_{k1})(1 - g_{k2}), \quad (3.191)$$

i.e. it can be seen as an example where  $L = 3$  and  $k = 3$ . For this function, the Taylor expansion of order 2 around the point  $(\bar{g}_k, \bar{g}_k)$  can be given as the following

$$\begin{aligned} f(g_{k1}, g_{k2}) &= f(\bar{g}_k, \bar{g}_k) + (g_{k1} - \bar{g}_k) \frac{\partial f}{\partial g_{k1}}|_{(\bar{g}_k, \bar{g}_k)} + (g_{k2} - \bar{g}_k) \frac{\partial f}{\partial g_{k2}}|_{(\bar{g}_k, \bar{g}_k)} \\ &\quad + \frac{1}{2} (g_{k1} - \bar{g}_k)^2 \frac{\partial^2 f}{\partial g_{k1}^2}|_{(\bar{g}_k, \bar{g}_k)} + \frac{1}{2} (g_{k2} - \bar{g}_k)^2 \frac{\partial^2 f}{\partial g_{k2}^2}|_{(\bar{g}_k, \bar{g}_k)} \\ &\quad + (g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k) \frac{\partial f}{\partial g_{k1}}|_{(\bar{g}_k, \bar{g}_k)} \frac{\partial f}{\partial g_{k2}}|_{(\bar{g}_k, \bar{g}_k)} \\ &= (1 - \bar{g}_k)(1 - \bar{g}_k) - (g_{k1} - \bar{g}_k)(1 - \bar{g}_k) - (g_{k2} - \bar{g}_k)(1 - \bar{g}_k) + (g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k) \\ &= (1 - \bar{g}_k)^{3-1} - (1 - \bar{g}_k)^{3-2} (g_{k1} - \bar{g}_k + g_{k2} - \bar{g}_k) + (g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k). \end{aligned} \quad (3.192)$$

Note that in the first line of the above equation we use operator  $=$  (and not  $\approx$ ) since, as it can be easily noticed, we have all the expansions of  $f$  for order  $\geq 3$  are equal to the one for order 2. This latter statement results from the fact that  $\frac{\partial^\alpha f}{\partial g_{ki}^\alpha} = 0, \forall \alpha \geq 2$ . Since we work under the condition that all the  $g_{ki}$  are close to  $\bar{g}_k$ , we can claim that the term  $(g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k)$  (in (3.192)) is sufficiently small compared with the other terms. Hence, we get

$$f(g_{k1}, g_{k2}) \approx (1 - \bar{g}_k)^{3-1} - (1 - \bar{g}_k)^{3-2} (g_{k1} - \bar{g}_k + g_{k2} - \bar{g}_k). \quad (3.193)$$

The obtained result can be easily generalized, and thus we can write

$$\prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}) \approx (1 - \bar{g}_k)^{L-1} - (1 - \bar{g}_k)^{L-2} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k). \quad (3.194)$$

By recalling that the approximated average rate expression is  $(1 - L\theta) dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki})$  and replacing  $\prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki})$  with its approximation given in (3.194), we eventually obtain

$$r_k \approx (1 - L\theta) dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \left[ (1 - \bar{g}_k)^{L-1} - (1 - \bar{g}_k)^{L-2} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right]. \quad (3.195)$$

### 3.9.12 Proof of Theorem 9

The proof consists of two parts. The first part provides the explicit expression of fraction  $\beta_A$ . Then, in the second part we show that the approximate policy  $\Delta^A$  achieves at least a fraction  $\beta_A$  of the stability region achieved by  $\Delta_{GI}^*$ .

• **Step 1:** As in the first part of the proof of Theorem 8, depending on the number of scheduled pairs under each policy, we consider (the same) four cases.

1) *If the Number of Scheduled Pairs is 1 under both of  $\Delta^A$  and  $\Delta_{GI}^*$ :* We recall that in this case the scheduled pair is denoted by  $j$  and its average rate is given by  $r_{\text{svd},j}$ , independently of whether we consider the perfect or imperfect case. Thus, the dot product  $\mathbf{r} \cdot \mathbf{q}$  can be written as  $r_{\text{svd},j} q_j$  under  $\Delta^A$  as well as under  $\Delta_{GI}^*$ . Hence, we have  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} = (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)}$ , and for any positive constant  $\beta_A \leq 1$  we can deduce that the following inequality holds

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq \beta_A (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)}. \quad (3.196)$$

2) *If the Number of Scheduled Pairs is strictly greater than 1 under both of  $\Delta^A$  and  $\Delta_{GI}^*$*  Using Rate Approximation 2, under policy  $\Delta^A$  the dot product  $\mathbf{r} \cdot \mathbf{q}$  can be expressed as

$$(1 - L_A \theta) dR \left[ \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A - 1} q_k - \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A - 2} q_k \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right], \quad (3.197)$$

whereas under  $\Delta_{GI}^*$  this dot product is given by

$$(1 - L \theta) dR \left[ \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L - 1} q_k - \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L - 2} q_k \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right]. \quad (3.198)$$

Since the approximate policy  $\Delta^A$  schedules the subset  $\mathcal{L}_A$  that maximizes  $\phi \cdot \mathbf{q}$ , and recalling that

$$\phi_k(l) = (1 - l \theta) dR e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{l - 1}, \quad (3.199)$$

it can be easily seen that

$$(1 - L_A \theta) dR \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A - 1} q_k \geq (1 - L \theta) dR \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L - 1} q_k. \quad (3.200)$$

Similarly, using the definition of the optimal policy  $\Delta_{GI}^*$  under which the dot product  $\mathbf{r} \cdot \mathbf{q}$  is maximized, the following inequality holds

$$\begin{aligned} & (1 - L \theta) dR \left[ \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L - 1} q_k - \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L - 2} q_k \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right] \geq \\ & (1 - L_A \theta) dR \left[ \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A - 1} q_k - \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A - 2} q_k \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right]. \end{aligned} \quad (3.201)$$



### 3.9. Appendix

These last two inequalities, in (3.200) and (3.201), lead us to the simple observation

$$\begin{aligned} & - (1 - L\theta)dR \left[ \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L-2} q_k \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right] \geq \\ & - (1 - L_A \theta)dR \left[ \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A-2} q_k \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right]. \end{aligned} \quad (3.202)$$

For the rest of this proof, we define  $o_1, o_2, p_1, p_2$  as

$$o_1 = (1 - L_A \theta)dR \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A-1} q_k, \quad o_2 = (1 - L\theta)dR \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L-1} q_k,$$

$$p_1 = -(1 - L_A \theta)dR \left[ \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A-2} q_k \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right],$$

$$p_2 = -(1 - L\theta)dR \left[ \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L-2} q_k \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right].$$

We can easily notice that  $p_1$  and  $p_2$  can be rewritten, respectively, as

$$p_1 = -(1 - L_A \theta)dR \left[ \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{-1} (1 - \bar{g}_k)^{L_A-1} q_k \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right], \quad (3.203)$$

$$p_2 = -(1 - L\theta)dR \left[ \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{-1} (1 - \bar{g}_k)^{L-1} q_k \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right]. \quad (3.204)$$

We next point out two simple but important remarks that will help us complete the proof.

- ◇ For any policy  $\Delta_2$  that approximates any policy  $\Delta_1$  to a fraction  $\beta$  ( $\leq 1$ ), we have

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_2)} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta_1)}. \quad (3.205)$$

If w.r.t. the approximate policy  $(\Delta_2)$  there exists a scheduling policy  $\Delta_{22}$  such that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{22})} \leq (\mathbf{r} \cdot \mathbf{q})^{(\Delta_2)}, \quad (3.206)$$

then we can derive a fraction based on  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{22})}$  instead of  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_2)}$ . We can easily notice that this fraction is lower than or equal to  $\beta$ , therefore, w.r.t. the stability region achieved by  $\Delta_1, \Delta_2$  reaches a fraction larger than that achieved by  $\Delta_{22}$ .

- ◇ If w.r.t. the approximated policy  $(\Delta_1)$  there exists a scheduling policy  $\Delta_{11}$  such that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_1)} \leq (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{11})},$$

### 3.9. Appendix

then we can derive an achievable fraction based on  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{11})}$ , and this fraction will be lower than or equal to  $\beta$ . The key idea here is that sometimes it is easier to find the fraction using  $\Delta_{22}$  (resp.,  $\Delta_{11}$ ) instead of  $\Delta_2$  (resp.,  $\Delta_1$ ), but this will be to the detriment of finding an achievable fraction that is, in general, lower than the exact solution.

To proceed further, we consider the extreme cases of  $p_1$  and  $p_2$  which correspond to define

$$p_{1e} = \min_{\mathcal{L}_A \in \mathfrak{L}} \left\{ \min_{k \in \mathcal{L}_A} \left\{ -(1 - \bar{g}_k)^{-1} \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right\} \right\} (1 - L_A \theta) dR \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A - 1} q_k,$$

$$p_{2e} = \max_{\mathcal{L} \in \mathfrak{L}} \left\{ \max_{k \in \mathcal{L}} \left\{ -(1 - \bar{g}_k)^{-1} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right\} \right\} (1 - L \theta) dR \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L - 1} q_k.$$

It is obvious that  $p_1 \geq p_{1e}$  and  $p_2 \leq p_{2e}$ . Let us define  $\tilde{m}_3$  and  $\tilde{m}_4$  as

$$\tilde{m}_3 = \min_{\mathcal{L}_A \in \mathfrak{L}} \left\{ \min_{k \in \mathcal{L}_A} \left\{ -(1 - \bar{g}_k)^{-1} \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right\} \right\},$$

$$\tilde{m}_4 = \max_{\mathcal{L} \in \mathfrak{L}} \left\{ \max_{k \in \mathcal{L}} \left\{ -(1 - \bar{g}_k)^{-1} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right\} \right\}.$$

Then, it is easy to see that  $p_{1e} = \tilde{m}_3 o_1$  and  $p_{2e} = \tilde{m}_4 o_2$ . This yields the following

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} = o_1 + p_1 \geq o_1 + p_{1e} = o_1 + \tilde{m}_3 o_1, \quad (3.207)$$

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)} = o_2 + p_2 \leq o_2 + p_{2e} = o_2 + \tilde{m}_4 o_2. \quad (3.208)$$

As mentioned earlier,  $\Delta^A$  approximates  $\Delta_{GI}^*$  to a fraction  $\beta$  if the following inequality holds

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)}. \quad (3.209)$$

In our case, it is difficult to derive  $\beta$ , however we can compute a fraction  $\beta_A \leq \beta$ . In detail, based on the two properties about the achievable fraction given in the above paragraph, and combining (3.207) with (3.208), the problem turns out to find  $\beta_A$  such that

$$o_1(1 + \tilde{m}_3) \geq \beta_A o_2(1 + \tilde{m}_4). \quad (3.210)$$

Using the fact that  $o_2 \leq o_1$ , which was shown at the beginning of this proof, it suffices to have  $\beta_A \leq \frac{1 + \tilde{m}_3}{1 + \tilde{m}_4}$ , to satisfy the inequality in (3.210). Let us consider the upper bound  $\beta_A = \frac{1 + \tilde{m}_3}{1 + \tilde{m}_4}$ . Therefore, we can deduce that  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq \beta_A (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{GI}^*)}$ .

**3)** *If the Number of Scheduled Pairs is 1 under  $\Delta^A$  and strictly greater than 1 under  $\Delta_{GI}^*$*  Denoting by  $j$  the index of the scheduled pair under  $\Delta^A$ , we can write  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} = r_{\text{svd},j} q_j$ . On the other side,

### 3.9. Appendix

as in the second case,  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)}$  can be expressed as

$$(1 - L\theta)dR \left[ \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L-1} q_k - \sum_{k \in \mathcal{L}} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L-2} q_k \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k) \right]. \quad (3.211)$$

Using the expression of  $\tilde{m}_4$  that was given in the second case, it can be shown that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq \frac{1}{1 + \tilde{m}_4} (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)}. \quad (3.212)$$

The proof is similar to the proof of the second case and thus is omitted for the sake of brevity.

4) *If the Number of Scheduled Pairs is 1 under  $\Delta_{\text{GI}}^*$  and strictly greater than 1 under  $\Delta^A$*  Denoting by  $j$  the index of the scheduled pair under  $\Delta_{\text{GI}}^*$ , we have  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} = r_{\text{svd},j} q_j$ . On the other hand, as in the second case,  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)}$  can be expressed as

$$(1 - L_A \theta)dR \left[ \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A-1} q_k - \sum_{k \in \mathcal{L}_A} e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L_A-2} q_k \sum_{i \in \mathcal{L}_A, i \neq k} (g_{ki} - \bar{g}_k) \right]. \quad (3.213)$$

Using the expression of  $\tilde{m}_3$  that was provided in the second case, it can be shown that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq (1 + \tilde{m}_3)(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)}. \quad (3.214)$$

The proof here is also similar to the proof of the second case and thus is omitted.

• **Step 2:** We first recall the assumptions  $-1 < \tilde{m}_3 < 0$  and  $\tilde{m}_4 > 0$ . Thus, it can be easily seen that we have  $\frac{1+\tilde{m}_3}{1+\tilde{m}_4} < 1 + \tilde{m}_3$  and  $\frac{1+\tilde{m}_3}{1+\tilde{m}_4} < \frac{1}{1+\tilde{m}_4}$ . Hence, based on these inequalities and on Step 1, and recalling that  $\beta_A = \frac{1+\tilde{m}_3}{1+\tilde{m}_4}$ , under the four cases we can write

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^A)} \geq \beta_A (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)}. \quad (3.215)$$

Now, to complete the proof, we use a similar approach to that used in Step 3 of the proof for Theorem 8. Specifically, the drift under  $\Delta^A$  can be expressed as

$$Dr^{(\Delta^A)}(\mathbf{q}(t)) \leq E - \sum_{k=1}^N q_k(t) \left[ \mathbb{E} \left\{ D_k^{(\Delta^A)}(t) \mid \mathbf{q}(t) \right\} - a_k \right], \quad (3.216)$$

for some finite constant  $E$ . Using equation (3.215), we can write

$$\sum_{k=1}^N q_k(t) \mathbb{E} \left\{ D_k^{(\Delta^A)}(t) \mid \mathbf{q}(t) \right\} \geq \sum_{k=1}^N q_k(t) \beta_A \mathbb{E} \left\{ D_k^{(\Delta_{\text{GI}}^*)}(t) \mid \mathbf{q}(t) \right\}. \quad (3.217)$$

Plugging this directly into (3.216) yields

$$Dr^{(\Delta^A)}(\mathbf{q}(t)) \leq E - \beta_A \sum_{k=1}^N q_k(t) \left[ \mathbb{E} \left\{ D_k^{(\Delta_{\text{GI}}^*)}(t) \mid \mathbf{q}(t) \right\} - a'_k \right], \quad (3.218)$$

in which  $a_k = \beta_A a'_k$ . After some manipulations, which are very similar to those used in the proof of

### 3.9. Appendix

Theorem 8, we eventually obtain

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^N \mathbb{E} \{q_k(t)\} \leq \frac{E}{\epsilon_{\max}(\mathbf{a}')}, \quad (3.219)$$

for some  $\epsilon_{\max}(\mathbf{a}')$  and  $T$ . It follows that  $\Delta^A$  stabilizes any arrival rate vector  $\mathbf{a} = \beta_A \mathbf{a}'$ . Hence, since  $\mathbf{a}'$  can be any point in the stability region of  $\Delta_{\text{GI}}^*$ , we can state that  $\Delta^A$  stabilizes any arrival rate vector interior to fraction  $\beta_A$  of the stability region of  $\Delta_{\text{GI}}^*$ . A last point to note is that the term “at least” in the theorem is justified by the fact that  $\beta_A$  is lower than or equal to the exact solution ( $\beta$ ). Therefore, the desired statement follows.

#### 3.9.13 Proof of Theorem 10

As in the first part of the proof of Theorem 8, depending on the number of scheduled pairs under each policy, we distinguish (the same) four cases.

*1) If the Number of Scheduled Pairs is equal to 1 under both of  $\Delta_{\text{GI}}^*$  and  $\Delta_{\text{GP}}^*$ :* We recall that in this case the scheduled pair is denoted by  $j$  and its average rate is given by  $r_{\text{svd},j}$ , independently of whether we consider the perfect case or the imperfect case. Hence, the dot product  $\mathbf{r} \cdot \mathbf{q}$  can be written as  $r_{\text{svd},j} q_j$  under  $\Delta_{\text{GP}}^*$  as well as under  $\Delta_{\text{GI}}^*$ . Hence, we have  $(\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)} = (\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)}$ , and for any (positive constant)  $\beta_P \leq 1$  the following inequality holds

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} \geq \beta_P (\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}. \quad (3.220)$$

*2) If the Number of Scheduled Pairs is strictly greater than 1 under both of  $\Delta_{\text{GI}}^*$  and  $\Delta_{\text{GP}}^*$ :* Under policy  $\Delta_{\text{GI}}^*$  and using the approximate expression of  $r_k$  given in (3.63), the product  $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)}$  can be written as the following

$$(1 - L\theta) \left[ \sum_{k \in \mathcal{L}} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}) \right]. \quad (3.221)$$

On the other hand, using the definition of  $\Delta_{\text{GP}}^*$ , the product  $(\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}$  can be expressed as

$$(1 - L_P \theta) \sum_{k \in \mathcal{L}_P} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k. \quad (3.222)$$

One can easily remark that this latter expression has the following equivalent representation, which results from multiplying and dividing by the same term  $\prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki})$ ,

$$(1 - L_P \theta) \sum_{k \in \mathcal{L}_P} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \left[ \prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki}) \right] \left[ \prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki}) \right]^{-1}. \quad (3.223)$$

### 3.9. Appendix

The extreme case of  $(\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}$  corresponds to

$$\tilde{m}_5(1 - L_P\theta) \left[ \sum_{k \in \mathcal{L}_P} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki}) \right], \quad (3.224)$$

where  $\tilde{m}_5$  is defined as follows

$$\tilde{m}_5^{-1} = \min_{\mathcal{L}_P \in \mathfrak{L}} \left\{ \min_{k \in \mathcal{L}_P} \left\{ \prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki}) \right\} \right\}. \quad (3.225)$$

Since, by definition, policy  $\Delta_{\text{GI}}^*$  produces the subset  $\mathcal{L}$  and maximizes the product  $(\mathbf{r} \cdot \mathbf{q})$ , it yields

$$(1 - L\theta) \left[ \sum_{k \in \mathcal{L}} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}) \right] \geq (1 - L_P\theta) \left[ \sum_{k \in \mathcal{L}_P} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki}) \right]. \quad (3.226)$$

As explained earlier, the stability region achieved by  $\Delta_{\text{GI}}^*$  approximates the one achieved by  $\Delta_{\text{GP}}^*$  to a fraction  $\beta$  if the following holds

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} \geq \beta (\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}. \quad (3.227)$$

It is hard to find  $\beta$  based on the product  $(\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}$ , however, using a similar observation to that provided at the end of the proof of Theorem 9, we can compute a fraction  $\beta_P \leq \beta$  based on an upper bound on this product. In detail, using (3.224), which represents this upper bound, our problem turns out to find  $\beta_P$  such that

$$\tilde{m}_5(1 - L_P\theta) \left[ \sum_{k \in \mathcal{L}_P} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \prod_{i \in \mathcal{L}_P, i \neq k} (1 - g_{ki}) \right] \leq \beta_P^{-1}(1 - L\theta) \left[ \sum_{k \in \mathcal{L}} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}) \right]. \quad (3.228)$$

It suffices to take  $\beta_P^{-1} \geq \tilde{m}_5$ , or equivalently  $\beta_P \leq \tilde{m}_5^{-1}$ , to satisfy the above inequality. By considering  $\beta_P = \tilde{m}_5^{-1}$ , the inequality in (3.228) becomes an equality.

Combining the above with equation (3.226), we get

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} \geq \beta_P (\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}. \quad (3.229)$$

**3)** *If the Number of Scheduled Pairs is 1 under  $\Delta_{\text{GI}}^*$  and strictly greater than 1 under  $\Delta_{\text{GP}}^*$ :* Denoting by  $j$  the index of the scheduled pair under  $\Delta_{\text{GI}}^*$ , we can write

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} = r_{\text{svd},j} q_j. \quad (3.230)$$

On the other hand, as in the second case,  $(\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}$  can be expressed as

$$(1 - L_P\theta) \sum_{k \in \mathcal{L}_P} dRe^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} q_k. \quad (3.231)$$

### 3.9. Appendix

---

Using the expression of  $\tilde{m}_5$  that was given in the second case, it can be shown that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} \geq \tilde{m}_5^{-1}(\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}. \quad (3.232)$$

The proof is similar to the proof of the second case and thus is omitted for the sake of brevity.

4) *If the Number of Scheduled Pairs is 1 under  $\Delta_{\text{GP}}^*$  and strictly greater than 1 under  $\Delta_{\text{GI}}^*$ :* It can be seen that such a case cannot arise. This can be shown using the concept of proof by contradiction. The proof is simple and will not be provided for the sake of brevity.

Based on the different cases given above, we can deduce that the following inequality holds

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{\text{GI}}^*)} \geq \beta_{\text{P}} (\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta_{\text{GP}}^*)}, \quad (3.233)$$

where  $\beta_{\text{P}} = \tilde{m}_5^{-1}$ . Finally, we note that the rest of the proof can be done in a similar way as in the proofs of Theorem 8 and Theorem 9, so we omit this part to avoid repetition.

This completes the proof.

#### 3.9.14 Proof of Lemma 6

Using Lemma 2, we can claim that  $(1 - L\theta)(F(\tau_j))^{L-1}$  decreases with  $L$ . Thus,  $r_{\text{M}}(L)$  is a decreasing function since it is the sum of  $M$  decreasing functions. On the other hand, it can be seen that  $r_{\text{MT}}(L)$  is nothing but the weighted sum (with positive coefficients) of  $M$  functions that have similar behavior as  $r_{\text{T}}(L)$ , where we recall that this latter function increases from point  $r_{\text{T}}(0) = 0$ , reaches its maximum and then decreases to point  $r_{\text{T}}(\frac{1}{\theta}) = 0$ . We can then observe that for  $L = 0$  or  $L = \frac{1}{\theta}$ ,  $r_{\text{MT}}(L) = 0$ . Also, the resulting function is positive since the  $M$  functions and their corresponding coefficients are positive. Moreover, since each one of these  $M$  functions reaches one maximum, we can claim that  $r_{\text{T}}(L)$  might have several maxima, and consequently one of these maxima will be a global maximum. Therefore, the desired result follows.

## Chapter 4

# Opportunistic Feedback Reporting and Scheduling Scheme for Multichannel Wireless Networks

### 4.1 Overview

In this chapter, we address the problem of joint feedback reporting and scheduling for multiuser downlink wireless networks employing multiple parallel channels, i.e. multi-carrier techniques, to serve the users. Such a setting corresponds for example to a single cell OFDMA scheme, which is implemented in the long term evolution (LTE) standards [9] and was shown to deliver a substantial increase in the system's performance. To exploit multiuser diversity in multichannel downlink networks, the BS needs to acquire CSI from users. These CSIs are usually unknown at the BS, especially in FDD systems which lack of channel reciprocity. A common method to acquire the downlink CSI is to allocate a part of the uplink resources to the users to report their CSIs. However, the more CSIs are decided to be reported, the more resources are needed, thus resulting in a bigger overhead in the system. In this regard, in [111–116] different approaches are proposed to reduce the feedback load while still achieving the benefits of multiuser diversity. However, these works do not take into account the incoming traffic processes of the users. In this chapter, an important factor that is considered is the traffic pattern for each user, which is stored in a respective queue at the BS. So the system stability (i.e. when all the queue lengths are finite) is an important property the scheduling mechanism should take into account.

The feedback allocation algorithm directly impacts the scheduling mechanism and thus the system stability since, in general, a user cannot be scheduled unless its CSI is reported to the scheduler. In realistic scenarios, complete feedback knowledge is not readily available at the scheduler. Different limitations result in such incomplete information, such as estimation error, delay and limited feedback resources, so it is important to analyze the impact of such imperfections on the system stability. In this regard, in [99], where a TDD mode of feedback is used, the authors have explored the resulting

trade-off between acquiring CSI (which consumes a fraction of the time-slot) and exploiting channel diversity to the various receivers; in this work, optimal stopping theory is used and it is assumed that the distributions of the channel gains are known to the BS. For a similar model, the authors in [42,117] have proposed a feedback scheme for a single channel system. This scheme sets as threshold the rate of the user with the maximum queue length and requires no knowledge of channel and traffic statistic. In addition, distributed scheduling approach is considered in [75] where the authors propose a Greedy Maximal Scheduling, for which the achievable stability performance depends on the network topology. Some other distributed schemes that approach the system stability region are provided in [76,77], but are very complex to implement. Moreover, the work in [41] derives the optimal feedback scheme for a single-channel downlink system under partial channel state information. In [118], the authors study centralized scheduling with rate adaptation under imperfect channel-estimator joint statistics.

Under a multichannel downlink system where an FDD mode is adopted, obtaining complete feedback would require a prohibitive portion of the overall uplink capacity, especially for a large number of users. Under such a system, the authors in [44] study the impact of limited feedback resources on the achievable stability region. However, note that in this work the delay in the feedback process is not accounted for; the delay and the amount of feedback resources are normally coupled, i.e. more feedback resources incur more delay in the feedback process. Also in the context of multichannel wireless downlink networks, the authors in [46] propose a set of low computational complexity scheduling algorithms with a large number of users and proportionally large bandwidth. Furthermore, in [56] the problem of routing/scheduling in a wireless network with delayed network (channel and queue) state information (NSI) is studied. Specifically, two cases are considered: the centralized routing/scheduling, where a central controller obtains heterogeneously delayed information from each of the nodes, and the decentralized routing/scheduling, where each node makes a decision based on its current channel and queue states along with homogeneous delayed NSI from other nodes. The authors in [58] analyze the effect that delayed CSI has on the throughput performance of scheduling in wireless networks. Specifically, by accounting for the delays in CSI as they relate to the network topology, the authors revisit the comparison between centralized and distributed scheduling, which is analyzed as a trade-off between using delayed CSI and making imperfect scheduling decisions.

Another interesting case study is the system with *cooperative relaying*, or more generally the multi-hop wireless system. In fact, one of the main challenges facing the networking community is the provision of high throughput for users at the cell edge. Such users often suffer from bad channel conditions. The use of cooperative relaying is a very promising solution to tackle this problem as it provides throughput gains as well as coverage extension [11]. Combining multi-carrier techniques (e.g. orthogonal frequency division multiplexing (OFDM)) and cooperative relaying ensures high throughput requirements, particularly for users at the cell edge. Moreover, relaying is considered as a cost effective throughput enhancement in IEEE 802.16j and LTE-A standards. To greatly exploit the benefits of relaying, efficient resource allocation is crucial in multiuser environment. Investigating the performance of systems with cooperative relaying has been the subject of extensive research [119–125]. However, relatively speaking, the works that consider a system with cooperative relaying, and more generally a multi-hop system, and where the traffic processes of the users are taken into account are not that many, because of the difficulty of studying such systems; usually, for such works, no imperfections are considered in the model since otherwise another dimension of complexity will be



added to the analysis. An example of these works is [74] where, in the context of multichannel relay networks, the authors study an iterative Max-Weight algorithm for routing and scheduling and show that this algorithm can stabilize the system in several large-scale settings. Also, it was shown that this algorithm outperforms the Back-Pressure algorithm from a queue-length/packet-delay perspective. In [70], optimal algorithms for minimizing the end-to-end buffer usage in a multi-flow multi-hop wireless network are proposed. From a network stability viewpoint for Max-Weight, in [71] the authors show that the network is stable if the routes are fixed.

In this chapter, we consider a multichannel multiuser wireless downlink network where both limited feedback resources and delayed feedback information are accounted for. To the best of our knowledge, this is the first work to account for these two imperfections at the same time for a multichannel system where the incoming traffic processes of the users are taken into consideration. Two scenarios are considered and examined, namely (i) the system without relaying, and (ii) the system with relaying. For the latter scenario, an additional imperfection is accounted for: the users do not have complete knowledge of the fading coefficients of the links between the BS and the relay; the only information a user know about any one of these fading coefficients is whether its corresponding SNR is greater than or equal to a certain threshold. Under the two considered scenarios, the incoming data for each user is stored in a respective buffer at the BS. Let  $L$  be the total number of channels in the system. Note that in this chapter the term "channel" denotes a certain frequency band, whereas the term "link" refers to the wireless connection between a user and the BS over a specific frequency band. Regarding the feedback resources, we adopt a setting where the feedback capacity per slot is limited. Specifically, only  $\bar{F}$  link states (i.e. CSI) can be reported to the BS per slot. However, the system can decide to report an amount of feedback  $F > \bar{F}$ , thus the feedback process will require more than one slot in order to be accomplished. Hence, the amount of feedback resources and the delay in the feedback process are coupled. One can notice the importance of the trade-off between having more  $F$ , which leads to a greater number of reported link states but which are more outdated, and having less  $F$ , which means a lower number of reported links but which are more accurate (i.e. less outdated). Since the feedback process directly impacts the scheduling mechanism, we design efficient joint feedback reporting and scheduling algorithms. It is worth mentioning that the performances of such algorithms are measured w.r.t. the performance of the *ideal* system (where all the link states are fully and perfectly known at the BS at no cost).

## Contributions

The *main contributions* and the *organization* of this chapter are summarized as follows.

- In Section 4.2 we present the system model for the scenarios with and without relaying, the limited and delayed CSI scheme, and the queueing model.
- In Section 4.3 we present the stability analysis for the system without relaying taking into account the delay in the feedback process. Specifically, for this system:
  - ★ We provide an algorithm that uses exactly  $L$  feedback resources, where the feedback and scheduling decisions are done at the users side; such an approach is considered to take advantage of the local CSI knowledge of these users in order to achieve better gains. This algorithm works

## 4.2. System Model

under the assumption of continuous time for contention. Further, we investigate the stability performance of this algorithm and show that it can guarantee a certain fraction from the stability region of the ideal system.

- ★ We propose a second algorithm that does not make the continuous contention time assumption, i.e. it is adapted to a discrete-time contention scheme, and that imitates the first algorithm to a great extent. The proposed algorithm here is based on a threshold-based concept and consists in having  $F$  ( $> L$ ) feedback resources and in letting the users decide whether they should send their CSIs or not, and then the BS performs scheduling over each channel. An efficient approach to update the threshold value is also provided. We point out that in the numerical results we find the best trade-off in terms of feedback resources (i.e. best  $F$ ) under various system setups.
- In Section 4.4, the stability analysis for the system with relaying is presented. Specifically:
  - ★ We first consider the imperfect knowledge of the fading coefficients between the BS and the relay at the users without accounting for the delay in the feedback process. Under this setting, we provide a joint feedback and scheduling algorithm, and we analyze the stability performance of this algorithm. In the special case of a single rate level, we characterize the minimum fraction this algorithm can achieve w.r.t. the stability region of the ideal system; note that the fraction for the multiple rate levels case is extremely hard to characterize.
  - ★ We then add the impact of the delay in the feedback process and propose two feedback reporting and scheduling algorithms which are simple variations of the two algorithms that are proposed under the system without relaying.
- Numerical results and relevant discussions for the systems with and without relaying are carried out in Section 4.5.
- We finally conclude the chapter in Section 4.6.

## 4.2 System Model

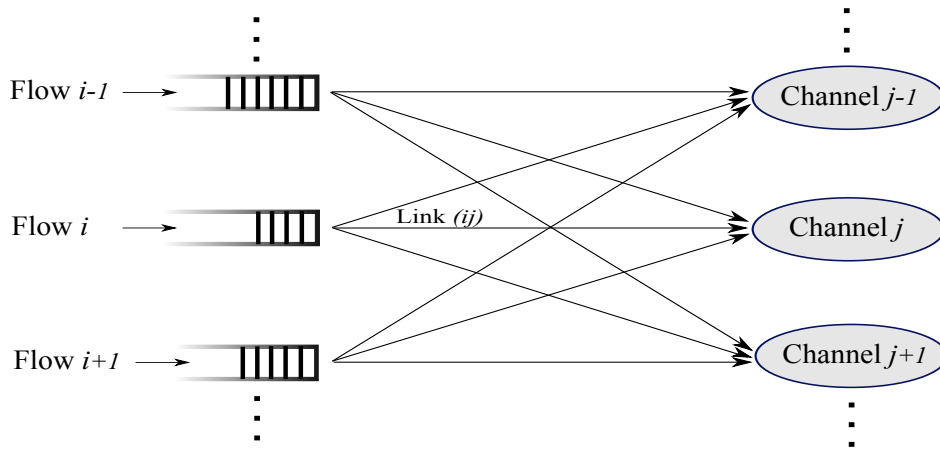


Figure 4.1: Multi-channel multi-user queueing system.

## 4.2. System Model

We consider an FDD cellular wireless network, with one single-antenna BS, one relay,  $N$  single-antenna mobile users and  $L$  channels; e.g. LTE-OFDMA system. Data packets to be transmitted to the users are stored in  $N$  separate queues at the BS. Time is slotted with the slots of all users synchronized. Let  $q_i(t)$  denote the length of queue  $i$  at the beginning of time-slot  $t$ , and let  $\mathbf{q}(t) = (q_1(t), \dots, q_N(t))$ . This system can be seen as a multi-queue, multi-server, discrete-time queueing system. It can be noticed that since we work under an OFDMA-like system, at a given slot a channel can be allocated to one and at most one user. Under the adopted network, and for clarity of exposition and analysis, we consider two different scenarios:

1. *System without relaying*: the relay is not considered under this scenario.
2. *System with relaying*: the relay is used under this scenario.

We next present the system model under each of the two scenarios given above.

### 4.2.1 System without Relaying

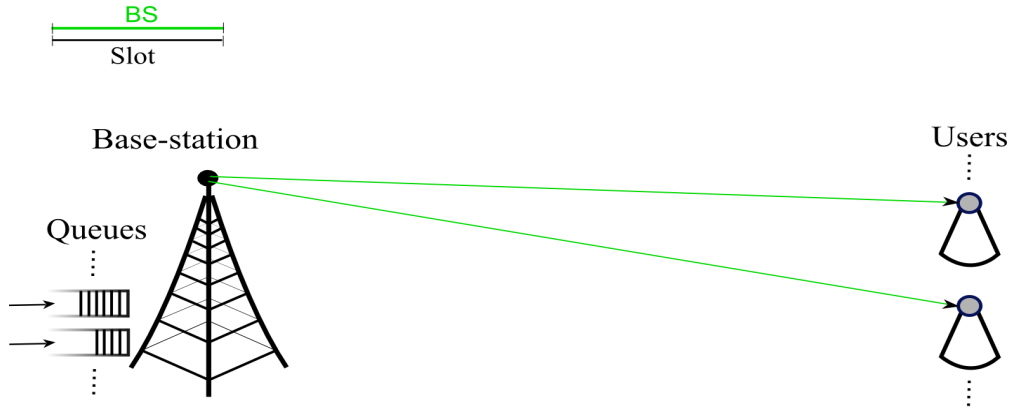


Figure 4.2: System model without relaying.

We denote by  $h_{ij}(t)$  the fading of the link connecting the BS and user  $i$  using channel  $j$ . The received signal for the  $i$ th user if it gets scheduled on the  $j$ th channel at time-slot  $t$  can be given by the following expression

$$\sqrt{P}h_{ij}(t)x_{ij}(t) + z_{ij}(t), \quad (4.1)$$

where  $x_{ij}(t)$  is the corresponding complex-valued data stream, with  $\mathbb{E}\{|x_{ij}|^2\} = 1$ ,  $P$  denotes the transmission power on each channel, and  $z_{ij}(t)$  is the additive white Gaussian noise with zero mean and variance  $\sigma_{z,ij}^2$ , i.e.  $z_{ij}(t) \sim \mathcal{CN}(0, \sigma_{z,ij}^2)$ . We assume that at time-slot  $t$  user  $i$  has perfect knowledge of the  $h_{ij}(t)$ ,  $\forall j$ . The corresponding SNR is denoted and given as

$$\gamma_{ij}(t) = \rho_{ij}|h_{ij}(t)|^2, \text{ with } \rho_{ij} = \frac{P}{\sigma_{z,ij}^2}. \quad (4.2)$$

## 4.2. System Model

In any time-slot, the link state of a user on a channel represents the number of packets that can be successfully transmitted without outage to that user, on that channel. Transmission at a rate above the link state always fails, whereas transmission at a rate below or equal to the link state always succeeds. We use  $C_{ij}(t)$  to represent the state of link  $(ij)$  at time-slot  $t$ . We assume that each link state can take  $K$  possible values  $\{R_1, R_2, \dots, R_K\}$ , where rate  $R_k$  represents state  $k$  and corresponds to the case where  $\tau_k \leq \gamma_{ij}(t) < \tau_{k+1}$ , for some non-negative thresholds  $\tau_k$ . This setting is very used in practice where usually  $K$  modulation and coding schemes are used (i.e. the instantaneous transmitted rate can take only  $K$  possible values). For ease of exposition, we suppose that the  $R_k$  are sorted in a descending order such that  $R_k < R_m$  if  $k > m$ .

Finally, the fading process, which is represented by the coefficients  $h_{ij}(t)$ , is assumed to be *channel convergent* [28]. For channel convergent processes, it is known that the time average fraction of time in each state converges to the steady state distribution, and the expected time average is arbitrarily close to this distribution if sampled over a suitably large interval (of time). Note that in the special case of i.i.d. fading process, the steady state averages are achieved every time-slot.

### 4.2.2 System with Relaying

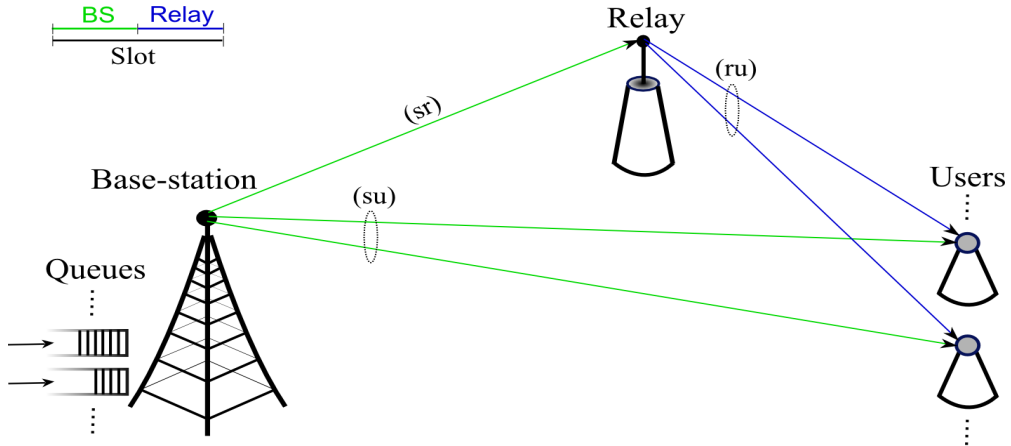


Figure 4.3: System model with relaying.

Under this scenario, we add a relay that can help the BS transmit the information to the users. The users can hear both the BS and the relay transmissions and apply the maximum ratio combining (MRC) of the available information. We assume that the relay operates in half-duplex mode, i.e. the relay cannot receive and transmit simultaneously over the same channel (i.e. frequency band). Moreover, if for a specific channel the relay was selected to help the BS, the overall system is assumed to be orthogonal in time, i.e. the transmission slot is divided in two mini-slots: the corresponding wireless channel is allocated for the BS transmission during the first mini-slot and for the relay transmission during the second mini-slot. Obviously, this orthogonality constraint induces the need for full synchronization among the nodes in the system, which is assumed here.

We now describe the two main relaying strategies (i.e. protocols) normally used in the literature: Decode-and-Forward (DF) and Amplify-and-Forward (AF).

## 4.2. System Model

---

- For DF strategy, the BS transmits to both the receiver and the relay in the first mini-slot. The relay decodes the signal in the first mini-slot. If decoding is successful, the relay re-encodes the decoded data and sends it out in the second mini-slot.
- For AF strategy, the BS transmits to both the receiver and the relay in the first mini-slot. The relay amplifies the received signal in the first mini-slot and sends it out in the second mini-slot.

Under both DF and AF strategies, since the transmission takes place in two mini-slots, the achievable rate should be halved. In the literature, there is a lot of interest in AF because of its simplicity in terms of analysis and its low complexity compared to other relaying strategies. For these reasons, we adopt the AF strategy in our work here.

If at time-slot  $t$  and over channel  $j$  user  $i$  was selected to be served and the relay gets selected to help the BS, the signal corresponding to the BS transmission received at the relay is denoted and given as the following

$$y_{ij}^{(\text{sr})}(t) = \sqrt{P^{(\text{s})}} h_j^{(\text{sr})}(t) x_{ij}(t) + z_j^{(\text{r})}(t), \quad (4.3)$$

and the signal corresponding to the BS transmission received at user  $i$  is denoted and given by

$$y_{ij}^{(\text{su})} = \sqrt{P^{(\text{s})}} h_{ij}^{(\text{su})}(t) x_{ij}(t) + z_{ij}^{(\text{u})}(t). \quad (4.4)$$

In the above equations,  $x_{ij}(t)$  denotes the complex-valued data stream, with  $\mathbb{E}\{|x_{ij}(t)|^2\} = 1$ ,  $P^{(\text{s})}$  is the BS power per channel,  $h_j^{(\text{sr})}(t)$  and  $h_{ij}^{(\text{su})}(t)$  represent the fading coefficients corresponding to the BS-relay and BS-user  $i$  links on channel  $j$  at time-slot  $t$ ; note that these coefficients can capture, in addition to fading, the effects of path-loss. In addition,  $z_j^{(\text{r})}(t)$  and  $z_{ij}^{(\text{u})}(t)$  stand for the noises at the relay and user  $i$  (over channel  $j$ ), respectively, which are assumed to be additive white Gaussian processes, with zero mean and variances  $(\sigma_{z,j}^{(\text{r})})^2$  and  $(\sigma_{z,ij}^{(\text{u})})^2$ , respectively. As alluded earlier, under the AF strategy the relay amplifies (without any further processing) the signal received from the BS such that it fulfills the relay's power constraint per channel, which is represented by  $P^{(\text{r})}$ , and then retransmits the signals towards the user. We assume that at time-slot  $t$  the relay has perfect knowledge of  $h_j^{(\text{sr})}(t)$ ,  $\forall j$ . The signal received at user  $i$  (over channel  $j$ ) corresponding to the relay transmission, which we denote by  $y_{ij}^{(\text{ru})}$ , can be expressed as [126]

$$\begin{aligned} y_{ij}^{(\text{ru})}(t) &= \sqrt{\frac{P^{(\text{r})}}{\mathbb{E}\{|y_{ij}^{(\text{sr})}(t)|^2\}}} h_{ij}^{(\text{ru})}(t) y_{ij}^{(\text{sr})}(t) + z_{ij}^{(\text{u})}(t) \\ &= \sqrt{\frac{P^{(\text{r})}}{P^{(\text{s})}|h_j^{(\text{sr})}(t)|^2 + (\sigma_{z,j}^{(\text{r})})^2}} h_{ij}^{(\text{ru})}(t) \left( \sqrt{P^{(\text{s})}} h_j^{(\text{sr})}(t) x_{ij}(t) + z_j^{(\text{r})}(t) \right) + z_{ij}^{(\text{u})}(t) \\ &= \sqrt{\frac{P^{(\text{r})} P^{(\text{s})}}{P^{(\text{s})}|h_j^{(\text{sr})}(t)|^2 + (\sigma_{z,j}^{(\text{r})})^2}} h_{ij}^{(\text{ru})}(t) h_j^{(\text{sr})}(t) x_{ij}(t) + \\ &\quad \sqrt{\frac{P^{(\text{r})}}{P^{(\text{s})}|h_j^{(\text{sr})}(t)|^2 + (\sigma_{z,j}^{(\text{r})})^2}} h_{ij}^{(\text{ru})}(t) z_j^{(\text{r})}(t) + z_{ij}^{(\text{u})}(t). \end{aligned} \quad (4.5)$$

## 4.2. System Model

---

Note that in the above equation  $h_{ij}^{(\text{ru})}(t)$  represents the fading coefficient corresponding to the relay-user  $i$  link on channel  $j$  at time-slot  $t$ . We assume that at time-slot  $t$  user  $i$  has perfect knowledge of coefficients  $h_{ij}^{(\text{su})}(t)$  and  $h_{ij}^{(\text{ru})}(t)$ ,  $\forall j$ . As in the system without relaying, the fading processes here are also assumed to be *channel convergent* [28].

Depending on whether the use of the relay can provide a gain to the system compared to the case where only the BS is used for transmission, for each user we have two possible schemes:

1. ***The BS serves the user without using the relay.*** In other words, the BS transmits over the direct link during all the slot period. For this scheme, the SNR at user  $i$  and corresponding to channel  $j$  and time-slot  $t$  is defined as follows (see equation (4.4))

$$\gamma_{ij}^{(\text{d})}(t) = \rho_{ij}^{(\text{su})} |h_{ij}^{(\text{su})}(t)|^2, \text{ with } \rho_{ij}^{(\text{su})} = \frac{P^{(\text{s})}}{(\sigma_{z,ij}^{(\text{u})})^2}. \quad (4.6)$$

Let  $C_{ij}^{(\text{d})}(t)$  be the amount of data that the BS can transmit to user  $i$  over channel  $j$  without using the relay at time-slot  $t$ . As in the system without relay, we suppose that  $C_{ij}^{(\text{d})}(t)$  can take  $K$  possible values  $\{R_1, R_2, \dots, R_K\}$ , where we recall that rate  $R_k$  represents state  $k$  and results from the case where  $\tau_k \leq \gamma_{ij}^{(\text{d})}(t) < \tau_{k+1}$ , for some non-negative thresholds  $\tau_k$ .

2. ***The relay helps the BS serving the user.*** Specifically, the BS transmits during the first half of the slot (i.e. first mini-slot) and in the second half (i.e. second mini-slot) the relay transmits. The SNR at user  $i$  that results from the first mini-slot of time-slot  $t$  and that corresponds to channel  $j$  is denoted by  $\gamma_{ij}^{(\text{su})}(t)$ . It can be easily noticed that we have

$$\gamma_{ij}^{(\text{su})}(t) = \gamma_{ij}^{(\text{d})}(t). \quad (4.7)$$

From equation (4.5), the term

$$\sqrt{\frac{P^{(\text{r})}}{P^{(\text{s})} |h_j^{(\text{sr})}(t)|^2 + (\sigma_{z,j}^{(\text{r})})^2}} h_{ij}^{(\text{ru})}(t) z_j^{(\text{r})}(t) + z_{ij}^{(\text{u})}(t) \quad (4.8)$$

is considered as the total noise at user  $i$ .

The SNR at user  $i$  that results from the second mini-slot can be defined as

$$\begin{aligned} \gamma_{ij}^{(\text{sr})}(t) &= \frac{P^{(\text{r})} P^{(\text{s})} |h_{ij}^{(\text{ru})}(t)|^2 |h_j^{(\text{sr})}(t)|^2}{P^{(\text{s})} |h_j^{(\text{sr})}(t)|^2 + (\sigma_{z,j}^{(\text{r})})^2} \left( \frac{P^{(\text{r})} |h_{ij}^{(\text{ru})}(t)|^2 (\sigma_{z,j}^{(\text{r})})^2}{P^{(\text{s})} |h_j^{(\text{sr})}(t)|^2 + (\sigma_{z,j}^{(\text{r})})^2} + (\sigma_{z,ij}^{(\text{u})})^2 \right)^{-1} \\ &= \frac{P^{(\text{r})} P^{(\text{s})} |h_{ij}^{(\text{ru})}(t)|^2 |h_j^{(\text{sr})}(t)|^2}{P^{(\text{r})} |h_{ij}^{(\text{ru})}(t)|^2 (\sigma_{z,j}^{(\text{r})})^2 + P^{(\text{s})} |h_j^{(\text{sr})}(t)|^2 (\sigma_{z,ij}^{(\text{u})})^2 + (\sigma_{z,j}^{(\text{r})})^2 (\sigma_{z,ij}^{(\text{u})})^2} \\ &= \frac{\gamma_{ij}^{(\text{ru})}(t) \gamma_j^{(\text{sr})}(t)}{\gamma_{ij}^{(\text{ru})}(t) + \gamma_j^{(\text{sr})}(t) + 1}, \end{aligned} \quad (4.9)$$

## 4.2. System Model

in which  $\gamma_{ij}^{(\text{ru})}(t)$  and  $\gamma_j^{(\text{sr})}(t)$  are defined as follows

$$\gamma_{ij}^{(\text{ru})}(t) = \rho_{ij}^{(\text{ru})} |h_{ij}^{(\text{ru})}(t)|^2, \text{ with } \rho_{ij}^{(\text{ru})} = \frac{P^{(\text{r})}}{(\sigma_{z,ij}^{(\text{u})})^2}, \quad (4.10)$$

$$\gamma_j^{(\text{sr})}(t) = \rho_j^{(\text{sr})} |h_j^{(\text{sr})}(t)|^2, \text{ with } \rho_j^{(\text{sr})} = \frac{P^{(\text{s})}}{(\sigma_{z,j}^{(\text{r})})^2}. \quad (4.11)$$

Hence, we can define the SNR at user  $i$  and corresponding to channel  $j$  and time-slot  $t$  as

$$\gamma_{ij}^{(\text{r})}(t) = \gamma_{ij}^{(\text{su})}(t) + \gamma_{ij}^{(\text{sr})}(t). \quad (4.12)$$

Let  $C_{ij}^{(\text{r})}(t)$  be the amount of data that the BS and the relay can transmit to user  $i$  over channel  $j$  at time-slot  $t$ . As mentioned earlier, in this case the achievable rate should be halved, so  $C_{ij}^{(\text{r})}(t)$  can take  $K$  possible values  $\{\frac{1}{2}R_1, \frac{1}{2}R_2, \dots, \frac{1}{2}R_K\}$ , where we note that rate  $\frac{1}{2}R_k$  represents state  $k$  and results from the case where  $\tau_k \leq \gamma_{ij}^{(\text{r})}(t) < \tau_{k+1}$ .

We use  $C_{ij}(t)$  to represent the state of link  $(ij)$  at time-slot  $t$ , i.e. the (maximum) amount of data that can be transmitted without outage to user  $i$  on channel  $j$ . Since there are two possible transmission schemes, we choose  $C_{ij}(t)$  to be the maximum amount of data that can be delivered under these schemes, meaning that we have

$$C_{ij}(t) = \max\{C_{ij}^{(\text{d})}(t), C_{ij}^{(\text{r})}(t)\}. \quad (4.13)$$

It is worth recalling that a transmission at a rate above the link state always fails, whereas transmission at a rate below or equal to the link state always succeeds.

### 4.2.3 Limited and Delayed CSI Knowledge Scheme

Before presenting the limited and delayed CSI scheme, we first provide the conditions under which the systems with and without relaying are called *ideal* (in terms of CSI knowledge):

- **System without relaying.** This system is called ideal when at each time-slot the BS has perfect and full knowledge of all the link states at no cost. In other words, at time-slot  $t$  the BS knows (perfectly)  $C_{ij}(t)$ ,  $\forall i, j$ . Note that the term 'perfect' means that each  $C_{ij}(t)$  is known perfectly, whereas the term 'full' refers to the fact that all these link states are known. Under the system without relaying, we recall that at time-slot  $t$  user  $i$  knows perfectly  $h_{ij}(t)$ ,  $\forall j$ , which implies that this user has perfect knowledge of  $C_{ij}(t)$ ,  $\forall j$ . Hence, for the system to be called ideal we should assume that: (a) all the users send their perfect knowledge of their link states to the BS, (b) if the users send these states at time-slot  $t$ , they arrive to the BS also at time-slot  $t$ , and (c) there is no cost for this feedback (i.e. CSI acquisition) process.
- **System with relaying.** This system is called ideal if at each time-slot the BS receives perfect and full knowledge of all the link states at no cost, i.e. at time-slot  $t$  the BS knows all the

## 4.2. System Model

$C_{ij}(t)$  perfectly. Under the system with relaying, we recall that at time-slot  $t$  user  $i$  has perfect knowledge of  $h_{ij}^{(\text{su})}(t)$  and  $h_{ij}^{(\text{ru})}(t)$ ,  $\forall j$ . Using equations (4.12) and (4.13), it can be noticed that  $C_{ij}(t)$  depends on  $h_{ij}^{(\text{su})}(t)$ ,  $h_{ij}^{(\text{ru})}(t)$ , and  $h_j^{(\text{sr})}(t)$ ; recall that we suppose the relay knows  $h_j^{(\text{sr})}(t)$ ,  $\forall j$ . Hence, for the system with relaying to be called ideal we should assume that: (a) user  $i$  knows (in addition to  $h_{ij}^{(\text{su})}(t)$  and  $h_{ij}^{(\text{ru})}(t)$ )  $h_j^{(\text{sr})}(t)$  perfectly, which implies that this user can compute  $C_{ij}(t)$ ,  $\forall j$ , (b) all the users send their perfect knowledge of their link states to the BS, (c) if the users send these states at time-slot  $t$ , they arrive to the BS also at time-slot  $t$ , and (d) there is no cost for this feedback process.

It is worth mentioning that if the users have imperfect knowledge of  $h_j^{(\text{sr})}(t)$ , then this will lead user  $i$  (for e.g.) to possibly mis-estimate its exact corresponding link state (given by  $C_{ij}(t)$ ), and thus the users will (possibly) send an imperfect version of their link states to the BS.

### Limited and Delayed CSI

We now present the *limited* and *delayed* CSI scheme. To this end, we define  $d$  to be the delay in number of slots. We consider a realistic context where the feedback capacity per slot is limited. We assume that at most  $\bar{F}$  link rates can be reported per slot. Therefore, if the system decides to report  $F > \bar{F}$  link rates, a delay of  $d$  slots is required, where  $d = \lceil F(\bar{F})^{-1} \rceil$ . Hence, the amount of feedback resources we allocate and the delay are coupled: the more we allocate feedback resources (i.e. more  $F$ ), the more time is needed to finish the feedback process, which implies more delay in the system. For the system without relaying, we use  $\hat{h}_{ij}(t)$  to denote  $h_{ij}(t-d)$ , which is the fading of link  $(ij)$  at time-slot  $t-d$ . We recall that in each slot each user has a perfect estimation of the  $h_{ij}$ ,  $\forall j$ , at this slot. For example, in time-slot  $t-d$  user  $i$  knows perfectly  $\hat{h}_{ij}(t)$  (i.e.  $h_{ij}(t-d)$ ),  $\forall j$ . The SNR corresponding to  $\hat{h}_{ij}(t)$  is denoted and given by  $\hat{\gamma}_{ij}(t) = \rho_{ij}|\hat{h}_{ij}(t)|^2$ . We use  $\hat{C}_{ij}(t)$  to denote the state of link  $(ij)$  at time-slot  $t-d$ , i.e.  $\hat{C}_{ij}(t) = C_{ij}(t-d)$ . Hence,  $\hat{C}_{ij}(t)$  depends on the value of  $\hat{\gamma}_{ij}(t)$ , e.g.  $\hat{C}_{ij}(t) = R_k$  if  $\tau_k \leq \hat{\gamma}_{ij}(t) < \tau_{k+1}$ . For the system with relaying, let  $\hat{h}_{ij}^{(\text{su})}(t) = h_{ij}^{(\text{su})}(t-d)$ ,  $\hat{h}_{ij}^{(\text{ru})}(t) = h_{ij}^{(\text{ru})}(t-d)$ , and  $\hat{h}_j^{(\text{sr})}(t) = h_j^{(\text{sr})}(t-d)$ . In addition, let  $\hat{\gamma}_{ij}^{(\cdot)}(t) = \gamma_{ij}^{(\cdot)}(t-d)$ , where the notation  $(\cdot)$  is used to represent the different SNRs (at the users), and  $\hat{C}_{ij}(t) = C_{ij}(t-d)$ .

It is worth recalling that at a given time-slot user  $i$  has a perfect estimation of  $h_{ij}^{(\text{su})}$  and  $h_{ij}^{(\text{ru})}$ ,  $\forall j$ , of this slot. However, we stress out that, until this stage, no specific assumption is made on the knowledge of  $h_j^{(\text{sr})}(t)$  at the users; note that such an assumption will be made later.

### 4.2.4 Queueing Model

We here present the queueing model (i.e. queue dynamics) for the ideal system as well as for the system with limited and delayed feedback. To this end, we consider that each of the  $N$  users has an incoming traffic process  $A_i(t)$ , which is an integer-valued process, measured in bits, i.i.d. (independent and identically distributed) in time and independent across users, with  $A_i(t) < A_{\max}$ , for some finite constant  $A_{\max}$ . The mean rate of this process is  $a_i = \mathbb{E}\{A_i(t)\}$ . We assume that packets arrive at the BS at the beginning of the time-slot and are served at the end of the time-slot.

We also assume that a transmission over link  $(ij)$  can only be fulfilled if the corresponding link state (i.e.  $C_{ij}(t)$  in the ideal case for ex) is reported to the BS.



#### 4.2.4.1 Queue Dynamics under the Ideal System

We first note that the model that will be provided here can be used under the system without relaying as well as the system with relaying. To begin with, we define  $S_{ij}(t)$  to be the scheduling decision associated with link  $(ij)$  at time-slot  $t$ . So, we have

$$S_{ij}(t) = \begin{cases} 1, & \text{if user } i \text{ gets scheduled on channel } j \\ 0, & \text{otherwise.} \end{cases} \quad (4.14)$$

As mentioned earlier, in one time-slot and on a given channel one and at most one user can be scheduled. As a result, for all  $t$  and all  $j$ , any valid service policy must obey  $\sum_{i=1}^N S_{ij}(t) \leq 1$ . It should be pointed out that under the ideal system there is no need to define a feedback decision since all the link states are reported (without delay and at no cost).

Based on the above, the queues in the system evolve according to the following equation

$$q_i(t+1) = \max\{q_i(t) + A_i(t) - \sum_{j=1}^L C_{ij}(t)S_{ij}(t), 0\}, \text{ for } 1 \leq i \leq N, \quad (4.15)$$

where  $\sum_{j=1}^L C_{ij}(t)S_{ij}(t)$  represents the service rate allocated for user  $i$  at time-slot  $t$ .

#### 4.2.4.2 Queue Dynamics under the System with Delayed and Limited CSI

Also here the model can be used under the system with or without relaying. Let us define  $S_{ij}(t)$  to be the scheduling decision associated with link  $(ij)$  at time-slot  $t$ . In addition, we define  $Y_{ij}(t)$  as the feedback decision associated with link  $(ij)$  at time-slot  $t$ .

We draw the attention to the fact that the delay  $d$  is the required period to report the CSI for the users selected for the feedback. Then, the period between the feedback decision and the scheduling mechanism will not be always equal to  $d$  slots. Specifically, the scheduling decisions of, for ex, time-slots  $t_1, t_1 + 1, \dots, t_1 + d - 1$  will be based on the feedback decision of time-slot  $t_1 - d$ , where we note that this feedback arrives at the BS at time-slot  $t_1$ . To simplify the notation and the presentation of the analysis, we denote by  $t$  the slots at which the scheduling is done and by  $t - d$  the slot at which the feedback decision is done. That is to say, a fixed delay value is considered between the feedback decision and the scheduling time. This delay value could correspond to the worst case delay, average delay or best case delay. It is worth noting that the theoretical analysis conducted in this paper still holds for all of those cases.

Let  $\hat{Y}_{ij}(t) = Y_{ij}(t - d)$  denote the feedback decision associated with link  $(ij)$  at time-slot  $t - d$ . So, we can write

$$\hat{Y}_{ij}(t) = \begin{cases} 1, & \text{if } \hat{C}_{ij}(t) \text{ gets reported to the scheduler} \\ 0, & \text{otherwise,} \end{cases} \quad (4.16)$$

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

where we recall that  $\hat{C}_{ij}(t) = C_{ij}(t - d)$ . On the other side, for  $S_{ij}(t)$ , which represents the scheduling decision at time-slot  $t$ , we have

$$S_{ij}(t) = \begin{cases} 1, & \text{if user } i \text{ gets scheduled on channel } j \\ 0, & \text{otherwise.} \end{cases} \quad (4.17)$$

The queueing dynamics are then given as follows

$$q_i(t+1) = \max\{q_i(t) + A_i(t) - \sum_{j=1}^L \hat{C}_{ij}(t) \hat{Y}_{ij}(t) S_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))}, 0\}, \text{ for } 1 \leq i \leq N, \quad (4.18)$$

where  $\mathbb{1}_{(\cdot)}$  is the indicator function and the expression

$$\sum_{j=1}^L \hat{C}_{ij}(t) \hat{Y}_{ij}(t) S_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \quad (4.19)$$

is nothing but the service rate allocated for user  $i$  at time-slot  $t$ . Note that the use of indicator function  $\mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))}$  is necessary due to a possible mismatch between the reported rate,  $\hat{C}_{ij}(t)$ , and the actual link state,  $C_{ij}(t)$ , meaning that it is possible that the reported rate is greater than the actual link rate, thus leading to outage.

## 4.3 Proposed Algorithms and Stability Analysis for the System without Relaying

In this section, we provide a stability analysis for the system without relaying. Specifically, we propose a joint feedback allocation and scheduling algorithm, named FSA, under which the allocations are done at the users side. We derive the fraction that this algorithm reaches of the stability region of the ideal system. Although it provides good stability performance, this algorithm is adapted only for continuous-time contention schemes. For this reason, we propose a second algorithm, termed as FMA, that is adapted for discrete-time contention schemes and that provides a good imitation of FSA. For FMA, an amount of  $F$  feedback resources will be available, i.e. at most  $F$  links can report their CSI to the BS. Note that since we consider a bursty traffic, the metric we use to evaluate the performance is the stability of the queues. It is worth mentioning that the scheduling mechanisms under the different schemes are mainly based on the Max-Weight policy, which was defined in Chapter 2, where we can also find the definitions of stability region and optimal scheduling policy.

Before proceeding in the presentation, in the following we discuss the importance of finding a good trade-off between the amount of feedback resources and the delay. If more link states are reported to the BS, the Max-Weight has more CSIs that can be used in the decision, which improves the performance of the scheduling. On the other hand, a delay occurs between the estimation time of the CSIs and the scheduling decision time, and this delay increases with the number of reported link states, which results in performance reduction. The trade-off between the number of reported CSIs

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

---

and the delay is a challenging problem that is considered in the numerical results.

In order for the BS to perform scheduling, it needs to receive the rates of the links that were selected for the feedback. The feedback allocation process is thus of great importance since it directly impacts the scheduling mechanism. Hence, a necessary step to conduct the stability analysis is to specify/develop a joint feedback allocation and scheduling algorithm and to characterize its performance w.r.t. the ideal system where full and perfect CSI is available at no cost. As alluded earlier, the second algorithm we propose, FMA, is designed in such a way as to greatly imitate another algorithm we develop, FSA, that provides good stability performance but that is challenging to implement. Obviously, under the ideal system there is no need for a feedback allocation algorithm since it is supposed that all the link states are instantaneously reported to the BS (at not cost).

We now provide the motivation behind the reasoning regarding the feedback decision used in algorithms FSA and FMA (which will be presented afterwards).

**Motivation.** We first note that a Max-Weight-like rule is adopted for the scheduling mechanism, under which the decisions are affected by the reported CSIs as well as the queue lengths; i.e. the feedback allocation algorithm directly impacts the scheduling mechanism. One could consider that the BS must decide of the feedback allocation. However, letting the users make this decision provides a gain due the following: if the BS decides which user should feed back its CSI, this decision will be based on the channel statistics. In other words, the BS can ask a user with bad CSI to report its feedback, since the BS cannot know beforehand if the current CSI is good or bad. However, the users estimate their CSI at each time-slot and therefore have perfect knowledge of their current CSIs, so including them in the feedback process would enhance the system performance. On the other hand, the users lack the queue lengths information, which should be taken into consideration by the feedback allocation since the scheduling mechanism is directly impacted by the feedback decisions. In general, the stability of the system is more sensitive to the channels variation than to the queues variation, so we can provide the users with the queue lengths knowledge every period of time and not every slot. This is done by the BS which broadcasts the queue lengths information every  $T_b$  slots, where  $T_b$  is typically high so that the (signalization) cost of such a broadcast stays negligible.

#### 4.3.1 Algorithm FSA and its Stability Performance

Let 'mdl' represent the system where over each channel only one user reports its feedback (and then uses this feedback to transmit after  $d$  slots) and where this user is selected using algorithm FSA. We use 'pf' to denote the ideal system, i.e. perfect and full CSI is available at the BS, where Max-Weight policy is used to schedule users for transmission. Let  $\mathcal{M}^{\text{pf}}(t)$  and  $\mathcal{M}^{\text{mdl}}(t)$  be the subsets of scheduled links at time-slot  $t$  under 'pf' and 'mdl', respectively. Note that here the delay  $d = \lceil L(\bar{F})^{-1} \rceil$  since FSA uses an amount of feedback resources equal to the number of channels,  $L$ .

Before presenting algorithm FSA, and to better understand the reasoning that is used in this algorithm, we next provide the Max-Weight scheduling used under the *ideal* system. First, recall that under this system the BS has full and perfect knowledge (at no cost) of all the link states. That is to say, at time-slot  $t$  the BS knows  $C_{ij}(t)$ ,  $\forall i, j$ . Thus, at time-slot  $t$  the Max-Weight policy schedules

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

over channel  $j$  the user that results from the following operation

$$\arg \max_i \{q_i(t)C_{ij}(t)\}. \quad (4.20)$$

For the ideal system, we define  $g_{\text{pf}}$  to be the *expected weighted throughput*, which is given as follows

$$g_{\text{pf}}(\mathbf{q}(t)) = \mathbb{E} \left\{ \sum_{j=1}^L \sum_{i=1}^N q_i(t)C_{ij}(t)S_{ij}(t) \mid \mathbf{q}(t) \right\}, \quad (4.21)$$

where here  $S_{ij}(t)$  represents the scheduling decision of link  $(ij)$  at time-slot  $t$  (under the ideal system), which must obey  $\sum_{i=1}^N S_{ij}(t) \leq 1$ . An alternative way to write  $g_{\text{pf}}$  is

$$g_{\text{pf}}(\mathbf{q}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{pf}}(t)} q_i(t)C_{ij}(t) \mid \mathbf{q}(t) \right\}. \quad (4.22)$$

---

**Algorithm FSA:** For system 'mdl' the feedback and scheduling decisions are based on the following procedure.

1. *Queue lengths broadcast every  $T_b$  slots:*

Every  $T_b$  time-slots, that is to say, at time  $0, T_b, 2T_b, \dots, nT_b, \dots$ , the BS broadcasts the queue lengths of all users, where  $T_b$  is typically high. So each user has outdated knowledge of the state of its queue (and all other queues). Let  $\tilde{q}_i(t)$  be the (outdated) queue length the users know at time  $t$ , i.e.  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b[$ .

2. *Feedback and scheduling decisions at time-slot  $t - d$ :*

To simplify the notation, we denote  $\tilde{q}_i(t - d)$  as  $\hat{q}_i(t)$ .

For each channel  $j \in \{1, \dots, L\}$ , only one user sends its CSI to the BS. This user is the result of the Max-Weight rule [19], [56] and can be given by

$$\arg \max_i \left\{ \hat{q}_i(t)\hat{C}_{ij}(t)\mathbb{P} \left\{ C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t) \right\} \right\}. \quad (4.23)$$

To detect this user, we use an approach that consists in letting the users contend among each other for a certain period of time. Let  $T_c$  be the contention period for each channel. Assuming that contention can be done in continuous time [68], meaning that there is no collision between users, for each channel  $j \in \{1, \dots, L\}$  the contention is performed as the following:

User  $i$ ,  $\forall i \in \{1, \dots, N\}$ , waits until time

$$T_c \left( \hat{q}_i(t)\hat{C}_{ij}(t)\mathbb{P} \left\{ C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t) \right\} \right)^{-1}, \quad (4.24)$$

then broadcasts a signal (of negligible duration). It is obvious that the user that broadcasts its signal first will be the result of (4.23) since this user will wait the smallest period of time. After the broadcast of the first signal, the contention procedure (of channel  $j$ ) terminates and the corresponding user reports its CSI.

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

---

Once this procedure finishes, the contention of another channel gets started; clearly, two channels cannot have their contentions over the same period of time.

#### 3. *Transmission at time-slot $t$ :*

After getting the CSI (of time-slot  $t-d$ ) of each channel, the BS uses these CSIs for transmission at time-slot  $t$ . It is obvious that the user that is selected to report its feedback (at time-slot  $t-d$ ) of channel  $j$  will also be the user transmitting over this channel (at time-slot  $t$ ).

---

The main idea behind algorithm FSA is to approach the scheduling in the ideal system, while noting that the feedback decision is done by the users; the motivation behind letting the users be implicated in the decision was provided earlier. For each channel, this algorithm lets the users contend among each other in order to determine the user that will report its state, and this user will also be scheduled for transmission on the corresponding channel. As indicated in the algorithm, for channel  $j$  the user that is selected for feedback (and consequently for transmission on channel  $j$ ), which results from the contention procedure, is the one that maximizes the expression

$$\hat{q}_i(t)\hat{C}_{ij}(t)\mathbb{P}\left\{C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t)\right\}. \quad (4.25)$$

The choice to maximize the above quantity can be explained as follows. First, recall that under the ideal system on channel  $j$  the user that is scheduled for transmission is the one that maximizes  $q_i(t)C_{ij}(t)$  (using the Max-Weight concept). Under system 'mdl', and accounting for delay  $d$ , the feedback decisions corresponding to the transmission of time-slot  $t$  are made at time-slot  $t-d$ . The available information at user  $i$  at time-slot  $t-d$  includes all the queue lengths and in particular  $\hat{q}_i(t)$ , where we recall that  $\hat{q}_i(t) = \tilde{q}(t-d)$ , and  $\hat{C}_{ij}(t)$  (which results from  $\hat{h}_{ij}(t)$ ),  $\forall j$ , with  $\hat{C}_{ij}(t) = C_{ij}(t-d)$ . Based on the above and taking into account the outage possibility, which occurs when  $\hat{C}_{ij}(t) > C_{ij}(t)$ , it can be noticed that the best way to approximate the ideal scheduling on channel  $j$  is by selecting the user that maximizes the quantity given in (4.25).

### Stability Performance Analysis for FSA

As mentioned earlier, here delay  $d = \lceil L(\bar{F})^{-1} \rceil$  since FSA uses  $L$  feedback resources. However, for FMA we have  $d = \lceil F(\bar{F})^{-1} \rceil$  because under this algorithm an amount of  $F$  feedback resources is used. The same notation of delay,  $d$ , is used in both algorithms for the sake of notational simplicity. As a result of algorithm FSA, the expected weighted throughput, termed as  $g_{\text{mdl}}$ , at time-slot  $t$  is defined as follows

$$g_{\text{mdl}}(\hat{\mathbf{q}}(t)) = \mathbb{E}\left\{\sum_{j=1}^L \sum_{i=1}^N \hat{q}_i(t)\hat{C}_{ij}(t)\hat{Y}_{ij}(t)S_{ij}(t)\mathbf{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t)\right\}, \quad (4.26)$$

in which  $S_{ij}(t)$  represents the scheduling decision of link  $(ij)$  at time-slot  $t$ , so here  $S_{ij}(t) = \hat{Y}_{ij}(t)$  and is equal to 1 if user  $i$  is selected (at time  $t-d$ ) by algorithm FSA to report its CSI (and consequently

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

to transmit at time-slot  $t$ ) and 0 otherwise. Recall that over each channel one and at most one channel is scheduled for transmission, i.e.

$$\sum_{i=1}^N S_{ij}(t) \leq 1. \quad (4.27)$$

We stress out that the indicator function  $\mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))}$  is used because of a possible mismatch between the reported link state  $\hat{C}_{ij}(t)$  and the actual link state  $C_{ij}(t)$ , i.e. it is possible that  $\hat{C}_{ij}(t) > C_{ij}(t)$  which leads to an outage since a transmission at a rate strictly greater than the actual link state always fails.

An alternative way to express  $g_{\text{mdl}}$  is by using the definition of  $\mathcal{M}^{\text{mdl}}(t)$  as

$$g_{\text{mdl}}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mdl}}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t) \right\}. \quad (4.28)$$

Next, we want to investigate the stability performance of system 'mdl' by characterizing the minimum fraction that the stability region of algorithm FSA can achieve w.r.t. the stability region of the ideal system (i.e. 'pf'); clearly, this fraction is lower than 1. To this end, we define  $\eta$  as

$$\eta = \max_{t, (ij)} \left\{ \frac{C_{ij}(t)}{\hat{C}_{ij}(t)} \right\}. \quad (4.29)$$

We here assume that all the possible values of  $C_{ij}(t)$  (and thus of  $\hat{C}_{ij}(t)$ ), given by  $R_1, \dots, R_K$ , are different than zero. In practice, it is very rare not to be able to transmit at a non-zero rate. Moreover, we define  $p_c^{\min}$  as follows

$$p_c^{\min} = \min_{(ij)} p_{cij}^{\min}, \quad (4.30)$$

in which  $p_{cij}^{\min}$  is given by

$$p_{cij}^{\min} = \min_{t, \hat{h}_{ij}(t)} \left\{ \mathbb{P}\{C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t)\} \right\}, \quad (4.31)$$

where obviously the value of  $\hat{C}_{ij}(t)$  depends on  $\hat{h}_{ij}(t)$ . It is worth recalling that under system 'mdl' the scheduling decision at time-slot  $t$  is based on the feedback decision made at time  $t - d$ . Also, recall that the BS broadcasts the queue lengths to the users each  $T_b$  slots.

Based on the above, we now provide the stability region that system 'mdl' can achieve compared with the stability region of the ideal system.

**Theorem 14.** *Algorithm FSA can achieve at least a fraction  $\beta$  of the stability region achieved by the ideal system (with perfect and full feedback), where*

$$\beta = \left(1 - \frac{1}{T_b}\right) \frac{p_c^{\min}}{\eta}. \quad (4.32)$$

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

*Proof.* We first provide an important result that will be useful for the rest of the proof. For this purpose, we define 'pf<sub>1</sub>' to be the system where the BS has perfect and full knowledge of the link states and then applies the Max-Weight rule for scheduling at time-slot  $t$  using  $\hat{\mathbf{q}}(t)$ . Recall that

$$\hat{q}_i(t) = \tilde{q}_i(t - d), \text{ where } \tilde{q}_i(t) = q_i(nT_b) \text{ for } t \in [nT_b, (n+1)T_b[, \quad (4.33)$$

and that here  $d = \lceil L(\bar{F})^{-1} \rceil$ . We stress out that the only difference between 'pf<sub>1</sub>' and 'pf' (i.e. ideal system) is the fact that at time-slot  $t$  the scheduling process (at the BS) under 'pf<sub>1</sub>' will be done based on the queue lengths information  $\hat{\mathbf{q}}(t)$ , whereas under 'pf' the scheduling is based on  $\mathbf{q}(t)$ . Let  $\Lambda_{\text{pf}}$  and  $\Lambda_{\text{pf}_1}$  be the stability regions achieved under systems 'pf' and 'pf<sub>1</sub>', respectively. Based on the above, the following result can be stated.

**Lemma 7.** *As long as  $T_b$  is finite, system 'pf<sub>1</sub>' is strongly stable if and only if system 'pf' is strongly stable. That is to say, the following stability region result holds*

$$\Lambda_{\text{pf}_1} = \Lambda_{\text{pf}}. \quad (4.34)$$

From the above statement, it can be deduced that even if the scheduling is done based on an infrequent (i.e. outdated) queue lengths information, the stability of the system can still be achieved as long as the difference between the estimate (i.e. outdated value) and the exact value is bounded by a constant. It can be seen that this latter condition is satisfied for system 'pf<sub>1</sub>' since  $T_b$  is finite and the maximum per-time-slot length change in any queue is bounded (because the maximum arrival and service rates are bounded).

Please refer to Appendix 4.7.1 for the complete proof of the theorem, which also contains the proof of Lemma 7. □

If we denote  $\Lambda_{\text{pf}}$  and  $\Lambda_{\text{mdl}}$  as the stability regions of systems 'pf' and 'mdl', respectively, the above theorem implies that  $\Lambda_{\text{mdl}}$  can be bounded as

$$\beta \Lambda_{\text{pf}} \subseteq \Lambda_{\text{mdl}} \subseteq \Lambda_{\text{pf}}. \quad (4.35)$$

That is to say, fraction  $\beta$  just represents a lower bound on the performance guarantee of FSA, meaning that this algorithm may deliver better stability performance than that guaranteed by the lower bound. Based on this observation and depending on the setting under consideration, we can state that FSA generally provides good stability performance. However, implementing this algorithm is a complicated task due to the assumption that the contention can be done in continuous time. This assumption is not realistic since in practice the contention is done in discrete time, leading to a possible collision between the users, and this problem is not handled in FSA. Thus, we next propose another algorithm that adopts a different approach and that provides a good imitation of FSA, and therefore good stability performance.

### 4.3.2 Algorithm FMA and its Stability Performance

Let 'dl' represent the system with delayed and limited feedback, and where the feedback and scheduling decisions are determined by algorithm FMA. Under FMA, the feedback allocation is done at the users side, and an amount of  $F$  feedback resources is available. Specifically, under this algorithm, at time-slot  $t - d$  the users decide to report at most  $F$  link states; typically,  $F$  should be greater than  $L$ . Let us denote the set of these links by  $\hat{\mathcal{F}}(t)$ , where the cardinality  $|\hat{\mathcal{F}}(t)| \leq F$ . Then, due to the delay of  $d$  slots, the BS receives the totality of this feedback at time-slot  $t$ . We recall that here  $d = \lceil F(\bar{F})^{-1} \rceil$ . Let us define  $F_i(t - d)$  to be the number of link states user  $i$  decides at time  $t - d$  to report under FMA, so we have  $\sum_{i=1}^N F_i(t - d) \leq F$ . Also, we define  $\alpha(t)$  to be some threshold that the system updates with the time; the update process will be provided later. To simplify the notation, we let  $\hat{F}_i(t)$  and  $\hat{\alpha}(t)$  represent  $F_i(t - d)$  and  $\alpha_i(t - d)$ , respectively. Algorithm FMA is based on the idea of combining the Max-Weight and threshold-based concepts [56], [127]. It is worth mentioning that the steps in FMA are different from what is proposed in these latter works.

---

**Algorithm FMA:** For system 'dl', the feedback and scheduling decisions are based on the following procedure.

1. *Queue lengths broadcast every  $T_b$  slots:*

Every  $T_b$  time-slots, that is to say, at time  $0, T_b, 2T_b, \dots, nT_b, \dots$ , the BS broadcasts the queue lengths of all users, where  $T_b$  is typically high. So each user has outdated knowledge of the state of its queue (and all other queues). Let  $\tilde{q}_i(t)$  be the (outdated) queue length the users know at time  $t$ , i.e.  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b[$ .

2. *Queue lengths sorting at each user:*

Each user knows then the queues of all other users and sorts all the queue lengths (including its queue) in a descending order. Let  $i_m$  be the index of the user at the  $m$ th position; e.g.  $i_1$  is the index of the user with the largest queue length. A tie is broken by picking the user with the smallest index. We define  $k(i_m j)$  as the state of link  $(i_m j)$ , so  $\hat{C}_{i_m j}(t) = R_{k(i_m j)}$ , where  $k(i_m j) \in \{1, \dots, K\}$ .

3. *Feedback decisions at time-slot  $t - d$ :*

To simplify the notation, we denote  $\tilde{q}_i(t - d)$  as  $\hat{q}_i(t)$ . Set  $\hat{F}_{i_m}(t) = 0, \forall i_m \in \{1, \dots, N\}$ .

For  $m = 1$ , which yields index  $i_1$  and thus corresponds to the user with the largest queue length, the allocation process starts as the following:

(a) For all  $j \in \{1, \dots, L\}$ :

- If there are enough feedback resources, i.e. if  $\sum_{i_m=1}^N \hat{F}_{i_m}(t) < F$ :
  - \* If channel  $j$  satisfies the following inequality

$$\hat{q}_{i_m}(t) \hat{C}_{i_m j}(t) \mathbb{P} \left\{ C_{i_m j}(t) \geq \hat{C}_{i_m j}(t) \mid \hat{h}_{i_m j}(t) \right\} \geq \hat{\alpha}(t), \quad (4.36)$$

then  $\hat{C}_{i_m j}(t)$  will be reported to the BS and  $\hat{F}_{i_m}(t) = \hat{F}_{i_m}(t) + 1$ .



### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

---

\* Else,  $\hat{C}_{ij}(t)$  will not be reported.

- Otherwise, the allocation process stops since no feedback resources are available.

- (b) At this step user  $i_m$  has just finished its part of the algorithm, so he broadcasts a special symbol to inform the BS that he finished the feedback allocation. Now, if the condition  $\sum_{i_m=1}^N \hat{F}_{i_m}(t) < F$  is still satisfied and  $m < N$ , increment  $m$  by 1, and go to Step (a). Otherwise, the allocation process stops.

#### 4. Scheduling decisions and transmission at time-slot $t$ :

At time-slot  $t$ , the BS receives the link states that were selected at time-slot  $t - d$ . Then, for each channel, the BS applies the Max-Weight rule for scheduling. Specifically, over channel  $j$  the BS schedules the user that results from the following

$$\arg \max_{i: (ij) \in \hat{\mathcal{F}}(t)} \{q_i(t) \hat{C}_{ij}(t)\}, \quad (4.37)$$

where we recall that  $\hat{\mathcal{F}}(t)$  is the set of reported link states and that if for some channel no feedback is reported, then no transmission can take place over this channel.

---

Roughly speaking, the idea behind the feedback decision approach in the proposed algorithm here is to allocate the feedback resources to the links which are more likely to be scheduled under the Max-Weight policy if all the link states are available at the BS. Hence, when the BS applies the Max-Weight rule for a given channel based on the subset of reported links states (resulting from the feedback decision under FMA), it will be very unlikely not to schedule the right user (which is the result of scheduling under a full CSI knowledge scenario).

### Stability Performance Analysis for FMA

We now investigate the stability performance of algorithm FMA. As mentioned earlier, this algorithm is proposed to imitate FSA which in its turn approaches the decisions of the ideal system. To analytically illustrate the good stability performance that FMA yields, we show that this algorithm imitates FSA to a great extent. For this purpose, we first discuss the conditions that an algorithm should guarantee in order to provide a good imitation of FSA.

In order to design an algorithm that greatly imitates FSA, two essential points should be accounted for. The first point is to ensure that over channel  $j$  the user that will be scheduled is the result of

$$\arg \max_i \left\{ \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{P} \left\{ C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t) \right\} \right\}. \quad (4.38)$$

The second point is to make sure that a transmission occurs over each channel, which can be ensured by having at least one reported link state for each channel. This is necessary since FSA guarantees a reported link state for each channel, leading to a transmission over all the channels. Based on these two conditions, the following remark can be made.

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

**Remark 3.** *Algorithm FMA can imitate algorithm FSA to a great extent. Hence, FMA, as FSA, generally provides good stability performance.*

The motivation behind the statement in Remark 3 is provided as follows. By allocating a sufficient amount of feedback resources,  $F$ , with  $F$  sufficiently large, and using condition (4.36) which ensures the selection (for the feedback) of links with large value of

$$\hat{q}_i(t)\hat{C}_{ij}(t)\mathbb{P}\left\{C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t)\right\}, \quad (4.39)$$

we can claim that algorithm FMA guarantees the first point (to ensure a good imitation of FSA) with high probability. On the other side, by a judicious choice of threshold  $\hat{\alpha}(t)$  (or  $\alpha(t)$ ), the second point (for a good imitation of FSA) can be guaranteed, that is to say, this threshold can be updated (with time) in such a way as to ensure that the probability of any one of the channels to have zero feedback is very low. We next provide how the threshold value can be updated.

#### Threshold Update

We consider an approach under which the BS updates the threshold each  $T_u$  slots (e.g. at slots  $T_u, 2T_u, 3T_u, \dots$ ) and then broadcasts it to the users, where the choice of  $T_u$  depends on the system under consideration. For notational convenience, suppose that  $t_u - d$  is the time-slot at which the BS has to calculate the threshold, i.e.  $t_u - d = nT_u$  for some non-negative integer  $n$ . Hence, at time-slot  $t_u - d$  the BS calculates threshold  $\hat{\alpha}(t_u)$ , where we recall that  $\hat{\alpha}(t_u) = \alpha(t_u - d)$ . Note that at this slot the BS has complete knowledge of the queue lengths of this slot but knows nothing about the link states of this slot; recall that these lengths and states are denoted by  $\hat{q}_{ij}(t_u)$  and  $\hat{C}_{ij}(t_u)$ , respectively. Further, a reasonable assumption that can be made here is that the BS knows the probabilities  $\mathbb{P}\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{C}_{ij}(t_u) = R_k\}$ . As indicated earlier, the threshold should ensure that it is very likely that every channel gets at least one feedback, or equivalently, that it is very unlikely that a channel gets zero feedback.

Before proceeding further, we here explain a simple but important property about *necessary conditions*, which will be useful for the analysis. Let  $\mathcal{V}$  be an event defined as the intersection of two other events, namely  $\mathcal{V}_1$  and  $\mathcal{V}_2$ . Event  $\mathcal{V}_1$  can be seen as a necessary condition for event  $\mathcal{V}$  to happen; the same remark can be made for  $\mathcal{V}_2$ . It is plain to see that  $\mathbb{P}\{\mathcal{V}\} \leq \mathbb{P}\{\mathcal{V}_1\}$ . Hence, if event  $\mathcal{V}_1$  is unlikely to happen, then this implies that  $\mathcal{V}$  is also unlikely to happen.

Under algorithm FMA, a necessary condition for link  $(ij)$  to get selected for the feedback at time-slot  $t_u - d$  is to have (see (4.36))

$$\hat{q}_i(t_u)\hat{C}_{ij}(t_u)\mathbb{P}\left\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{h}_{ij}(t_u)\right\} \geq \hat{\alpha}(t_u). \quad (4.40)$$

Based on this observation and the available information at the BS, the necessary condition for channel  $j$  to get at least one feedback can be approximated by the BS as

$$\sum_i \sum_{k: \hat{q}_i(t_u)R_k \mathbb{P}\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{C}_{ij}(t_u)=R_k\} \geq \hat{\alpha}(t_u)} p_{ijk} \geq 1, \quad (4.41)$$

### 4.3. Proposed Algorithms and Stability Analysis for the System without Relaying

where  $p_{ijk}$  is the probability that link  $(ij)$  is at state  $k$ , which is supposed to be known by the BS. Define  $\Omega_{ijk}$  and  $\omega_j$  as

$$\Omega_{ijk} = \hat{q}_i(t_u) R_k \mathbb{P}\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{C}_{ij}(t_u) = R_k\}, \quad (4.42)$$

$$\omega_j = \sum_i \sum_{k: \Omega_{ijk} \geq \hat{\alpha}(t_u)} p_{ijk}. \quad (4.43)$$

Hence, a necessary condition for every channel to get at least one feedback can be stated as

$$\omega_j \geq 1, \forall j. \quad (4.44)$$

The following constraint should be added when calculating  $\hat{\alpha}(t_u)$

$$\sum_j \omega_j \leq F \text{ and as close as possible to } F. \quad (4.45)$$

The conditions in (4.45) ensure that the average number of links that are eligible to report their states (see (4.40)), which is given by  $\sum_j \omega_j$ , is pushed as close as possible to  $F$ . This is important due to the following reasons: (a) at most  $F$  (with  $F > L$ ) link states can be reported, due to an amount of feedback resources equal to  $F$ , (b) the more the number of links that are eligible for the feedback is greater than  $F$ , the more likely it is that algorithm FMA does not report the best links in terms of maximizing the product  $\hat{q}_i(t_u) \hat{C}_{ij}(t_u) \mathbb{P}\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{h}_{ij}(t_u)\}$ , and (c) the more the number of links that are eligible for the feedback is lower than  $F$ , the less the feedback resources are exploited efficiently, which may lead to one or multiple channels with zero feedback. Note that the lower the threshold is, the greater  $\omega_j$  and  $\sum_j \omega_j$  will be.

Based on all the above, the following simple procedure can be adopted to compute  $\hat{\alpha}(t_u)$ :

1. The BS chooses any threshold value such that, using this threshold, the condition  $\omega_j \geq 1$  is not satisfied for all  $j$ . Denote this value by  $\hat{\alpha}_0(t_u)$ .
2. Then, starting from  $\hat{\alpha}_0(t_u)$ , it decreases the threshold value until the condition  $\omega_j \geq 1$  is satisfied for all  $j$ . Denote the threshold at this stage by  $\hat{\alpha}_1(t_u)$ .
3. If using  $\hat{\alpha}_1(t_u)$  we have  $\sum_j \omega_j < F$ , then the BS decreases the threshold value (starting from  $\hat{\alpha}_1(t_u)$ ) until having  $\sum_j \omega_j$  as close as possible to  $F$  and such that  $\sum_j \omega_j \leq F$ .

Finally, we point out that if the BS knows

$$\mathbb{E}_{\hat{h}_{ij}(t_u)} \left\{ \mathbb{P}\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{C}_{ij}(t_u) = R_k, \hat{h}_{ij}(t_u)\} \right\}, \quad (4.46)$$

where we recall that  $\hat{h}_{ij}(t_u) = h_{ij}(t_u - d)$ , then, clearly, it is better to define  $\Omega_{ijk}$  as

$$\Omega_{ijk} = \hat{q}_i(t_u) R_k \mathbb{E}_{\hat{h}_{ij}(t_u)} \left\{ \mathbb{P}\{C_{ij}(t_u) \geq \hat{C}_{ij}(t_u) \mid \hat{C}_{ij}(t_u) = R_k, \hat{h}_{ij}(t_u)\} \right\}. \quad (4.47)$$

Taking this new definition into account, the same procedure as before can be used to find  $\hat{\alpha}(t_u)$ .

## 4.4 Proposed Algorithms and Stability Analysis for the System with Relaying

In this section, the stability analysis for the system with relaying is provided; the detailed model of this system can be found in Section 4.2. We consider a scheme that assumes the users have imperfect knowledge of  $h_j^{(\text{sr})}(t)$  (which represents the fading coefficient between the BS and the relay under channel  $j$ ),  $\forall j$ ; recall that  $h_{ij}^{(\text{su})}(t)$  and  $h_{ij}^{(\text{ru})}(t)$ ,  $\forall j$ , are perfectly known at user  $i$  at time-slot  $t$ . For clarity of exposition, we first analyze the case without delay in the feedback process, then we present the analysis for the case where the delay is accounted for.

### CSI Imperfection

The imperfection concerning the knowledge of  $h_j^{(\text{sr})}(t)$  at the users is detailed as follows. At time-slot  $t$  we assume that the only information the users know about  $h_j^{(\text{sr})}(t)$  is if  $\gamma_j^{(\text{sr})}(t) \geq \delta$  or  $\gamma_j^{(\text{sr})}(t) < \delta$ , for some threshold  $\delta > 0$ , where we recall that

$$\gamma_j^{(\text{sr})}(t) = \rho_j^{(\text{sr})} |h_j^{(\text{sr})}(t)|^2, \text{ with } \rho_j^{(\text{sr})} = \frac{P^{(\text{s})}}{(\sigma_{\text{z},j}^{(\text{r})})^2}. \quad (4.48)$$

Obviously, it is the relay who broadcasts the information to the users once it estimates  $h_j^{(\text{sr})}(t)$ , where we note that this estimation is assumed to be perfect. It is worth noting that the cost of broadcasting such information is negligible.

### 4.4.1 Case without Delay

Here, we assume that there is no delay in the feedback process. As explained earlier, the feedback and scheduling decisions are coupled, and it is of great importance to design an efficient algorithm to make these decisions. Such an algorithm will be provided here, and is termed RW, for which we study the stability performance by characterizing the minimum fraction this algorithm can achieve w.r.t. the stability region of the *ideal* system. In this latter system, in addition to the facts that there is no delay and that  $h_{ij}^{(\text{su})}(t)$  and  $h_{ij}^{(\text{ru})}(t)$ ,  $\forall j$ , are perfectly known at user  $i$  at time-slot  $t$ , the users have perfect knowledge of the fading coefficients of the links between the BS and the relay, i.e. at time-slot  $t$  user  $i$  knows  $h_j^{(\text{sr})}(t)$ ,  $\forall j$ . Thus, under the ideal system, each user can calculate  $C_{ij}^{(\text{r})}(t)$  and consequently  $C_{ij}(t)$ ; i.e.  $C_{ij}(t) = \max\{C_{ij}^{(\text{d})}(t), C_{ij}^{(\text{r})}(t)\}$ .

Unlike the ideal system, under the adopted scheme, at time-slot  $t$  and for channel  $j$  user  $i$  may have an imperfect estimation of  $C_{ij}^{(\text{r})}(t)$  since the only information user  $i$  receives about  $h_{ij}^{(\text{sr})}(t)$  is if  $\gamma_j^{(\text{sr})}(t) \geq \delta$  or  $\gamma_j^{(\text{sr})}(t) < \delta$ . Let us denote this imperfect knowledge of the link state as  $\check{C}_{ij}^{(\text{sr})}(t)$ . It can be noticed that, however,  $C_{ij}^{(\text{d})}(t)$  is perfectly estimated at user  $i$ . Define  $\check{C}_{ij}(t)$  as

$$\check{C}_{ij}(t) = \max\{C_{ij}^{(\text{d})}(t), \check{C}_{ij}^{(\text{r})}(t)\}. \quad (4.49)$$

#### 4.4. Proposed Algorithms and Stability Analysis for the System with Relaying

It is worth recalling that  $\gamma_{ij}^{(d)}(t) = \gamma_{ij}^{(su)}(t)$  and  $\gamma_{ij}^{(r)}(t) = \gamma_{ij}^{(su)}(t) + \gamma_{ij}^{(sru)}(t)$ , where

$$\gamma_{ij}^{(sru)}(t) = \frac{\gamma_{ij}^{(ru)}(t) \gamma_j^{(sr)}(t)}{\gamma_{ij}^{(ru)}(t) + \gamma_j^{(sr)}(t) + 1}. \quad (4.50)$$

Before presenting algorithm RW, we first explain some important points that are used in this algorithm. If  $\gamma_j^{(sr)}(t) \geq \delta$ , then to calculate  $\check{C}_{ij}^{(r)}$  user  $i$  replaces  $\gamma_j^{(sr)}(t)$  by  $\delta$ . In detail, user  $i$  calculates  $\gamma_{ij}^{(su)}(t)$ ,  $\check{\gamma}_{ij}^{(sru)}(t)$  and  $\check{\gamma}_{ij}^{(r)}(t)$ , where

$$\check{\gamma}_{ij}^{(sru)}(t) = \frac{\gamma_{ij}^{(ru)}(t) \delta}{\gamma_{ij}^{(ru)}(t) + \delta + 1} \text{ and } \check{\gamma}_{ij}^{(r)}(t) = \gamma_{ij}^{(su)}(t) + \check{\gamma}_{ij}^{(sru)}(t), \quad (4.51)$$

based on which  $C_{ij}^{(d)}(t)$  and  $\check{C}_{ij}^{(r)}(t)$  can be determined; i.e. if  $\tau_k \leq \gamma_{ij}^{(d)}(t) < \tau_{k+1}$ , then  $C_{ij}^{(d)}(t) = R_k$ , and if  $\tau_k \leq \check{\gamma}_{ij}^{(r)}(t) < \tau_{k+1}$ , then  $\check{C}_{ij}^{(r)}(t) = \frac{1}{2} R_k$ , where  $k \in \{1, \dots, K\}$ . This user can then find  $\check{C}_{ij}(t)$ . On the other hand, if  $\gamma_j^{(sr)}(t) < \delta$ , user  $i$  sets  $\check{C}_{ij}^{(r)}(t) = 0$ , that is to say, the choice of using the relay to transmit to user  $i$  on channel  $j$  is not possible at time-slot  $t$ .

---

**Algorithm RW:** For the adopted system here, the decisions for feedback and scheduling are based on the following procedure.

1. *Queue lengths broadcast every  $T_b$  slots:*

Every  $T_b$  time-slots, that is to say, at time  $0, T_b, 2T_b, \dots, nT_b, \dots$ , the BS broadcasts the queue lengths of all users, where  $T_b$  is typically high. So each user has outdated knowledge of the state of its queue (and all other queues). Let  $\tilde{q}_i(t)$  ( $\forall i$ ) represent the (outdated) queue length the users know at time  $t$ , i.e.  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b[$ .

2. *Feedback and scheduling decisions at time-slot  $t$ :*

Define  $\phi_{ij}(t) = 0$ .

After the relay broadcasts the information (i.e. if  $\gamma_j^{(sr)}(t) \geq \delta$  or  $\gamma_j^{(sr)}(t) < \delta$ , for all  $j$ ), user  $i$  calculates  $\check{C}_{ij}(t) = \max\{C_{ij}^{(d)}(t), \check{C}_{ij}^{(r)}(t)\}$ .

If  $\check{C}_{ij}(t) = C_{ij}^{(d)}(t)$ , then user  $i$  sets  $\phi_{ij}(t) = \check{C}_{ij}(t)$ .

Otherwise (i.e. if  $\check{C}_{ij}(t) = \check{C}_{ij}^{(r)}(t)$ ), user  $i$  sets  $\phi_{ij}(t)$  as

$$\phi_{ij}(t) = \check{C}_{ij}(t) \mathbb{P} \left\{ C_{ij}(t) \geq \check{C}_{ij}(t) \mid \gamma_j^{(sr)}(t) \geq \delta, h_{ij}^{(su)}(t), h_{ij}^{(ru)}(t) \right\}. \quad (4.52)$$

For each channel  $j \in \{1, \dots, L\}$ , only one user sends its CSI to the BS. This user is the result of the Max-Weight rule and can be given by

$$\arg \max_i \{ \tilde{q}_i(t) \phi_{ij}(t) \}. \quad (4.53)$$

To detect this user, we use an approach that consists in letting the users contend among each other for a certain period of time. Let  $T_c$  be the contention period for each channel. Assuming that contention can be done in continuous time, i.e. there is no collision between users, for each

#### 4.4. Proposed Algorithms and Stability Analysis for the System with Relaying

---

channel  $j \in \{1, \dots, L\}$  the contention is performed as the following:

User  $i$ ,  $\forall i \in \{1, \dots, N\}$ , waits until time

$$T_c (\tilde{q}_i(t) \phi_{ij}(t))^{-1}, \quad (4.54)$$

then broadcasts a signal (of negligible duration). It is obvious that the user that broadcasts its signal first will be the result of (4.53) since this user will wait the smallest period of time. After the broadcast of the first signal, the contention procedure (of channel  $j$ ) terminates and the corresponding user reports its CSI. Also, this user sends to the BS the decision about using or not the relay for transmission, depending on whether  $\check{C}_{ij}^{(r)}(t) > C_{ij}^{(d)}(t)$  or  $\check{C}_{ij}^{(r)}(t) < C_{ij}^{(d)}(t)$ . Once this procedure finishes, the contention of another channel gets started.

##### 3. *Transmission at time-slot $t$ :*

The BS receives the CSI of each channel and uses these CSIs for transmission at time-slot  $t$ . It is obvious that the user that is selected to report its feedback of channel  $j$  will also be the user transmitting over this channel.

---

In the above algorithm, each user compares the achievable rates and decides whether the transmission with or without relaying is better. Then, for the feedback decision where for each channel (at most) one user reports its feedback, the users contend among each other in such a way as to ensure that the user with the greatest product  $\tilde{q}_i(t) \phi_{ij}(t)$  reports its CSI of channel  $j$ ; note that a similar approach was used in algorithm FSA. Obviously, this user will be scheduled for transmission over channel  $j$ , where we note that this same user sends its decision about using or not the relay for transmission so that the BS knows what scenario to adopt. Recall that if at a given slot the decision is not to use the relay, then the BS transmits during all the slot period, whereas the decision of using the relay implies that the BS transmits during the first half of the slot and then the relay amplifies the received signal and forwards it to the user in the second half. We finally point out that in equation (4.52)  $\check{C}_{ij}(t)$  is multiplied by the probability

$$\mathbb{P} \left\{ C_{ij}(t) \geq \check{C}_{ij}(t) \mid \gamma_j^{(\text{sr})}(t) \geq \delta, h_{ij}^{(\text{su})}(t), h_{ij}^{(\text{ru})}(t) \right\} \quad (4.55)$$

due to a possible mismatch between  $C_{ij}(t)$  and  $\check{C}_{ij}(t)$  in the corresponding case; recall that a transmission at a rate strictly greater than  $C_{ij}(t)$  always fails.

#### Stability Performance Analysis for RW

Let us denote the system here and the ideal system by 're' and 'rp', respectively. Also, let  $\mathcal{M}^{\text{rp}}(t)$  and  $\mathcal{M}^{\text{re}}$  be the subsets of scheduled links under systems 're' and 'rp', respectively. For the ideal system (i.e. 'rp'), define  $g_{\text{rp}}$  to be the expected weighted throughput, which can be written as

$$g_{\text{rp}}(\mathbf{q}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{rp}}(t)} q_i(t) C_{ij}(t) \mid \mathbf{q}(t) \right\}. \quad (4.56)$$

#### 4.4. Proposed Algorithms and Stability Analysis for the System with Relaying

In addition, the expected weighted throughput for system 're', denoted by  $g_{\text{re}}$ , can be expressed as

$$g_{\text{re}}(\tilde{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{re}}(t)} \tilde{q}_i(t) \check{C}_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \check{C}_{ij}(t))} \mid \tilde{\mathbf{q}}(t) \right\}. \quad (4.57)$$

Now, we characterize the minimum fraction that algorithm RW guarantees to achieve w.r.t. the stability region of the ideal system. Because the characterization of this fraction for the multiple rate levels case is very complex, here we constrain ourself to the *single* rate level case (i.e. one modulation). Denote this rate as  $R$  and let  $\tau$  be its corresponding threshold. Thus, if  $\gamma_{ij}^{(\text{d})}(t) \geq \tau$ , then  $C_{ij}^{(\text{d})}(t) = R$ , whereas if  $\gamma_{ij}^{(\text{d})}(t) < \tau$  and  $\gamma_{ij}^{(\text{r})}(t) \geq \tau$ , then  $C_{ij}^{(\text{r})}(t) = \frac{1}{2}R$ ; also, if  $\gamma_{ij}^{(\text{d})}(t) < \tau$ ,  $\gamma_j^{(\text{sr})}(t) \geq \delta$  and  $\check{\gamma}_{ij}^{(\text{r})}(t) \geq \tau$ , then  $\check{C}_{ij}^{(\text{r})}(t) = \frac{1}{2}R$ . Under the above scheme, the following result can be stated.

**Theorem 15.** *For the imperfect system with relaying and under a single rate level scheme, algorithm RW can achieve at least a fraction  $\beta_r$  of the stability region of the ideal system, where*

$$\beta_r = (1 - \frac{1}{T_b}) \min_{j \in \{1, \dots, L\}} \left\{ \min_{i \in \{1, \dots, N\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\} \right\}. \quad (4.58)$$

*Proof.* We first provide a result that will be useful for the rest of the proof. To this end, define 'rp<sub>1</sub>' to be the system in which the BS has perfect knowledge of all the link states and then applies the Max-Weight rule for scheduling at time-slot  $t$  using  $\tilde{q}(t)$ . Recall that  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b]$ . The only difference between 'rp<sub>1</sub>' and 'rp' (i.e. ideal system) is that at time-slot  $t$  the scheduling process under 'rp<sub>1</sub>' is done based on  $\tilde{\mathbf{q}}(t)$ , while under 'rp' the scheduling is based on  $\mathbf{q}(t)$ . Let  $\Lambda_{\text{rp}}$  and  $\Lambda_{\text{rp}_1}$  represent the stability regions achieved under systems 'rp' and 'rp<sub>1</sub>', respectively. Using the above definitions, the following result can be given.

**Lemma 8.** *System 'rp<sub>1</sub>' is strongly stable if and only if system 'rp' is strongly stable. In other words, in terms of stability regions, we have  $\Lambda_{\text{rp}} = \Lambda_{\text{rp}_1}$ .*

The complete proof of the theorem can be found in Appendix 4.7.2, where a remark about the proof of Lemma 8 is given.  $\square$

Note that for the above theorem no specific assumptions are made regarding the channel statistics and the fading model. By assuming that the fading coefficients are modeled as complex Gaussian processes and that their distributions, as well as the noises distributions, are the same independently of the channels (i.e. frequency bands), the following result can be given.

**Corollary 1.** *Recalling that  $z_j^{(\text{r})}(t)$  and  $z_{ij}^{(\text{u})}(t)$  are the noises corresponding to channel  $j$  at the relay and at user  $i$ , which are assumed to be additive white Gaussian processes with zero mean and variances  $(\sigma_{z,j}^{(\text{r})})^2$  and  $(\sigma_{z,ij}^{(\text{u})})^2$ , respectively, and assuming that*

- $\sigma_{z,j}^{(\text{r})}$  and  $\sigma_{z,ij}^{(\text{u})}$  are independent of  $j$  (which represents the channel index) and are thus denoted by  $\sigma_z^{(\text{r})}$  and  $\sigma_{z,i}^{(\text{u})}$ , respectively,
- The fading coefficients  $h_{ij}^{(\text{su})}(t)$ ,  $h_{ij}^{(\text{ru})}(t)$  and  $h_j^{(\text{sr})}(t)$ , for all  $j$ , are i.i.d. in time and follow complex Gaussian distributions with zero mean and variances  $(\sigma_i^{(\text{su})})^2$ ,  $(\sigma_i^{(\text{ru})})^2$  and  $(\sigma^{(\text{sr})})^2$ , respectively, i.e. the distribution of each one of these coefficients is the same for all the channels,

#### 4.4. Proposed Algorithms and Stability Analysis for the System with Relaying

the fraction  $\beta_r$  reduces to

$$\beta_r = \left(1 - \frac{1}{T_b}\right) \min_{i \in \{1, \dots, N\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\}, \quad (4.59)$$

where for  $\delta < \tau$  the probability  $\mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\}$  is given as

$$\begin{aligned} \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} = & \exp \left( -\frac{\tau}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) + \\ & \left[ \int_{\tau-\delta}^{\tau} \exp \left( -\frac{(\tau-\gamma)(\delta+1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta-(\tau-\gamma))} \right) \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp \left( -\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) d\gamma \right] \exp \left( -\frac{\delta}{\rho^{(\text{sr})}(\sigma^{(\text{sr})})^2} \right), \end{aligned} \quad (4.60)$$

whereas for  $\delta \geq \tau$  we have

$$\begin{aligned} \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} = & \exp \left( -\frac{\tau}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) + \\ & \left[ \int_0^{\tau} \exp \left( -\frac{(\tau-\gamma)(\delta+1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta-(\tau-\gamma))} \right) \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp \left( -\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) d\gamma \right] \exp \left( -\frac{\delta}{\rho^{(\text{sr})}(\sigma^{(\text{sr})})^2} \right), \end{aligned} \quad (4.61)$$

with  $\rho_i^{(\text{su})}$ ,  $\rho_i^{(\text{ru})}$ , and  $\rho^{(\text{sr})}$  defined as

$$\rho_i^{(\text{su})} = \frac{P^{(\text{s})}}{(\sigma_{z,i}^{(\text{u})})^2}, \quad \rho_i^{(\text{ru})} = \frac{P^{(\text{r})}}{(\sigma_{z,i}^{(\text{u})})^2}, \quad \rho^{(\text{sr})} = \frac{P^{(\text{s})}}{(\sigma_z^{(\text{r})})^2}. \quad (4.62)$$

*Proof.* The proof can be found in Appendix 4.7.3. □

Regarding the above corollary, some remarks are in order. First, the integrals in equations (4.60) and (4.61) do not have closed form solutions, however these integrals can be simply evaluated using numerical integration. In addition, if  $\rho_i^{(\text{su})}$  and  $\rho_i^{(\text{ru})}$  are both independent of the user index  $i$ , then  $\beta_r$  reduces to the following expression

$$\beta_r = \left(1 - \frac{1}{T_b}\right) \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\}, \quad (4.63)$$

since in this case  $\mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\}$  is independent of the user's identity. Moreover, to decide what value of  $\delta$  should be used, one approach is to choose this threshold in such a way as to maximize  $\beta_r$ , since a greater  $\beta_r$  implies better guaranteed stability performance.

Another possible and even more generalized approach (i.e. it holds independently of the assumptions in Corollary 1) regarding the selection of  $\delta$  is to let the BS update its value each period of time and then broadcast it to the users and the relay, where the update process is done as follows.



#### 4.4. Proposed Algorithms and Stability Analysis for the System with Relaying

Supposing that an update should be done at time-slot  $t$  and that the expressions of  $\mathbb{P}\{\check{C}_{ij}^{(d)}(t) > 0\}$  and  $\mathbb{P}\{\check{C}_{ij}^{(r)}(t) > 0\}$  are known by the BS,  $\delta$  is determined according to

$$\delta = \arg \max_{\delta} \left\{ \max_{(ij)} \left\{ q_i(t) \max \left\{ R \mathbb{P}\{\check{C}_{ij}^{(d)}(t) > 0\}, \frac{1}{2} R \mathbb{P}\{\check{C}_{ij}^{(r)}(t) > 0\} \right\} \right\} \right\}, \quad (4.64)$$

where we note that  $\mathbb{P}\{\check{C}_{ij}^{(r)}(t) > 0\}$  depends on  $\delta$  whereas  $\mathbb{P}\{\check{C}_{ij}^{(d)}(t) > 0\}$  is independent of  $\delta$ . Usually, the choice of  $\delta$  is from a set of possible values, in which case the solution of the optimization problem given before can be easily found numerically.

#### 4.4.2 Case with Delay

Unlike the previous section, here the delay is accounted for. Recall that the amount of feedback resources and the delay are coupled, that is to say, the more amount of feedback resources are allocated, the more delay is incurred in the feedback process. In a similar fashion to the case without relaying, in this subsection we provide two algorithms, termed as RFSA and RFMA, for making feedback and scheduling decisions, where we note that these algorithms are simple variations of FSA and FMA.

Recall that at time-slot  $t$  the only information the users have about  $h_{ij}^{(sr)}(t)$ ,  $\forall j$ , is if  $\gamma_j^{(sr)}(t) \geq \delta$  or not. As in the previous subsection, in order for the users to calculate  $\check{C}_{ij}^{(r)}(t)$ , the following approach is adopted: (a) if  $\gamma_j^{(sr)}(t) \geq \delta$ , then user  $i$  replaces  $\gamma_j^{(sr)}(t)$  by  $\delta$ , and calculates  $\check{C}_{ij}^{(r)}(t)$ , whereas (b) if  $\gamma_j^{(sr)}(t) < \delta$ , then user  $i$  sets  $\check{C}_{ij}^{(r)}(t) = 0$ . After finding  $\check{C}_{ij}^{(r)}(t)$  and  $C_{ij}^{(d)}(t)$ , user  $i$  can compute  $\check{C}_{ij}(t)$ , where we have  $\check{C}_{ij}(t) = \max\{C_{ij}^{(d)}(t), \check{C}_{ij}^{(r)}(t)\}$ . Recall that if  $\tau_k \leq \gamma_{ij}^{(d)}(t) < \tau_{k+1}$ , then  $C_{ij}^{(d)}(t) = R_k$ , and if  $\tau_k \leq \check{\gamma}_{ij}^{(r)}(t) < \tau_{k+1}$  (and  $\gamma_j^{(sr)}(t) \geq \delta$ ), then  $\check{C}_{ij}^{(r)}(t) = \frac{1}{2}R_k$ , where  $k \in \{1, \dots, K\}$ . Furthermore, we use  $\hat{C}_{ij}(t)$  to represent  $\check{C}_{ij}(t - d)$ , i.e.  $\hat{C}_{ij}(t) = \check{C}_{ij}(t - d)$ .

In a similar way to algorithm FSA, RFSA uses  $L$  feedback resources, each of which is reserved for one of the  $L$  channels; hence, the delay  $d = \lceil F(\bar{F})^{-1} \rceil$ . Due to the high similarity between these two algorithms, we only present the outline for RFSA to avoid repetition.

**Algorithm RFSA:** Under this algorithm, the feedback and scheduling decisions are done based on the following procedure.

1. *Queue lengths broadcast every  $T_b$  slots:*

This step is the same as in FSA.

2. *Feedback and scheduling decisions at time-slot  $t - d$  (made at the users side):*

This step is similar to the second step in FSA, except that  $\hat{C}_{ij}(t)$  should be replaced with  $\hat{C}_{ij}(t)$ . Obviously, the fading coefficients used here are those of time-slot  $t - d$ .

3. *Transmission at time-slot  $t$ :*

This step is the same as in FSA.

Note that algorithm RFSA assumes a continuous contention time scheme, that is to say, there is no collision between the users when they contend to find, on each channel, which user should be selected. Alternatively, we propose algorithm RFMA that is adapted for a discrete contention time scheme and that tries to greatly imitate RFSA. As FMA, under RFMA an amount of  $F$  feedback

## 4.5. Validation of the Proposed Model

resources is used, i.e. at most  $F$  link states can be reported; here, it can be seen that  $d = \lceil F(\bar{F})^{-1} \rceil$ .

**Algorithm RFMA:** Under this algorithm, the feedback and scheduling decisions are done based on the following procedure.

1. *Queue lengths broadcast every  $T_b$  slots:*  
This step is identical to that in FMA.
2. *Queue lengths sorting at each user:*  
This step is identical to that in FMA.
3. *Feedback decisions at time-slot  $t - d$  (made at the users side):*  
This step is similar to that in FMA, except that  $\hat{C}_{ij}(t)$  should be replaced with  $\hat{\check{C}}_{ij}(t)$ .
4. *Scheduling decisions (made at the BS) and transmissions at time-slot  $t$ :*  
This step is similar to that in FMA. Clearly, the BS relies on the subset of reported  $\hat{\check{C}}_{ij}(t)$  in order to make the scheduling decisions.

## 4.5 Validation of the Proposed Model

### System without Relaying

Table 4.1: List of simulation parameters for the system without relaying

Parameter	Description	Value
$N$	Number of users	30
$L$	Number of channels	30
$P$	Power per channel	10 dB
$BW$	Bandwidth per channel	180 KHz
$f_c$	Carrier frequency	2.1 GHz
$c$	Speed of light	$3 \times 10^8$ m/sec
$T_s$	Slot duration	1 msec
$M_s$	Number of slots per simulation	$10^4$

Here, the numerical results for the system without relaying are presented. We set  $N = L = 30$ ,  $P = 10 \log_{10}(10) = 10$  dB. We consider an LTE-like system with a bandwidth  $BW = 180$  KHz per channel and a carrier frequency  $f_c = 2.1$  GHz. Let  $T_s$  represent the slot period, which is set to 1 msec. Assume that the  $\sigma_{z,ij}$  are all equal to 1, i.e.  $\sigma_{z,ij} = 1, \forall i, j$ . To model the impact of delay, we consider the Gauss-Markov block fading process [128]. Based on this model,  $h_{ij}(t)$  can be expressed as follows

$$h_{ij}(t) = \sigma \hat{h}_{ij}(t) + e_{ij}(t), \quad (4.65)$$

where  $h_{ij}(t)$  is a complex normal random variable with zero mean and unit variance, i.e.  $h_{ij}(t) \sim \mathcal{CN}(0, 1)$ , and where  $e_{ij}(t) \sim \mathcal{CN}(0, \sigma_e^2)$  is the error due to delay. Notice that  $\hat{h}_{ij}(t) \sim \mathcal{CN}(0, 1)$  and

## 4.5. Validation of the Proposed Model

Table 4.2: Possible rates used for the simulations.

<b>Rate (bits/slot)</b>	505.32	570.58	622.69	666.08	703.24
<b>Rate (bits/slot)</b>	735.74	764.62	790.61	891.75	964.36

$\sigma_e^2 = 1 - \sigma^2$ . The correlation coefficient is given by

$$\sigma = J_0(2\pi f_{ds}T_s d), \quad (4.66)$$

with Doppler spread  $f_{ds}$ , where  $T_s$  is the slot duration,  $d$  is the delay in number of time-slots, and  $J_0(\cdot)$  is the zero-th order Bessel function of the first kind. The Doppler spread can be given by  $f_{ds} = \frac{f_c v}{c}$ , where  $v$  is the user velocity and  $c = 3 \times 10^8$  m/sec is the speed of light; we assume that all the users have the same velocity  $v$ . Hence, the correlation coefficient  $\sigma = J_0(2\pi f_{ds}T_s d) = J_0(2\pi f_c v c^{-1}T_s d)$ . The fading coefficients are assumed to be i.i.d. across users and frequencies. The set of possible rates  $\{R_1, \dots, R_K\}$  is given in Table 4.2. We suppose that all the users have Poisson incoming traffic with the same mean arrival rate  $a$ , i.e.  $a_i = a, \forall i \in \{1, \dots, N\}$ .

To show the stability performance of the system, we plot the total average queue length, defined as

$$\frac{1}{M_s} \sum_{t=0}^{M_s-1} \sum_{k=1}^N q_k(t) \quad (4.67)$$

for different values of mean arrival rate  $a$ , where  $M_s$  represents the number of time-slots each simulation lasts (i.e. per mean arrival rate). We set  $M_s = 10^4$ . It is worth noting that the point where the total average queue length begins to increase very steeply is the point where the system becomes unstable.

We provide four figures where in each one we consider a different value of the velocity,  $v$ . Specifically, in Figures 4.4 to 4.7 we set  $v = 1, 5, 10$ , and  $50$  km/hr, respectively. Each figure depicts the variation of total average queue length for different values of mean arrival rate under various system settings. The first setting is ideal system 'pf', i.e. the system with perfect and full feedback. The three remaining settings consider system 'dl', i.e. the system with delayed and limited feedback where algorithm FMA is used for feedback and scheduling decisions, for three different amounts of feedback resources  $F = 50, 100$ , and  $150$ . Recall that delay  $d$  depends on the value of  $F$ . Here we assume that  $F = 50, 100$ , and  $150$  result in  $d = 1, 2$ , and  $3$  slots, respectively.

The simulations show that for relatively small values of the velocity (Figure 4.4 and Figure 4.5), increasing the feedback resources  $F$  to some limit can provide better stability performance. This is due to the fact that for small values of  $v$ , the variance of the error given by  $\sigma_e^2 = 1 - \sigma^2$  is sufficiently small, so the gain coming from having more feedback information overcomes the loss due to delay. Evidently, the limit until which if we increase  $F$  we get better performance depends on  $\sigma_e$  and consequently on  $v$ . It can be noticed that for  $v = 1$  km/hr, the best  $F$  we can choose is  $150$ . However, for  $v = 5$  km/hr, the best  $F$  is  $100$ . If we keep increasing  $v$ , there will be no gain from taking  $F > 50$ . This can be seen in Figure 4.6 and more clearly in Figure 4.7. Specifically, for  $v = 10$  km/hr, setting  $F = 50$  (or  $F = 100$ ) yields better stability performance than the case where  $F = 150$ . Also, for  $v = 50$  km/hr,  $F = 50$  yields better performance than both of  $F = 100$  and  $F = 150$  cases. This is due to the fact that for relatively high velocities the variance of the (delay) error, i.e.  $\sigma_e^2$ , is sufficiently high, so the

## 4.5. Validation of the Proposed Model

loss due to delay is bigger than the gain coming from having more feedback information.

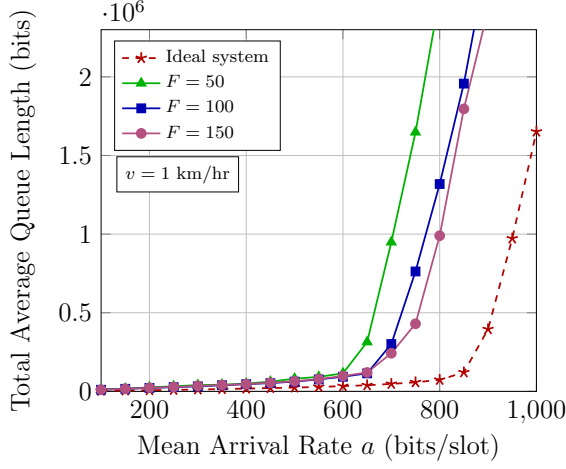


Figure 4.4: Total average queue length vs. mean arrival rate  $a$ . Here, the velocity  $v = 1$  km/hr.

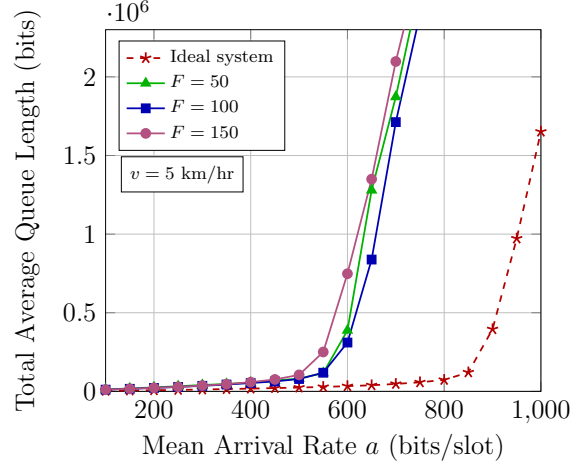


Figure 4.5: Total average queue length vs. mean arrival rate  $a$ . Here, the velocity  $v = 5$  km/hr.

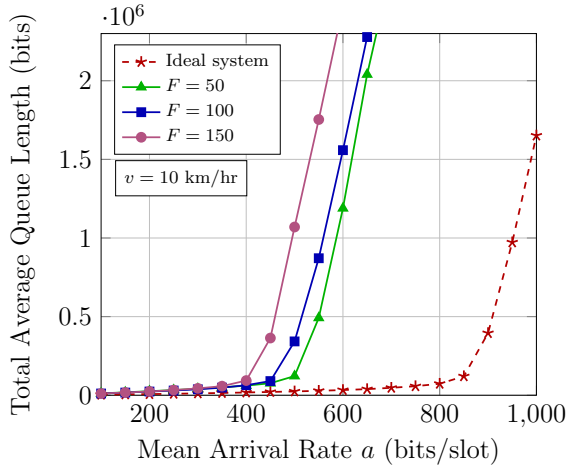


Figure 4.6: Total average queue length vs. mean arrival rate  $a$ . Here, the velocity  $v = 10$  km/hr.

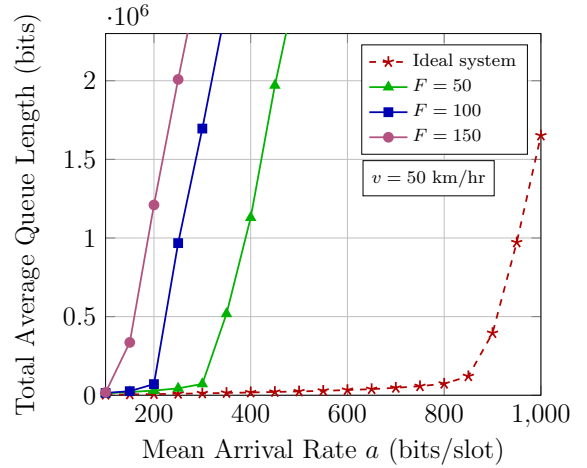


Figure 4.7: Total average queue length vs. mean arrival rate  $a$ . Here, the velocity  $v = 50$  km/hr.

## System with Relaying

The numerical results for the system with relaying are presented here. Specifically, we focus on the case without delay since the numerical results and the discussions for the case with delay are basically similar to those provided before for the system without relaying. We note that the scheme that is adopted here is essentially the one under which Corollary 1 holds.

We set  $N = L = 3$ ,  $P^{(s)} = P^{(r)} = 10$  dB. The bandwidth per channel is equal to 180 KHz. The slot duration is set to 1 msec. Assume that  $(\sigma_{z,ij}^{(r)})^2 = (\sigma_{z,ij}^{(u)})^2 = 1$ ,  $\forall i, j$ . We also assume that the coefficients  $h_{ij}^{(su)}(t)$ ,  $h_{ij}^{(sr)}(t)$ , and  $h_{ij}^{(ru)}(t)$  are i.i.d. across users, time, and frequencies and they follow

#### 4.5. Validation of the Proposed Model

Table 4.3: List of simulation parameters for the system with relaying

Parameter	Description	Value
$N$	Number of users	3
$L$	Number of channels	3
$P$	Power per channel	10 dB
$BW$	Bandwidth per channel	180 KHz
$f_c$	Carrier frequency	2.1 GHz
$T_s$	Slot duration	1 msec
$M_s$	Number of slots per simulation	$10^4$

complex Gaussian distributions with zero mean and variances  $\sigma^{(\text{su})}$ ,  $\sigma^{(\text{sr})}$ , and  $\sigma^{(\text{ru})}$ , respectively. As mentioned in the system model, these coefficients capture the effects of path-loss and fading. Typically, between the BS and the users the effect of the path-loss is higher than that between the relay and the users or that between the BS and the relay. This results from the fact that the distance between the BS and the users is greater than that between the BS and the relay or that between the relay and the users. Based on these observations, we set  $(\sigma^{(\text{sr})})^2 = (\sigma^{(\text{ru})})^2 = 1$  and  $(\sigma^{(\text{su})})^2 = 0.3$ . Regarding the number of modulations, we assume that there is only a single rate level, denoted as  $R$ , that corresponds to threshold  $\tau$ ; clearly, under the adopted setting we have  $R = 180 \log_2(1 + \tau)$  bits/slot. Such a scheme is used to be able to provide some numerical results on the variation of the minimum fraction  $\beta_r$  w.r.t.  $\delta$  because the explicit expression of  $\beta_r$  was found only under the single rate level assumption. We suppose that all the users have Poisson incoming traffic with the same mean arrival rate  $a$ . To show stability performance, one way is to adopt a similar approach as that used for the system without relaying, i.e. we plot the total average queue length for different values of  $a$ ; we set  $M_s = 10^4$ . Finally, the broadcast period  $T_b = 20$  slots, which is sufficiently high, i.e.  $1 - \frac{1}{T_b} \approx 1$ .

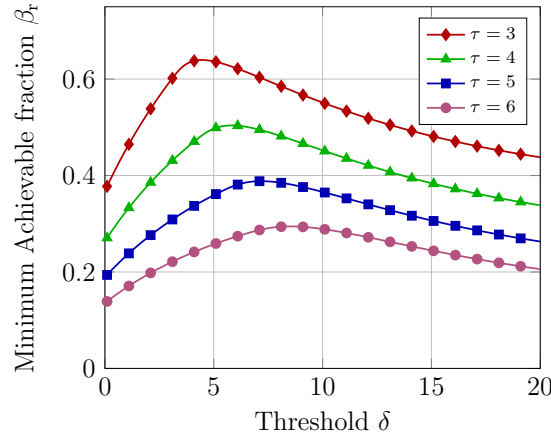


Figure 4.8: Minimum achievable fraction  $\beta_r$  vs. threshold  $\delta$ .

Let us recall and note a couple of things that will be useful for the description of the numerical results. First, recall that  $\beta_r$  represents the performance that algorithm RW guarantees w.r.t. the ideal system, that is to say, RW can achieve a fraction that is greater than or equal to  $\beta_r$ . One way to choose the value of  $\delta$  is to use the  $\beta_r$  expression, i.e. choose  $\delta$  that maximizes  $\beta_r$ . This value of  $\delta$  ensures the best guaranteed performance for algorithm RW, however this does not necessarily mean

#### 4.5. Validation of the Proposed Model

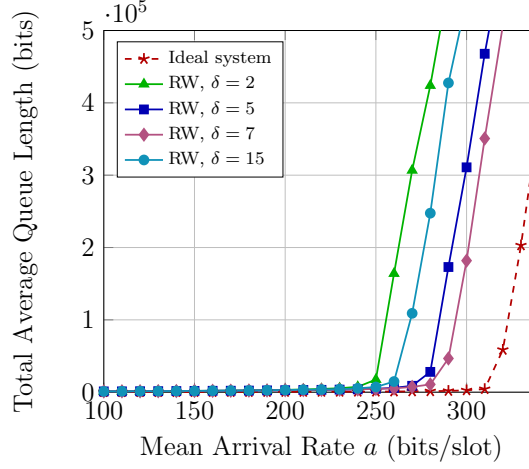


Figure 4.9: Total average queue length vs. mean arrival rate  $a$ . Here  $\tau = 5$ .

that this value yields the best (effective) performance for RW. Moreover, recall that  $\gamma_{ij}^{(d)}(t) = \gamma_{ij}^{(su)}(t)$ ,  $\check{\gamma}_{ij}^{(r)}(t) = \gamma_{ij}^{(su)}(t) + \check{\gamma}_{ij}^{(sru)}(t)$  and that under algorithm RW if  $\gamma_j^{(sr)}(t) \geq \delta$  then to calculate  $\check{C}_{ij}^{(r)}$  user  $i$  replaces  $\gamma_j^{(sr)}(t)$  by  $\delta$ , i.e. user  $i$  calculates  $\gamma_{ij}^{(su)}(t)$ ,  $\check{\gamma}_{ij}^{(sru)}(t)$  and  $\check{\gamma}_{ij}^{(r)}(t)$ , where

$$\check{\gamma}_{ij}^{(sru)}(t) = \frac{\gamma_{ij}^{(ru)}(t) \delta}{\gamma_{ij}^{(ru)}(t) + \delta + 1} \text{ and } \check{\gamma}_{ij}^{(r)}(t) = \gamma_{ij}^{(su)}(t) + \check{\gamma}_{ij}^{(sru)}(t). \quad (4.68)$$

On the other hand, if  $\gamma_j^{(sr)}(t) < \delta$  then user  $i$  sets  $\check{C}_{ij}^{(r)}(t) = 0$ , meaning that at time-slot  $t$  the choice of using the relay to transmit to user  $i$  on channel  $j$  is not possible. Under the adopted setting here, it can be noticed that if  $\gamma_{ij}^{(d)}(t) \geq \tau$  then  $C_{ij}^{(d)}(t) = R$  and if  $\check{\gamma}_{ij}^{(r)}(t) \geq \tau$  (with  $\gamma_j^{(sr)}(t) \geq \delta$ ) then  $\check{C}_{ij}^{(r)}(t) = \frac{1}{2}R$ . A final point to note here is that under the ideal system all the  $C_{ij}(t)$  are known at the BS, which depend on  $\gamma_{ij}^{(d)}(t)$  and  $\gamma_{ij}^{(r)}(t)$ .

Figure 4.8 depicts the variation of the minimum achievable fraction  $\beta_r$  (see Corollary 1) w.r.t. threshold  $\delta$  for different values of threshold  $\tau$ . From this figure, for a fixed  $\delta$  it is obvious that  $\beta_r$  increases when  $\tau$  decreases. This results from the fact that the lesser  $\tau$ , the more likely it is to get  $\gamma_{ij}^{(d)}(t) \geq \tau$ ,  $\forall i, j$ , and thus to select the transmission without relay, which implies that in most of the time RW makes the same scheduling decisions as those of the ideal system. On the other hand, it can be noticed that for a fixed  $\tau$  the fraction  $\beta_r$  reaches its maximum at one value of  $\delta$  and this fraction decreases as we move away from this value. Although, as indicated previously, choosing the values of  $\delta$  that maximize  $\beta_r$  is not sure to yield the best performance, one can still use this approach to find a judicious choice of  $\delta$  that guarantees RW to deliver good performance.

In Figure 4.9 we plot the variation of the total average queue length w.r.t. the mean arrival rate  $a$  for three different values of  $\delta$ ; here  $\tau = 5$ . It can be seen that a judicious choice of  $\delta$ , for e.g.  $\delta = 5$  or  $\delta = 7$ , permits algorithm RW to achieve very good stability performance w.r.t. the ideal system, whereas an injudicious choice of  $\delta$  leads to less good performance. The results in the figure show that the approach that consists in choosing the values of  $\delta$  that maximize  $\beta_r$  is actually an efficient approach. Specifically, by recalling that  $\tau = 5$  in Figure 4.9 and looking at the case  $\tau = 5$  in Figure

4.8, it can be seen that choosing  $\delta = 7$ , which maximizes  $\beta_r$ , delivers the best stability performance. Note that choosing any value in the interval  $[5, 10]$  can be considered as a judicious choice of  $\delta$  since, as it can be noticed from Figure 4.8, over this interval  $\beta_r$  does not change much.

## 4.6 Closing Remarks

In this chapter, we addressed the problems of feedback allocation and scheduling for a multiuser multichannel downlink cellular network under limited and delayed feedback. Two scenarios were investigated: (i) system without relaying, and (ii) system with relaying. For the system with relaying, we account for an additional imperfection: the users have incomplete knowledge of the fading coefficients between the BS and the relay. For each system, we proposed an efficient joint feedback allocation and scheduling algorithm, in which the decisions are made at the users side. This algorithm is shown to achieve good stability performance with respect to the ideal system, however it is suitable for a continuous-time contention scheme. In case the contention is only possible in discrete time, we proposed a second algorithm, which uses a threshold-based concept, that imitates the first one to a great extent and thus guarantees good stability performance. For this algorithm, the feedback decision is done at the users side, and then the BS uses this feedback to perform users scheduling. Regarding the choice of the amount of feedback resources for this algorithm, we provided numerical results that find the best trade-off (i.e. best  $F$ ) between having more feedback resources (thus, more knowledge at the scheduler) but longer delay (hence, less accurate CSI) and having less feedback resources but low delay; these results are given for various system setups, i.e. different values of users velocity. For the system with relaying, we investigated the special case where the delay in the feedback process is not accounted for. Specifically, under this scheme, we proposed a joint feedback and scheduling algorithm, we analyzed its stability performance and we derived the fraction this algorithm guarantees w.r.t. the stability region of the ideal system. Finally, using some numerical results, we validated that each of the proposed algorithms provides good stability performance.

## 4.7 Appendix

### 4.7.1 Proof of Theorem 14

The proof consists of three main steps. We first provide the proof of Lemma 7. Then, we show that  $g_{\text{mdl}}(\hat{\mathbf{q}}(t)) \geq \frac{p_c^{\min}}{\eta} g_{\text{pf}_1}(\hat{\mathbf{q}}(t))$ , where  $g_{\text{pf}_1}$  is defined later in the proof. After that, using this inequality, we demonstrate that the stability region achieved under 'mdl' (using FSA) reaches at least a fraction  $\beta$  of the stability region achieved under 'pf<sub>1</sub>'. Finally, based on the last two steps and using the result in Lemma 7, we deduce the statement given in the theorem.

• **Step 1 (Proof of Lemma 7):**

*Proof.* We first point out that the process  $\hat{\mathbf{q}}(t)$ , as  $\mathbf{q}(t)$  and  $\tilde{\mathbf{q}}(t)$ , can be seen as a discrete time Markov chain evolving on a countable state space. Recall that  $\hat{q}_i(t) = \tilde{q}_i(t - d)$ , where  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b[$ . We also recall that a system is called strongly stable if every queue in the system is strongly stable. Define 'pf<sub>2</sub>' to be a similar system to 'pf<sub>1</sub>' except that under 'pf<sub>2</sub>' the scheduling is done based on  $\tilde{\mathbf{q}}(t)$ . It can be noticed that the proof of the theorem is similar to the proof of the following statement:  $\tilde{\mathbf{q}}(t)$  is strongly stable if and only if  $\mathbf{q}(t)$ ; we term this statement as "latter statement" and the one in the theorem as "former statement". This results from the fact that in the two systems 'pf<sub>1</sub>' and 'pf<sub>2</sub>' the scheduling is done based on outdated queue lengths information. Because the proof using  $\hat{\mathbf{q}}(t)$  is more complicated to present as compared to the one using  $\tilde{\mathbf{q}}(t)$ , we will only provide the proof of the latter statement, while noting that the proof of the former statement can be done in a similar way.

Let us first assume that  $\mathbf{q}(t)$  is strongly stable. One can notice that

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} \{\tilde{q}_i(t)\} \leq \frac{T_b}{T_b T} \sum_{t=0}^{T_b(T-1)} \mathbb{E} \{q_i(t)\}. \quad (4.69)$$

Thus, the following holds

$$\lim_{T \rightarrow +\infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} \{\tilde{q}_i(t)\} \leq T_b \lim_{T \rightarrow +\infty} \sup \frac{1}{T_b T} \sum_{t=0}^{T_b(T-1)} \mathbb{E} \{q_i(t)\} < +\infty, \quad (4.70)$$

where the second inequality follows from the assumption that  $\mathbf{q}(t)$  is strongly stable. Hence, we can deduce that  $\tilde{\mathbf{q}}(t)$  is strongly stable.

We now proceed to show the converse, that is, if  $\tilde{\mathbf{q}}(t)$  is strongly stable then  $\mathbf{q}(t)$  is also strongly stable. For all  $i \in \{1, \dots, N\}$ , we can write

$$\begin{aligned} \lim_{T_1 \rightarrow +\infty} \sup \frac{1}{T_1} \sum_{t=0}^{T_1-1} \mathbb{E} \{q_i(t)\} &= \lim_{T \rightarrow +\infty} \sup \frac{1}{T_b T} \sum_{t=0}^{T_b T-1} \mathbb{E} \{q_i(t)\} \\ &= \lim_{T \rightarrow +\infty} \sup \frac{1}{T_b T} \left[ \sum_{n=0}^{T-1} \mathbb{E} \{q_i(nT_b)\} + \sum_{n=0}^{T-1} \sum_{t'=1}^{T_b-1} \mathbb{E} \{q_i(nT_b + t')\} \right]. \end{aligned} \quad (4.71)$$



## 4.7. Appendix

---

Based on the facts that  $\tilde{\mathbf{q}}(t)$  is strongly stable and  $\tilde{q}_i(t) = q_i(nT_b)$  (for  $t \in [nT_b, (n+1)T_b]$ ), there exists a finite positive constant  $E_b$  such that,  $\forall i \in \{1, \dots, N\}$ , the following holds

$$\lim_{T \rightarrow +\infty} \frac{1}{T_b T} \sum_{n=0}^{T-1} \mathbb{E} \{q_i(nT_b)\} \leq \frac{E_b}{T_b}. \quad (4.72)$$

Also, note that  $q_i(nT_b + t') \leq q_i(nT_b) + t' A_{\max}$ ,  $\forall t' \in \{1, \dots, T_b - 1\}$ . This implies that

$$\sum_{t'=1}^{T_b-1} \mathbb{E} \{q_i(nT_b + t')\} \leq \mathbb{E} \{q_i(nT_b)\} + \frac{T_b(T_b - 1)}{2} A_{\max}. \quad (4.73)$$

Plugging the above into (4.71) yields

$$\begin{aligned} \lim_{T_1 \rightarrow +\infty} \sup \frac{1}{T_1} \sum_{t=0}^{T_1-1} \mathbb{E} \{q_i(t)\} &\leq \lim_{T \rightarrow +\infty} \sup \frac{1}{T_b T} \left[ 2 \sum_{n=0}^{T-1} \mathbb{E} \{q_i(nT_b)\} + T_b T \frac{T_b(T_b - 1)}{2} A_{\max} \right] \\ &\leq 2 \frac{E_b}{T_b} + \frac{(T_b - 1)^2}{2} A_{\max} < +\infty. \end{aligned} \quad (4.74)$$

Therefore,  $\mathbf{q}(t)$  is strongly stable. □

• **Step 2:** Here we want to prove that the following relation holds.

$$g_{\text{mdl}}(\hat{\mathbf{q}}(t)) \geq \frac{p_c^{\min}}{\eta} g_{\text{pf}_1}(\hat{\mathbf{q}}(t)). \quad (4.75)$$

To this end, we define 'mpf<sub>1</sub>' to be the system with full but delayed CSI (i.e. delay of  $d$  slots) at the BS, and where at time-slot  $t$  the Max-Weigh policy is used for scheduling based on  $\hat{\mathbf{q}}(t)$ . Under this system, the expected weighted throughput, which we denote by  $g_{\text{mpf}_1}$ , can be expressed as

$$g_{\text{mpf}_1}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{j=1}^L \sum_{i=1}^N \hat{q}_i(t) \hat{C}_{ij}(t) S_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t) \right\}. \quad (4.76)$$

We first show that  $g_{\text{mpf}_1}(\hat{\mathbf{q}}(t)) \geq \frac{p_c^{\min}}{\eta} g_{\text{pf}_1}(\hat{\mathbf{q}}(t))$ . It can be easily seen that under 'pf<sub>1</sub>' the expected weighted throughput, termed as  $g_{\text{pf}_1}$ , can be given by

$$g_{\text{pf}_1}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{j=1}^L \sum_{i=1}^N \hat{q}_i(t) C_{ij}(t) S_{ij}(t) \mid \hat{\mathbf{q}}(t) \right\}. \quad (4.77)$$

Expected weighted throughputs  $g_{\text{pf}_1}(\hat{\mathbf{q}}(t))$  and  $g_{\text{mpf}_1}(\hat{\mathbf{q}}(t))$  can be re-expressed as the following

$$g_{\text{pf}_1}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{pf}_1}(t)} \hat{q}_i(t) C_{ij}(t) \mid \hat{\mathbf{q}}(t) \right\}, \quad (4.78)$$

## 4.7. Appendix

---

$$g_{\text{mpf}_1}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mpf}_1}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t) \right\}, \quad (4.79)$$

where  $\mathcal{M}^{\text{pf}_1}(t)$  and  $\mathcal{M}^{\text{mpf}_1}(t)$  stand for the subsets of scheduled users under 'pf<sub>1</sub>' and 'mpf<sub>1</sub>', respectively, at time-slot  $t$ . Let  $\mathbf{h}$  be a vector that represents the fading of all the links at time-slot  $t$ . We also define  $\hat{\mathbf{h}}$  to be the fading of these links at time-slot  $t - d$ . Then, we can write

$$\begin{aligned} g_{\text{mpf}_1}(\hat{\mathbf{q}}(t)) &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \mathbb{E}_{\mathbf{h} \mid \hat{\mathbf{h}}} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mpf}_1}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t) \right\} \right\} \\ &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mpf}_1}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{P}\{C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t)\} \mid \hat{\mathbf{q}}(t) \right\}. \end{aligned} \quad (4.80)$$

Let us define  $p_c^{\min}$  as follows

$$p_c^{\min} = \min_{(ij)} p_{cij}^{\min}, \quad (4.81)$$

in which  $p_{cij}^{\min}$  is given by

$$p_{cij}^{\min} = \min_{t, \hat{h}_{ij}(t)} \left\{ \mathbb{P}\{C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t)\} \right\}, \quad (4.82)$$

where we recall that the value of  $\hat{C}_{ij}(t)$  depends on  $\hat{h}_{ij}(t)$ . Based on the expression of  $g_{\text{mpf}_1}(\hat{\mathbf{q}}(t))$  and the definition of  $p_c^{\min}$ , we get

$$\mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mpf}_1}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mid \hat{\mathbf{q}}(t) \right\} \leq \frac{g_{\text{mpf}_1}(\hat{\mathbf{q}}(t))}{p_c^{\min}}. \quad (4.83)$$

By defining  $\eta$  as

$$\eta = \max_{t, (ij)} \left\{ \frac{C_{ij}(t)}{\hat{C}_{ij}(t)} \right\}, \quad (4.84)$$

we obtain the following inequality

$$\mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{pf}_1}(t)} \hat{q}_i(t) C_{ij}(t) \mid \hat{\mathbf{q}}(t) \right\} \leq \eta \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mpf}_1}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mid \hat{\mathbf{q}}(t) \right\}. \quad (4.85)$$

This inequality can be proved as follows. Let  $i_1$  and  $i_2$  be the scheduled users over channel  $j$  under systems 'pf<sub>1</sub>' and 'mpf<sub>1</sub>', respectively. These users are selected according to the following

$$i_1 = \arg \max_i \{\hat{q}_i(t) C_{ij}(t)\}, \quad i_2 = \arg \max_i \{\hat{q}_i(t) \hat{C}_{ij}(t)\}. \quad (4.86)$$

Based on the above and the definition of  $\eta$ , we get

$$\hat{q}_{i_1}(t) C_{i_1 j}(t) \leq \hat{q}_{i_1}(t) \hat{C}_{i_1 j}(t) \eta \leq \hat{q}_{i_2}(t) \hat{C}_{i_2 j}(t) \eta. \quad (4.87)$$

## 4.7. Appendix

---

Hence, the inequality in (4.85) follows by summing over all the channels  $j \in \{1, \dots, L\}$ . Combining (4.83) and (4.85) yields

$$g_{\text{mpf}_1}(\hat{\mathbf{q}}(t)) \geq \frac{p_c^{\min}}{\eta} g_{\text{pf}_1}(\hat{\mathbf{q}}(t)). \quad (4.88)$$

Now, we show that  $g_{\text{mdl}}(\hat{\mathbf{q}}(t)) \geq g_{\text{mpf}_1}(\hat{\mathbf{q}}(t))$ . By denoting  $\mathcal{M}^{\text{mdl}}(t)$  as the subset of users scheduled for transmission under 'mdl' at time-slot  $t$ , we can rewrite  $g_{\text{mdl}}(\hat{\mathbf{q}}(t))$  as

$$g_{\text{mdl}}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{mdl}}(t)} \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t) \right\}. \quad (4.89)$$

We define  $g_{\text{mdl},j}(\hat{\mathbf{q}}(t))$  and  $g_{\text{mpf}_1,j}(\hat{\mathbf{q}}(t))$  as the following

$$g_{\text{mdl},j}(\mathbf{q}(t)) = \mathbb{E} \left\{ \hat{q}_{i_l}(t) \hat{C}_{i_l j}(t) \mathbb{1}_{(C_{i_l j}(t) \geq \hat{C}_{i_l j}(t))} \mid \hat{\mathbf{q}}(t) \right\}, \quad (4.90)$$

$$g_{\text{mpf}_1,j}(\hat{\mathbf{q}}(t)) = \mathbb{E} \left\{ \hat{q}_{i_f}(t) \hat{C}_{i_f j}(t) \mathbb{1}_{(C_{i_f j}(t) \geq \hat{C}_{i_f j}(t))} \mid \hat{\mathbf{q}}(t) \right\}, \quad (4.91)$$

where  $i_l$  and  $i_f$  denote the scheduled users over channel  $j$  under 'mdl' and 'mpf<sub>1</sub>', respectively. These users can be determined according to the following

$$i_l = \arg \max_i \left\{ \hat{q}_i(t) \hat{C}_{ij}(t) \mathbb{P}\{C_{ij}(t) \geq \hat{C}_{ij}(t) \mid \hat{h}_{ij}(t)\} \right\}, \quad (4.92)$$

$$i_f = \arg \max_i \left\{ \hat{q}_i(t) \hat{C}_{ij}(t) \right\}. \quad (4.93)$$

Based on the above, it can be seen that

$$\hat{q}_{i_l}(t) \hat{C}_{i_l j}(t) \mathbb{P}\{C_{i_l j}(t) \geq \hat{C}_{i_l j}(t) \mid \hat{h}_{i_l j}(t)\} \geq \hat{q}_{i_f}(t) \hat{C}_{i_f j}(t) \mathbb{P}\{C_{i_f j}(t) \geq \hat{C}_{i_f j}(t) \mid \hat{h}_{i_f j}(t)\}. \quad (4.94)$$

Hence, we get

$$\begin{aligned} g_{\text{mpf}_1,j}(\hat{\mathbf{q}}(t)) &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \mathbb{E}_{\mathbf{h} \mid \hat{\mathbf{h}}} \left\{ \max_i \left\{ \hat{q}_i(t) \hat{C}_{ij}(t) \right\} \mathbb{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))} \mid \hat{\mathbf{q}}(t) \right\} \right\} \\ &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \mathbb{E}_{\mathbf{h} \mid \hat{\mathbf{h}}} \left\{ \hat{q}_{i_f}(t) \hat{C}_{i_f j}(t) \mathbb{1}_{(C_{i_f j}(t) \geq \hat{C}_{i_f j}(t))} \mid \hat{\mathbf{q}}(t) \right\} \right\} \\ &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \hat{q}_{i_f}(t) \hat{C}_{i_f j}(t) \mathbb{P}\{C_{i_f j}(t) \geq \hat{C}_{i_f j}(t) \mid \hat{h}_{i_f j}(t)\} \mid \hat{\mathbf{q}}(t) \right\} \\ &\stackrel{(a)}{\leq} \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \hat{q}_{i_l}(t) \hat{C}_{i_l j}(t) \mathbb{P}\{C_{i_l j}(t) \geq \hat{C}_{i_l j}(t) \mid \hat{h}_{i_l j}(t)\} \mid \hat{\mathbf{q}}(t) \right\} \\ &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \mathbb{E}_{\mathbf{h} \mid \hat{\mathbf{h}}} \left\{ \hat{q}_{i_l}(t) \hat{C}_{i_l j}(t) \mathbb{1}_{(C_{i_l j}(t) \geq \hat{C}_{i_l j}(t))} \mid \hat{\mathbf{q}}(t) \right\} \right\} \\ &= g_{\text{mdl},j}(\hat{\mathbf{q}}(t)), \end{aligned} \quad (4.95)$$

where inequality (a) results from the relation in (4.94). By taking the sum over all the channels, we can deduce that  $g_{\text{mdl}}(\hat{\mathbf{q}}(t)) \geq g_{\text{mpf}_1}(\hat{\mathbf{q}}(t))$ .

## 4.7. Appendix

---

• **Step 3:** We now want to show that system 'mdl' achieves a fraction  $(1 - \frac{1}{T_b}) \frac{P_c^{\min}}{\eta}$  of the stability region of system 'pf<sub>1</sub>'. Let us first define  $D_i(t)$  as the service rate allocated for user  $i$  at time-slot  $t$ . In addition, we define the quadratic *Lyapunov function* as the following

$$Ly(\mathbf{x}) \triangleq \frac{1}{2} (\mathbf{x} \cdot \mathbf{x}) = \frac{1}{2} \sum_{i=1}^N x_i^2. \quad (4.96)$$

The evolution equation for queue  $q_i$ , for all  $i \in \{1, \dots, N\}$ , can be given as the following

$$q_i((n+1)T_b + d) = \max \left\{ q_i(nT_b + d) + \sum_{t_1=0}^{T_b-1} A_i(nT_b + d + t_1) - \sum_{t_1=1}^{T_b-1} D_i(nT_b + d + t_1), 0 \right\}, \quad (4.97)$$

where we note that the sum over the  $D_i$  starts from 1 because every  $T_b$  slots the BS uses the first slot, i.e. slot  $nT_b$ , to broadcast the queue lengths, meaning that no transmission occurs during this slot. For notational convenience we sometimes will replace  $nT_b + d$  by  $t_2$ , i.e.  $t_2 = nT_b + d$ .

From (4.97) we have

$$\begin{aligned} q_i^2(t_2 + T_b) &\leq q_i^2(t_2) + \left[ \sum_{t_1=0}^{T_b-1} A_i(t_2 + t_1) \right]^2 + \left[ \sum_{t_1=1}^{T_b-1} D_i(t_2 + t_1) \right]^2 \\ &\quad + 2q_i(t_2) \left[ \sum_{t_1=0}^{T_b-1} A_i(t_2 + t_1) - \sum_{t_1=1}^{T_b-1} D_i(t_2 + t_1) \right] \\ &\leq q_i^2(t_2) + T_b^2 A_{\max}^2 + (T_b - 1)^2 R_1^2 + 2q_i(t_2) \left[ \sum_{t_1=0}^{T_b-1} A_i(t_2 + t_1) - \sum_{t_1=1}^{T_b-1} D_i(t_2 + t_1) \right], \end{aligned} \quad (4.98)$$

where the first inequality results from the following fact: for any  $q \geq 0$ ,  $A \geq 0$ ,  $D \geq 0$ , we have

$$(\max \{q + A - D, 0\})^2 \leq q^2 + A^2 + D^2 + 2q(A - D).$$

The second inequality holds since we have  $A_i(t) \leq A_{\max}$  and  $D_i(t) \leq R_1$ ,  $\forall t$ ; we recall that  $R_1$  stands for the highest rate. From the above, setting  $E = \frac{1}{2} N T_b^2 A_{\max}^2 + \frac{1}{2} N (T_b - 1)^2 R_1^2$ , it follows that

$$\begin{aligned} Ly(\mathbf{q}((n+1)T_b + d)) - Ly(\mathbf{q}(nT_b + d)) &= Ly(\mathbf{q}(t_2 + T_b)) - Ly(\mathbf{q}(t_2)) \\ &= \frac{1}{2} \sum_{i=1}^N [q_i^2(t_2 + T_b) - q_i^2(t_2)] \\ &\leq E + q_i(t_2) \left[ \sum_{t_1=0}^{T_b-1} A_i(t_2 + t_1) - \sum_{t_1=1}^{T_b-1} D_i(t_2 + t_1) \right]. \end{aligned} \quad (4.99)$$

Let us define  $Dr(\mathbf{q}(nT_b))$  as the *conditional Lyapunov drift* for time instance  $nT_b$ :

$$Dr(\mathbf{q}(nT_b)) \triangleq \mathbb{E} \{ Ly(\mathbf{q}(t_2 + T_b)) - Ly(\mathbf{q}(t_2)) \mid \mathbf{q}(nT_b) \}. \quad (4.100)$$

## 4.7. Appendix

---

Using (4.99), we have that  $Dr(\mathbf{q}(nT_b))$  for a general scheduling policy satisfies

$$\begin{aligned} Dr(\mathbf{q}(nT_b)) &\leq E + \sum_{i=1}^N q_i(t_2) \sum_{t_1=0}^{T_b-1} \mathbb{E}\{A_i(t_2+t_1) \mid \mathbf{q}(nT_b)\} - \sum_{i=1}^N q_i(t_2) \sum_{t_1=1}^{T_b-1} \mathbb{E}\{D_i(t_2+t_1) \mid \mathbf{q}(nT_b)\} \\ &= E + T_b \sum_{i=1}^N q_i(t_2) a_i - \sum_{i=1}^N q_i(t_2) \sum_{t_1=1}^{T_b-1} \mathbb{E}\{D_i(t_2+t_1) \mid \mathbf{q}(nT_b)\}, \end{aligned} \quad (4.101)$$

where we have used the fact that arrivals are i.i.d. over time-slots and thus independent of current queue lengths, meaning that  $\mathbb{E}\{A_i(t_2+t_1) \mid \mathbf{q}(nT_b)\} = \mathbb{E}\{A_i(t_2+t_1)\} = a_i$ . Note that the conditional expectation at the right-hand-side of (4.101) is w.r.t. the randomly observed channel states. Let  $\Delta_{\text{mdl}}$  denote the scheduling policy under system 'mdl'. Also, we use  $\Delta_{\text{pf}}$  and  $\Delta_{\text{pf}_1}$  to denote the scheduling policies under 'pf' and 'pf<sub>1</sub>', respectively. For the drift under  $\Delta_{\text{mdl}}$  we have

$$Dr^{(\Delta_{\text{mdl}})}(\mathbf{q}(nT_b)) \leq E + T_b \sum_{i=1}^N q_i(t_2) a_i - \sum_{i=1}^N q_i(t_2) \sum_{t_1=1}^{T_b-1} \mathbb{E}\left\{D_i^{(\Delta_{\text{mdl}})}(t_2+t_1) \mid \mathbf{q}(nT_b)\right\}, \quad (4.102)$$

where  $D_i^{(\Delta_{\text{mdl}})}(t)$  is the service rate allocated for user  $i$  at time-slot  $t$  under system 'mdl' and its expression can be given as

$$D_i^{(\Delta_{\text{mdl}})}(t) = \sum_{j=1}^L \hat{C}_{ij}(t) S_{ij}(t) \hat{Y}_{ij}(t) \mathbf{1}_{(C_{ij}(t) \geq \hat{C}_{ij}(t))}. \quad (4.103)$$

Based on the evolution equation of the queue lengths and the facts that  $A_i(t) \leq A_{\max}$  and  $D_i(t) \leq R_1$ , it can be seen that

$$-(s_2 - s_1)R_1 \leq q_i(s_2) - q_i(s_1) \leq (s_2 - s_1)A_{\max}. \quad (4.104)$$

Thus, the following holds  $-dR_1 \leq q_i(t_2) - q_i(nT_b) \leq dA_{\max}$ , or equivalently

$$q_i(nT_b) - dR_1 \leq q_i(t_2) \leq q_i(nT_b) + dA_{\max}, \quad (4.105)$$

where we recall that  $t_2 = nT_b + d$ . Plugging the above into (4.102), we get

$$\begin{aligned} Dr^{(\Delta_{\text{mdl}})}(\mathbf{q}(nT_b)) &\leq E + T_b \sum_{i=1}^N (q_i(nT_b) + dA_{\max}) a_i \\ &\quad - \sum_{i=1}^N (q_i(nT_b) - dR_1) \sum_{t_1=1}^{T_b-1} \mathbb{E}\left\{D_i^{(\Delta_{\text{mdl}})}(t_2+t_1) \mid \mathbf{q}(nT_b)\right\} \\ &\leq E + T_b N d A_{\max}^2 + (T_b - 1) N d R_1^2 + T_b \sum_{i=1}^N q_i(nT_b) a_i \\ &\quad - \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E}\left\{D_i^{(\Delta_{\text{mdl}})}(t_2+t_1) \mid \mathbf{q}(nT_b)\right\}. \end{aligned} \quad (4.106)$$

## 4.7. Appendix

---

Recall that

$$g_{\text{mdl}}(\hat{\mathbf{q}}(t)) = \sum_{i=1}^N \hat{q}_i(t) \mathbb{E} \left\{ D_i^{(\Delta_{\text{mdl}})}(t) \mid \hat{\mathbf{q}}(t) \right\}. \quad (4.107)$$

By setting  $t = t_2 + t_1$ , we have  $\hat{\mathbf{q}}(t) = \mathbf{q}(nT_b)$ ,  $\forall t_1$  such that  $1 \leq t_1 \leq T_b - 1$ . This can be explained as follows. For time-slot  $t$ , under 'mdl' (using FSA) the feedback decision at time-slot  $t - d$  is done based on  $\hat{\mathbf{q}}(t)$ ; we recall that  $\hat{\mathbf{q}}(t) = \tilde{\mathbf{q}}(t - d)$ , where  $\tilde{\mathbf{q}}(t - d) = \mathbf{q}(nT_b)$ , for  $t - d \in [nT_b, (n+1)T_b]$ . For each channel, only one link will report its feedback to the BS at time-slot  $t - d$ , so evidently this link will be transmitting over this channel at time-slot  $t$ . Here  $t = t_2 + t_1 = nT_b + d + t_1$ , thus at time-slot  $t - d = nT_b + t_1$  the feedback decision is based on  $\hat{\mathbf{q}}(t) = \tilde{\mathbf{q}}(t - d) = \tilde{\mathbf{q}}(nT_b + t_1)$ . But  $\tilde{\mathbf{q}}(nT_b + t_1) = \mathbf{q}(nT_b)$ ,  $\forall t_1$  such that  $1 \leq t_1 \leq T_b - 1$ , since the BS broadcasts the queue lengths at time-slots  $0, T_b, \dots, nT_b, \dots$ . Hence, the following holds

$$\begin{aligned} g_{\text{mdl}}(\hat{\mathbf{q}}(t)) &= g_{\text{mdl}}(\mathbf{q}(nT_b)) \\ &= \sum_{i=1}^N q_i(nT_b) \mathbb{E} \left\{ D_i^{(\Delta_{\text{mdl}})}(t) \mid \mathbf{q}(nT_b) \right\}. \end{aligned} \quad (4.108)$$

On the other hand, we have

$$g_{\text{pf}_1}(\hat{\mathbf{q}}(t)) = \sum_{i=1}^N \hat{q}_i(t) \mathbb{E} \left\{ D_i^{(\Delta_{\text{pf}_1})}(t) \mid \hat{\mathbf{q}}(t) \right\}, \quad (4.109)$$

where  $D_i^{(\Delta_{\text{pf}_1})}(t) = \sum_{j=1}^L C_{ij}(t) S_{ij}(t)$ .

In a similar manner to 'mdl', it can be shown that

$$\begin{aligned} g_{\text{pf}_1}(\hat{\mathbf{q}}(t)) &= g_{\text{pf}_1}(\mathbf{q}(nT_b)) \\ &= \sum_{i=1}^N q_i(nT_b) \mathbb{E} \left\{ D_i^{(\Delta_{\text{pf}_1})}(t) \mid \mathbf{q}(nT_b) \right\}. \end{aligned} \quad (4.110)$$

Using the relation  $g_{\text{mdl}}(\hat{\mathbf{q}}(t)) \geq \frac{p_c^{\min}}{\eta} g_{\text{pf}_1}(\hat{\mathbf{q}}(t))$ , which was shown earlier, we get

$$\sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{mdl}})}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\} \geq \frac{p_c^{\min}}{\eta} \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{pf}_1})}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\}. \quad (4.111)$$

Plugging the above directly into (4.106) yields

$$Dr^{(\Delta_{\text{mdl}})}(\mathbf{q}(nT_b)) \leq E_1 + T_b \sum_{i=1}^N q_i(nT_b) a_i - \frac{p_c^{\min}}{\eta} \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{pf}_1})}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\}, \quad (4.112)$$

## 4.7. Appendix

---

in which

$$E_1 = E + T_b N d A_{\max}^2 + (T_b - 1) N d R_1^2. \quad (4.113)$$

Since, by definition, policy  $\Delta_{\text{pf}_1}$  maximizes the weighted sum  $\sum_{i=1}^N q_i(nT_b) D_i(t_2 + t_1)$  over all alternative decisions, we can write

$$\sum_{i=1}^N q_i(nT_b) D_i^{(\Delta_{\text{pf}_1})}(t_2 + t_1) \geq \sum_{i=1}^N q_i(nT_b) D_i^{(\Delta)}(t_2 + t_1), \quad (4.114)$$

in which  $\Delta$  represents any alternative (possibly randomized) scheduling decision that can stabilize system 'pf<sub>1</sub>'. Taking a conditional expectation of the above inequality (given  $\mathbf{q}(nT_b)$ ) yields

$$\sum_{i=1}^N q_i(nT_b) \mathbb{E} \left\{ D_i^{(\Delta_{\text{pf}_1})}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\} \geq \sum_{i=1}^N q_i(nT_b) \mathbb{E} \left\{ D_i^{(\Delta)}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\}, \quad (4.115)$$

where we recall that  $t_2 = nT_b + d$ . Plugging the above into (4.112) yields

$$\begin{aligned} Dr^{(\Delta_{\text{mdl}})}(\mathbf{q}(nT_b)) &\leq E_1 + T_b \sum_{i=1}^N q_i(nT_b) a_i - \frac{p_c^{\min}}{\eta} \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta)}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\} \\ &= E_1 + T_b \sum_{i=1}^N q_i(nT_b) a_i - \frac{p_c^{\min}}{\eta} \sum_{i=1}^N q_i(nT_b) \mathbb{E} \left\{ \sum_{t_1=1}^{T_b-1} D_i^{(\Delta)}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\}. \end{aligned} \quad (4.116)$$

We now consider a particular policy  $\Delta$  that depends only on the channels states. It is worth recalling that here each channel process is not i.i.d. in time. We point out that the process  $D_i^{(\Delta)}(t)$ , defined over the *channel convergent* (fading) process, is *rate convergent* [28]. Using the definition of rate convergence for  $D_i^{(\Delta)}(t)$  with rate  $r_i^{(\Delta)}$ , we have (refer to [28] for more details)

$$\frac{1}{T_{\text{rc}}} \sum_{t=0}^{T_{\text{rc}}-1} D_i^{(\Delta)}(t) \rightarrow r_i^{(\Delta)}, \text{ with probability 1 as } T_{\text{rc}} \rightarrow \infty, \quad (4.117)$$

and for any  $\epsilon_i > 0$  there exists an interval  $T_{\text{rc}}$  such that, for all initial times  $t_0$ , and regardless of past history, the following holds

$$\left| r_i^{(\Delta)} - \mathbb{E} \left\{ \frac{1}{T_{\text{rc}}} \sum_{t=t_0}^{t_0+T_{\text{rc}}-1} D_i^{(\Delta)}(t) \right\} \right| \leq \epsilon_i. \quad (4.118)$$

By choosing  $T_{\text{rc}} = T_b - 1$  and  $t_0 = 1$ , we can deduce that

$$\mathbb{E} \left\{ \frac{1}{T_b - 1} \sum_{t_1=1}^{T_b-1} D_i^{(\Delta)}(t_1 + t_2) \mid \mathbf{q}(nT_b) \right\} \geq r_i^{(\Delta)} - \epsilon_i. \quad (4.119)$$

Note that  $T_b$  needs to be sufficiently large in order to get the  $\epsilon_i$  sufficiently small. Plugging the above

## 4.7. Appendix

---

into (4.116) yields

$$\begin{aligned} Dr^{(\Delta_{\text{mdl}})}(\mathbf{q}(nT_b)) &\leq E_1 + T_b \sum_{i=1}^N q_i(nT_b) a_i - \frac{p_c^{\min}}{\eta} (T_b - 1) \sum_{i=1}^N q_i(nT_b) \left[ r_i^{(\Delta)} - \epsilon_i \right] \\ &\leq E_1 - (T_b - 1) \frac{p_c^{\min}}{\eta} \sum_{i=1}^N q_i(nT_b) \left[ r_i^{(\Delta)} - \beta^{-1} a_i + \epsilon \right], \end{aligned} \quad (4.120)$$

where  $\beta = \left(1 - \frac{1}{T_b}\right) \frac{p_c^{\min}}{\eta}$  and  $\epsilon = \max_i \epsilon_i$ .

Recall that  $\Lambda_{\text{pf}_1}$  represents the stability region achieved under system 'pf<sub>1</sub>'. Let us suppose that the mean arrival rate vector  $\mathbf{a}$  is interior to fraction  $\beta$  of region  $\Lambda_{\text{pf}_1}$ , meaning that there exists an  $\epsilon_{\max}$ , which clearly depends on  $\mathbf{a}$ , such that

$$(a_1 + \epsilon_{\max}, \dots, a_N + \epsilon_{\max}) \in \beta \Lambda_{\text{pf}_1}. \quad (4.121)$$

The above can be equivalently re-expressed as

$$(\beta^{-1} a_1 + \beta^{-1} \epsilon_{\max}, \dots, \beta^{-1} a_N + \beta^{-1} \epsilon_{\max}) \in \Lambda_{\text{pf}_1}. \quad (4.122)$$

Thus, the following holds

$$r_i^{(\Delta)} \geq \beta^{-1} a_i + \beta^{-1} \epsilon_{\max}, \forall i \in \{1, \dots, N\}. \quad (4.123)$$

Plugging the above inequality into (4.120) yields

$$Dr^{(\Delta_{\text{mdl}})}(\mathbf{q}(nT_b)) \leq E_1 - T_b(\epsilon_{\max} + \beta\epsilon) \sum_{i=1}^N q_i(nT_b). \quad (4.124)$$

Let  $\check{\epsilon}_{\max} = T_b(\epsilon_{\max} + \beta\epsilon)$ . Taking an expectation of  $Dr^{(\Delta_{\text{mdl}})}$  over the randomness of the queue lengths and summing over  $n \in \{0, 1, \dots, T-1\}$  for some integer  $T > 0$ , we get

$$\mathbb{E}\{Ly(\mathbf{q}(TT_b + d))\} - \mathbb{E}\{Ly(\mathbf{q}(d))\} \leq E_1 T - \check{\epsilon}_{\max} \sum_{n=0}^{T-1} \sum_{i=1}^N \mathbb{E}\{q_i(nT_b)\}. \quad (4.125)$$

Rearranging terms, dividing by  $\check{\epsilon}_{\max} T$ , and using the fact that  $Ly(\mathbf{q}(TT_b + d)) \geq 0$  yields

$$\frac{1}{T} \sum_{n=0}^{T-1} \sum_{i=1}^N \mathbb{E}\{q_i(nT_b)\} \leq \frac{E_1}{\check{\epsilon}_{\max}} + \frac{\mathbb{E}\{Ly(\mathbf{q}(d))\}}{\check{\epsilon}_{\max} T}. \quad (4.126)$$

Assuming that  $\mathbb{E}\{Ly(\mathbf{q}(d))\} < \infty$  and taking a limsup, we eventually obtain

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{i=1}^N \mathbb{E}\{q_i(nT_b)\} \leq \frac{E_1}{\check{\epsilon}_{\max}}. \quad (4.127)$$

Based on the above inequality and the definition of strong stability, we can claim that  $\Delta_{\text{mdl}}$  stabilizes any mean arrival rate vector interior to fraction  $\beta$  of the stability region achieved under policy  $\Delta_{\text{pf}_1}$ ,



## 4.7. Appendix

---

meaning that  $\Delta_{\text{mdl}}$  stabilizes the system for any arrivals such that  $\mathbf{a} \in \beta \Lambda_{\text{pf}_1}$ .

Based on Lemma 7, we know that 'pf<sub>1</sub>' and 'pf' have the same stability region. Hence, we can deduce that  $\Delta_{\text{mdl}}$  stabilizes any arrival rate vector interior to fraction  $\beta$  of the stability region achieved under  $\Delta_{\text{pf}}$ . This completes the proof.

### 4.7.2 Proof of Theorem 15

The proof of this theorem is in the same spirit as the proof of Theorem 14.

The proof consists of three steps. We first recall the statement of Lemma 8 and give a remark about its proof. Then, we show that  $g_{\text{re}}(\tilde{\mathbf{q}}(t)) \geq \eta_r g_{\text{rp}_1}(\tilde{\mathbf{q}}(t))$ , where  $g_{\text{rp}_1}$  is defined later in the proof. After that, we show that the stability region achieved under 're' (using RW) reaches at least a fraction  $\eta_r$  of the stability region achieved under 'rp<sub>1</sub>'. Finally, based on the last two steps and using the result in Lemma 8, the statement given in the theorem can be deduced.

- **Step 1 (Proof of Lemma 8):** Recall that this lemma states that system 'rp<sub>1</sub>' is strongly stable if and only if system 'rp' is strongly stable. The proof of this statement is very similar to the proof of Lemma 7 (which can be found in Appendix 4.7.1) and is thus omitted to avoid repetition.
- **Step 2:** We here want to prove that  $g_{\text{re}}(\tilde{\mathbf{q}}(t)) \geq \eta_r g_{\text{rp}_1}(\tilde{\mathbf{q}}(t))$ . Recall that  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b[$  and  $\tilde{\mathbf{q}}(t) = (\tilde{q}_1(t), \dots, \tilde{q}_N(t))$ . It can be seen that under 'rp<sub>1</sub>' the expected weighted throughput, termed as  $g_{\text{rp}_1}$ , can be given as follows

$$g_{\text{rp}_1}(\tilde{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{rp}_1}(t)} \tilde{q}_i(t) C_{ij}(t) \mid \tilde{\mathbf{q}}(t) \right\}, \quad (4.128)$$

where  $\mathcal{M}^{\text{rp}_1}(t)$  stands for the subset of scheduled users under 'rp<sub>1</sub>' at time-slot  $t$ . Recall that the weighted throughput  $g_{\text{re}}(\tilde{\mathbf{q}}(t))$  is

$$g_{\text{re}}(\tilde{\mathbf{q}}(t)) = \mathbb{E} \left\{ \sum_{(ij) \in \mathcal{M}^{\text{re}}(t)} \tilde{q}_i(t) \check{C}_{ij}(t) \mathbb{1}_{(C_{ij}(t) \geq \check{C}_{ij}(t))} \mid \tilde{\mathbf{q}}(t) \right\}. \quad (4.129)$$

Let  $i_{\text{p}_1}$  and  $i_e$  be the scheduled users at time-slot  $t$  and over channel  $j$  under systems 'rp<sub>1</sub>' and 're', respectively. Define  $g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t))$  and  $g_{\text{re},j}(\tilde{\mathbf{q}}(t))$  as

$$g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t)) = \mathbb{E} \left\{ \tilde{q}_{i_{\text{p}_1}}(t) C_{i_{\text{p}_1}j}(t) \mid \tilde{\mathbf{q}}(t) \right\}, \quad (4.130)$$

$$g_{\text{re},j}(\tilde{\mathbf{q}}(t)) = \mathbb{E} \left\{ \tilde{q}_{i_e}(t) \check{C}_{i_ej}(t) \mathbb{1}_{(C_{i_ej}(t) \geq \check{C}_{i_ej}(t))} \mid \tilde{\mathbf{q}}(t) \right\}. \quad (4.131)$$

Let  $\mathcal{E}_j$  represent the event that occurs if on channel  $j$  systems 'rp<sub>1</sub>' and 're' have the same scheduling decision, i.e.  $i_{\text{p}_1} = i_e$ . Also, let  $\bar{\mathcal{E}}_j$  be the complementary event of  $\mathcal{E}_j$ , i.e.  $\bar{\mathcal{E}}_j$  occurs when  $i_{\text{p}_1} \neq i_e$ . We denote the probability of event  $\mathcal{E}_j$  to happen as  $\mathbb{P}\{\mathcal{E}_j\}$ .

To make this part of the proof easier to understand, we first consider that there are only two users in

## 4.7. Appendix

---

the system, denoted by indexes 1 and 2, we then present the desired result and finally we generalize this result to the  $N$  users case. Let  $\mathcal{E}_{1j}$  be the event that occurs if both 'rp<sub>1</sub>' and 're' schedule user 1 on channel  $j$  for transmission at time-slot  $t$ . Similarly, define  $\mathcal{E}_{2j}$  to be the event that occurs if both 'rp<sub>1</sub>' and 're' schedule user 2 on channel  $j$ . It can be noticed that  $\mathcal{E}_j$  is nothing but the union of  $\mathcal{E}_{1j}$  and  $\mathcal{E}_{2j}$ , i.e.  $\mathcal{E}_j = \mathcal{E}_{2j} \cup \mathcal{E}_{1j}$ . Hence,  $g_{re,j}(\tilde{\mathbf{q}}(t))$  can be rewritten as

$$\begin{aligned} g_{re,j}(\tilde{\mathbf{q}}(t)) &= \mathbb{E} \left\{ \tilde{q}_1(t) \check{C}_{1j}(t) \mathbb{1}_{(C_{1j}(t) \geq \check{C}_{1j}(t))} \mid \tilde{\mathbf{q}}(t), \mathcal{E}_{1j} \right\} \mathbb{P}\{\mathcal{E}_{1j}\} + \\ &\quad \mathbb{E} \left\{ \tilde{q}_2(t) \check{C}_{2j}(t) \mathbb{1}_{(C_{2j}(t) \geq \check{C}_{2j}(t))} \mid \tilde{\mathbf{q}}(t), \mathcal{E}_{2j} \right\} \mathbb{P}\{\mathcal{E}_{2j}\} + \\ &\quad \mathbb{E} \left\{ \tilde{q}_{i_e}(t) \check{C}_{i_e j}(t) \mathbb{1}_{(C_{i_e j}(t) \geq \check{C}_{i_e j}(t))} \mid \tilde{\mathbf{q}}(t), \bar{\mathcal{E}}_j \right\} \mathbb{P}\{\bar{\mathcal{E}}_j\}, \end{aligned} \quad (4.132)$$

where we note that here  $i_e$ , as  $i_{p_1}$ , can be 1 or 2. Note that in the above expression all the terms are non-negative. We also note that in the first term we have the same scheduling decision for both 'rp<sub>1</sub>' and 're', thus  $C_{1j}(t) = \check{C}_{1j}(t)$  since we work under the single rate level assumption, which implies that the indicator function  $\mathbb{1}_{(C_{1j}(t) \geq \check{C}_{1j}(t))} = 1$ . In a similar way, we can show that  $\mathbb{1}_{(C_{2j}(t) \geq \check{C}_{2j}(t))} = 1$ . Based on the above observations, we get

$$\begin{aligned} g_{re,j}(\tilde{\mathbf{q}}(t)) &\geq \mathbb{E} \left\{ \tilde{q}_1(t) \check{C}_{1j}(t) \mid \tilde{\mathbf{q}}(t), \mathcal{E}_{1j} \right\} \mathbb{P}\{\mathcal{E}_{1j}\} + \mathbb{E} \left\{ \tilde{q}_2(t) \check{C}_{2j}(t) \mid \tilde{\mathbf{q}}(t), \mathcal{E}_{2j} \right\} \mathbb{P}\{\mathcal{E}_{2j}\} \\ &= \mathbb{E} \left\{ \tilde{q}_1(t) C_{1j}(t) \mid \tilde{\mathbf{q}}(t), \mathcal{E}_{1j} \right\} \mathbb{P}\{\mathcal{E}_{1j}\} + \mathbb{E} \left\{ \tilde{q}_2(t) C_{2j}(t) \mid \tilde{\mathbf{q}}(t), \mathcal{E}_{2j} \right\} \mathbb{P}\{\mathcal{E}_{2j}\}. \end{aligned} \quad (4.133)$$

Under system 'rp<sub>1</sub>', user 1 is scheduled for transmission if  $\tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t)$ . This same user is scheduled under 're' if  $\tilde{q}_1(t) \check{C}_{1j}(t) \geq \tilde{q}_2(t) \check{C}_{2j}(t)$ . Thus,  $\mathbb{P}\{\mathcal{E}_{1j}\}$  can be expressed as

$$\mathbb{P}\{\mathcal{E}_{1j}\} = \mathbb{P} \left\{ \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t), \tilde{q}_1(t) \check{C}_{1j}(t) \geq \tilde{q}_2(t) \check{C}_{2j}(t) \mid \tilde{\mathbf{q}}(t) \right\}. \quad (4.134)$$

Similarly, we have

$$\mathbb{P}\{\mathcal{E}_{2j}\} = \mathbb{P} \left\{ \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t), \tilde{q}_2(t) \check{C}_{2j}(t) \geq \tilde{q}_1(t) \check{C}_{1j}(t) \mid \tilde{\mathbf{q}}(t) \right\}. \quad (4.135)$$

For notational convenience, we will omit the condition (that vector  $\tilde{\mathbf{q}}(t)$  is known) from the expressions of  $\mathbb{P}\{\mathcal{E}_{1j}\}$  and  $\mathbb{P}\{\mathcal{E}_{2j}\}$ . Using Bayes' theorem yields

$$\begin{aligned} \mathbb{P}\{\mathcal{E}_{1j}\} &= \mathbb{P} \left\{ \tilde{q}_1(t) \check{C}_{1j}(t) \geq \tilde{q}_2(t) \check{C}_{2j}(t) \mid \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t) \right\} \mathbb{P} \left\{ \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t) \right\}, \\ \mathbb{P}\{\mathcal{E}_{2j}\} &= \mathbb{P} \left\{ \tilde{q}_2(t) \check{C}_{2j}(t) \geq \tilde{q}_1(t) \check{C}_{1j}(t) \mid \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t) \right\} \mathbb{P} \left\{ \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t) \right\}. \end{aligned} \quad (4.136)$$

Given that  $\tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t)$ , in order to have  $\tilde{q}_1(t) \check{C}_{1j}(t) \geq \tilde{q}_2(t) \check{C}_{2j}(t)$  it suffices that  $\check{C}_{1j}(t) > 0$ . This can be explained as follows:

- ◊ Under system 'rp<sub>1</sub>', if user 1 (for example) is scheduled for transmission and the decision is to transmit without relaying, i.e.  $\gamma_{1j}^{(d)}(t) \geq \tau$  and thus  $C_{1j}(t) = C_{1j}^{(d)}(t)$ , then it is straightforward to see that this same user will also be scheduled under system 're'; recall that in this proof we assume that there is only one rate level, so here  $\check{C}_{1j}(t) = C_{1j}^{(d)}(t) = R > 0$ .
- ◊ If user 1 is scheduled for transmission under system 'rp<sub>1</sub>' and the decision is to transmit with

## 4.7. Appendix

---

relaying, i.e.  $\gamma_{1j}^{(d)}(t) < \tau$ ,  $\gamma_{1j}^{(r)}(t) \geq \tau$ , and thus  $C_{1j}(t) = C_{1j}^{(r)}(t)$ , then the sufficient conditions for user 1 to be scheduled under system 're' is to have  $\check{\gamma}_{1j}^{(r)}(t) \geq \tau$  and  $\gamma_j^{(sr)}(t) \geq \delta$ , which means that  $\check{C}_{1j}(t) > 0$ ; since there is only one single rate level, it is clear that here we have  $\check{C}_{1j}(t) = \check{C}_{1j}^{(r)}(t) = C_{1j}^{(r)}(t) = \frac{1}{2}R$ .

Based on the above, we obtain

$$\mathbb{P} \left\{ \tilde{q}_1(t) \check{C}_{1j}(t) \geq \tilde{q}_2(t) \check{C}_{2j}(t) \mid \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t) \right\} = \mathbb{P} \left\{ \check{C}_{1j}(t) > 0 \right\}. \quad (4.137)$$

In a similar way, we can show that

$$\mathbb{P} \left\{ \tilde{q}_2(t) \check{C}_{2j}(t) \geq \tilde{q}_1(t) \check{C}_{1j}(t) \mid \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t) \right\} = \mathbb{P} \left\{ \check{C}_{2j}(t) > 0 \right\}. \quad (4.138)$$

Plugging the above into (4.136) and using (4.133) yields

$$\begin{aligned} g_{\text{re},j}(\tilde{\mathbf{q}}(t)) &\geq \mathbb{E} \left\{ \tilde{q}_1(t) C_{1j}(t) \mid \tilde{\mathbf{q}}(t) \right\} \mathbb{P} \left\{ \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t) \right\} \mathbb{P} \left\{ \check{C}_{1j}(t) > 0 \right\} + \\ &\quad \mathbb{E} \left\{ \tilde{q}_2(t) C_{2j}(t) \mid \tilde{\mathbf{q}}(t) \right\} \mathbb{P} \left\{ \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t) \right\} \mathbb{P} \left\{ \check{C}_{2j}(t) > 0 \right\} \\ &\geq \mathbb{E} \left\{ \tilde{q}_1(t) C_{1j}(t) \mid \tilde{\mathbf{q}}(t) \right\} \mathbb{P} \left\{ \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t) \right\} \min_{i \in \{1,2\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\} + \\ &\quad \mathbb{E} \left\{ \tilde{q}_2(t) C_{2j}(t) \mid \tilde{\mathbf{q}}(t) \right\} \mathbb{P} \left\{ \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t) \right\} \min_{i \in \{1,2\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\}. \end{aligned} \quad (4.139)$$

On the other side, considering the possible scheduling decisions,  $g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t))$  can be re-expressed as

$$\begin{aligned} g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t)) &= \mathbb{E} \left\{ \tilde{q}_1(t) C_{1j}(t) \mid \tilde{\mathbf{q}}(t) \right\} \mathbb{P} \left\{ \tilde{q}_1(t) C_{1j}(t) \geq \tilde{q}_2(t) C_{2j}(t) \right\} + \\ &\quad \mathbb{E} \left\{ \tilde{q}_2(t) C_{2j}(t) \mid \tilde{\mathbf{q}}(t) \right\} \mathbb{P} \left\{ \tilde{q}_2(t) C_{2j}(t) \geq \tilde{q}_1(t) C_{1j}(t) \right\}. \end{aligned} \quad (4.140)$$

From the above equations, we can deduce that

$$g_{\text{re},j}(\tilde{\mathbf{q}}(t)) \geq \min_{i \in \{1,2\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\} g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t)). \quad (4.141)$$

The above result can be easily generalized to the  $N$  users case, and we get

$$g_{\text{re},j}(\tilde{\mathbf{q}}(t)) \geq \min_{i \in \{1, \dots, N\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\} g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t)). \quad (4.142)$$

Note that the above result is for (any) channel  $j$  and that

$$g_{\text{re}}(\tilde{\mathbf{q}}(t)) = \sum_{j=1}^L g_{\text{re},j}(\tilde{\mathbf{q}}(t)) \text{ and } g_{\text{rp}_1}(\tilde{\mathbf{q}}(t)) = \sum_{j=1}^L g_{\text{rp}_1,j}(\tilde{\mathbf{q}}(t)) \quad (4.143)$$

Therefore, the following result holds

$$g_{\text{re}}(\tilde{\mathbf{q}}(t)) \geq \min_{j \in \{1, \dots, L\}} \left\{ \min_{i \in \{1, \dots, N\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\} \right\} g_{\text{rp}_1}(\tilde{\mathbf{q}}(t)). \quad (4.144)$$

## 4.7. Appendix

---

By defining  $\eta_r$  as

$$\eta_r = \min_{j \in \{1, \dots, L\}} \left\{ \min_{i \in \{1, \dots, N\}} \left\{ \mathbb{P} \left\{ \check{C}_{ij}(t) > 0 \right\} \right\} \right\}, \quad (4.145)$$

the desired result follows, i.e.  $g_{\text{re}}(\tilde{\mathbf{q}}(t)) \geq \eta_r g_{\text{rp}_1}(\tilde{\mathbf{q}}(t))$ .

• **Step 3:** We now want to show that the stability region of 're' is at least a fraction  $\beta_r = (1 - \frac{1}{T_b})\eta_r$  of the stability region of system 'rp<sub>1</sub>'. It is worth noting that this part of the proof is along the lines of Step 3 in the proof of Theorem 14.

Define the quadratic *Lyapunov function* as

$$Ly(\mathbf{x}) \triangleq \frac{1}{2} (\mathbf{x} \cdot \mathbf{x}) = \frac{1}{2} \sum_{i=1}^N x_i^2. \quad (4.146)$$

Let  $D_i(t)$  be the service rate allocated for user  $i$  at time-slot  $t$ . The evolution equation for queue  $q_i$ , for all  $i \in \{1, \dots, N\}$ , can be given as

$$q_i((n+1)T_b) = \max \left\{ q_i(nT_b + d) + \sum_{t_1=0}^{T_b-1} A_i(nT_b + t_1) - \sum_{t_1=1}^{T_b-1} D_i(nT_b + t_1), 0 \right\}, \quad (4.147)$$

where we note that the sum over the  $D_i$  starts from 1 because every  $T_b$  slots the BS uses the first slot, i.e. slot  $nT_b$ , to broadcast the queue lengths, i.e. no transmission occurs during this slot.

From the evolution equation we have

$$\begin{aligned} q_i^2((n+1)T_b) &\leq q_i^2(nT_b) + \left[ \sum_{t_1=0}^{T_b-1} A_i(nT_b + t_1) \right]^2 + \left[ \sum_{t_1=1}^{T_b-1} D_i(nT_b + t_1) \right]^2 \\ &\quad + 2q_i(nT_b) \left[ \sum_{t_1=0}^{T_b-1} A_i(nT_b + t_1) - \sum_{t_1=1}^{T_b-1} D_i(nT_b + t_1) \right] \\ &\leq q_i^2(nT_b) + T_b^2 A_{\max}^2 + (T_b - 1)^2 R^2 + 2q_i(nT_b) \left[ \sum_{t_1=0}^{T_b-1} A_i(nT_b + t_1) - \sum_{t_1=1}^{T_b-1} D_i(nT_b + t_1) \right], \end{aligned} \quad (4.148)$$

in which the first inequality follows since: for any  $q \geq 0$ ,  $A \geq 0$ ,  $D \geq 0$ , we have

$$(\max \{q + A - D, 0\})^2 \leq q^2 + A^2 + D^2 + 2q(A - D).$$

The second inequality holds since we have  $A_i(t) \leq A_{\max}$  and  $D_i(t) \leq R$ ,  $\forall t$ .

Setting  $E = \frac{1}{2} N T_b^2 A_{\max}^2 + \frac{1}{2} N (T_b - 1)^2 R^2$ , it follows that

$$\begin{aligned} Ly(\mathbf{q}((n+1)T_b)) - Ly(\mathbf{q}(nT_b)) &= \frac{1}{2} \sum_{i=1}^N [q_i^2(nT_b + T_b) - q_i^2(nT_b)] \\ &\leq E + q_i(nT_b) \left[ \sum_{t_1=0}^{T_b-1} A_i(nT_b + t_1) - \sum_{t_1=1}^{T_b-1} D_i(nT_b + t_1) \right]. \end{aligned} \quad (4.149)$$

## 4.7. Appendix

---

Define  $Dr(\mathbf{q}(nT_b))$  to be the *conditional Lyapunov drift* for time instance  $nT_b$ :

$$Dr(\mathbf{q}(nT_b)) \triangleq \mathbb{E} \{Ly(\mathbf{q}(nT_b + T_b)) - Ly(\mathbf{q}(nT_b)) \mid \mathbf{q}(nT_b)\}. \quad (4.150)$$

Based on (4.149), for a general scheduling policy  $Dr(\mathbf{q}(nT_b))$  satisfies

$$Dr(\mathbf{q}(nT_b)) \leq E + T_b \sum_{i=1}^N q_i(nT_b) a_i - \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \{D_i(nT_b + t_1) \mid \mathbf{q}(nT_b)\}, \quad (4.151)$$

where we have used the fact that arrivals are i.i.d. over time-slots and thus independent of current queue lengths, i.e.  $\mathbb{E} \{A_i(nT_b + t_1) \mid \mathbf{q}(nT_b)\} = a_i$ . Let  $\Delta_{\text{re}}$  denote the scheduling policy under system 're'. We also use  $\Delta_{\text{rp}}$  and  $\Delta_{\text{rp}_1}$  to denote the scheduling policies under 'rp' and 'rp<sub>1</sub>', respectively. For the drift under policy  $\Delta_{\text{re}}$  we have

$$Dr^{(\Delta_{\text{re}})}(\mathbf{q}(nT_b)) \leq E + T_b \sum_{i=1}^N q_i(nT_b) a_i - \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{re}})}(nT_b + t_1) \mid \mathbf{q}(nT_b) \right\}, \quad (4.152)$$

in which  $D_i^{(\Delta_{\text{re}})}(t)$  is nothing but the service rate allocated for user  $i$  at time-slot  $t$  under system 're' (using algorithm RW). It is easy to see that  $g_{\text{re}}(\tilde{\mathbf{q}}(t))$  can be rewritten as

$$g_{\text{re}}(\tilde{\mathbf{q}}(t)) = \sum_{i=1}^N \tilde{q}_i(t) \mathbb{E} \left\{ D_i^{(\Delta_{\text{re}})}(t) \mid \tilde{\mathbf{q}}(t) \right\}. \quad (4.153)$$

Similarly, we can rewrite  $g_{\text{rp}_1}(\tilde{\mathbf{q}}(t))$  as

$$g_{\text{rp}_1}(\tilde{\mathbf{q}}(t)) = \sum_{i=1}^N \tilde{q}_i(t) \mathbb{E} \left\{ D_i^{(\Delta_{\text{rp}_1})}(t) \mid \tilde{\mathbf{q}}(t) \right\}, \quad (4.154)$$

where  $D_i^{(\Delta_{\text{rp}_1})}(t)$  is the service rate allocated for user  $i$  at time-slot  $t$  under system 'rp<sub>1</sub>'. Recall that  $\tilde{q}_i(t) = q_i(nT_b)$  for  $t \in [nT_b, (n+1)T_b]$ . Using the relation  $g_{\text{re}}(\tilde{\mathbf{q}}(t)) \geq \eta_r g_{\text{rp}_1}(\tilde{\mathbf{q}}(t))$ , which was shown earlier in this proof, the following holds

$$\sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{re}})}(nT_b + t_1) \mid \mathbf{q}(nT_b) \right\} \geq \eta_r \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{rp}_1})}(nT_b + t_1) \mid \mathbf{q}(nT_b) \right\}. \quad (4.155)$$

Plugging the above into (4.152) yields

$$Dr^{(\Delta_{\text{re}})}(\mathbf{q}(nT_b)) \leq E + T_b \sum_{i=1}^N q_i(nT_b) a_i - \eta_r \sum_{i=1}^N q_i(nT_b) \sum_{t_1=1}^{T_b-1} \mathbb{E} \left\{ D_i^{(\Delta_{\text{rp}_1})}(nT_b + t_1) \mid \mathbf{q}(nT_b) \right\}. \quad (4.156)$$

Furthermore, we can claim that  $\mathbb{E} \left\{ D_i^{(\Delta_{\text{rp}_1})}(nT_b + t_1) \mid \mathbf{q}(nT_b) \right\}$  is independent of the time index. This statement results from the fact that  $D_i^{(\Delta_{\text{rp}_1})}(nT_b + t_1)$  has the same distribution for every slot  $t_1$

## 4.7. Appendix

---

such that  $nT_b + 1 \leq t_1 \leq nT_b + T_b - 1$ , because the rate  $D_i^{(\Delta_{rp_1})}(nT_b + t_1)$  in the slot depends only on the queue length vector  $\mathbf{q}(nT_b)$  and the realization of the channels. Hence, the inequality in (4.156) can be re-expressed as

$$Dr^{(\Delta_{re})}(\mathbf{q}(nT_b)) \leq E + T_b \sum_{i=1}^N q_i(nT_b) a_i - (T_b - 1) \eta_r \sum_{i=1}^N q_i(nT_b) \mathbb{E} \left\{ D_i^{(\Delta_{rp_1})}(t_2 + t_1) \mid \mathbf{q}(nT_b) \right\}. \quad (4.157)$$

Define  $r_i^{(\Delta_{rp_1})}(nT_b) = \mathbb{E} \left\{ D_i^{(\Delta_{rp_1})}(nT_b + t_1) \mid \mathbf{q}(nT_b) \right\}$ . By definition, policy  $\Delta_{rp_1}$  maximizes the weighted sum  $\sum_{i=1}^N q_i(nT_b) D_i(nT_b + t_1)$  over all alternative decisions, we thus can write

$$\sum_{i=1}^N q_i(nT_b) D_i^{(\Delta_{rp_1})}(nT_b + t_1) \geq \sum_{i=1}^N q_i(nT_b) D_i^{(\Delta)}(nT_b + t_1), \quad (4.158)$$

where  $\Delta$  represents any alternative (possibly randomized) scheduling policy that can stabilize system 'rp<sub>1</sub>'. Fixing a particular alternative decision  $\Delta$  for comparison and taking a conditional expectation of the above inequality (given  $\mathbf{q}(nT_b)$ ) yields

$$\sum_{i=1}^N q_i(nT_b) r_i^{(\Delta_{rp_1})}(nT_b) \geq \sum_{i=1}^N q_i(nT_b) r_i^{(\Delta)}(nT_b), \quad (4.159)$$

in which  $r_i^{(\Delta)}(nT_b) = \mathbb{E} \{ D_i^{(\Delta)}(nT_b + t_1) \mid \mathbf{q}(nT_b) \}$ . Plugging the above into (4.157) yields

$$Dr^{(\Delta_{re})}(\mathbf{q}(nT_b)) \leq E - (T_b - 1) \eta_r \sum_{i=1}^N q_i(nT_b) \left[ r_i^{(\Delta)}(nT_b) - \beta_r^{-1} a_i \right], \quad (4.160)$$

in which  $\beta_r = \left(1 - \frac{1}{T_b}\right) \eta_r$ . Let  $\Lambda_{rp_1}$  and  $\Lambda_{re}$  represent the stability regions achieved under systems 'rp<sub>1</sub>' and 're', respectively. Suppose that the mean arrival rate vector  $\mathbf{a}$  is interior to fraction  $\beta_r$  of region  $\Lambda_{rp_1}$ , i.e. there exists an  $\epsilon_{\max}$ , which depends on  $\mathbf{a}$ , such that

$$(a_1 + \epsilon_{\max}, \dots, a_N + \epsilon_{\max}) \in \beta_r \Lambda_{rp_1}. \quad (4.161)$$

The above can be re-expressed as

$$(\beta_r^{-1} a_1 + \beta_r^{-1} \epsilon_{\max}, \dots, \beta_r^{-1} a_N + \beta_r^{-1} \epsilon_{\max}) \in \Lambda_{rp_1}. \quad (4.162)$$

By considering a particular policy  $\Delta$  that depends only on the channels states, it results that

$$r_i^{(\Delta)}(nT_b) = \mathbb{E} \{ D_i^{(\Delta)}(nT_b + t_1) \} \geq \beta_r^{-1} a_i + \beta_r^{-1} \epsilon_{\max}, \quad \forall i \in \{1, \dots, N\}. \quad (4.163)$$

Plugging the above inequality into (4.160) yields

$$Dr^{(\Delta_{re})}(\mathbf{q}(nT_b)) \leq E - \epsilon_{\max} T_b \sum_{i=1}^N q_i(nT_b). \quad (4.164)$$

## 4.7. Appendix

Let  $\check{\epsilon}_{\max} = \epsilon_{\max} T_b$ . Taking an expectation of  $Dr^{(\Delta_{\text{re}})}$  over the randomness of the queue lengths and summing over  $n \in \{0, 1, \dots, T-1\}$  for some integer  $T > 0$ , we get

$$\mathbb{E}\{Ly(\mathbf{q}(TT_b))\} - \mathbb{E}\{Ly(\mathbf{q}(0))\} \leq ET - \check{\epsilon}_{\max} \sum_{n=0}^{T-1} \sum_{i=1}^N \mathbb{E}\{q_i(nT_b)\}. \quad (4.165)$$

Rearranging terms, dividing by  $\check{\epsilon}_{\max} T$ , using the facts that  $Ly(\mathbf{q}(TT_b)) \geq 0$  and  $\mathbb{E}\{Ly(\mathbf{q}(0))\} < \infty$ , and taking a limsup, we eventually obtain

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{i=1}^N \mathbb{E}\{q_i(nT_b)\} \leq \frac{E}{\check{\epsilon}_{\max}}. \quad (4.166)$$

Using the definition of strong stability and based on the above result, it can be stated that  $\Delta_{\text{re}}$  stabilizes any mean arrival rate vector interior to fraction  $\beta_r$  of the stability region achieved under policy  $\Delta_{\text{rp}_1}$ , i.e.  $\Delta_{\text{re}}$  stabilizes the system for any arrivals such that  $\mathbf{a} \in \beta_r \Lambda_{\text{rp}_1}$ .

In Lemma 8, it was shown that systems 'rp<sub>1</sub>' and 'rp' have the same stability region. Therefore, we can deduce that  $\Delta_{\text{re}}$  stabilizes any arrival rate vector interior to fraction  $\beta_r$  of the stability region achieved under  $\Delta_{\text{rp}}$ . This completes the proof.

### 4.7.3 Proof of Corollary 1

Regarding the derivation of  $\mathbb{P}\{\check{C}_{ij}(t) > 0\}$ , the following observation can be made. Rate  $\check{C}_{ij}(t)$  is different than zero, i.e.  $\check{C}_{ij}(t) > 0$ , if (a)  $\gamma_{ij}^{(d)}(t) \geq \tau$  or (b)  $\gamma_{ij}^{(d)}(t) < \tau$ ,  $\check{\gamma}_{ij}^{(r)}(t) \geq \tau$  and  $\gamma_j^{(sr)}(t) \geq \delta$ . Define the event  $\mathcal{G}_1$  such that this event occurs if  $\gamma_{ij}^{(d)}(t) \geq \tau$ . Also, define  $\mathcal{G}_2$  to be the event that occurs if  $\gamma_{ij}^{(d)}(t) < \tau$ ,  $\check{\gamma}_{ij}^{(r)}(t) \geq \tau$  and  $\gamma_j^{(sr)}(t) \geq \delta$ . Events  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are disjoint (i.e. mutually exclusive) since they cannot occur at the same time, i.e. in the first one we have  $\gamma_{ij}^{(d)}(t) \geq \tau$  whereas in the second one  $\gamma_{ij}^{(d)}(t) < \tau$ , which implies that  $\mathbb{P}\{\mathcal{G}_1 \cap \mathcal{G}_2\} = 0$ . Thus, we can write

$$\begin{aligned} \mathbb{P}\{\check{C}_{ij}(t) > 0\} &= \mathbb{P}\{\mathcal{G}_1 \cup \mathcal{G}_2\} \\ &= \mathbb{P}\{\mathcal{G}_1\} + \mathbb{P}\{\mathcal{G}_2\}. \end{aligned} \quad (4.167)$$

Recalling that  $\gamma_{ij}^{(d)}(t)$  is defined as

$$\gamma_{ij}^{(d)}(t) = \rho_{ij}^{(\text{su})} |h_{ij}^{(\text{su})}(t)|^2, \text{ where } \rho_{ij}^{(\text{su})} = \frac{P^{(\text{s})}}{(\sigma_{z,ij}^{(\text{u})})^2} \quad (4.168)$$

and that  $\sigma_{z,i}^{(\text{u})} = \sigma_{z,ij}^{(\text{u})}$ , we can re-express  $\gamma_{ij}^{(d)}(t)$  as

$$\gamma_{ij}^{(d)}(t) = \rho_i^{(\text{su})} |h_{ij}^{(\text{su})}(t)|^2, \text{ where } \rho_i^{(\text{su})} = \frac{P^{(\text{s})}}{(\sigma_{z,i}^{(\text{u})})^2}. \quad (4.169)$$

## 4.7. Appendix

---

Based on the assumption that  $h_{ij}^{(\text{su})}(t)$  follows a (complex) Gaussian distribution with zero mean and variance  $(\sigma_i^{(\text{su})})^2$ , we can derive  $\mathbb{P}\{\mathcal{G}_1\}$  as follows

$$\begin{aligned}\mathbb{P}\{\mathcal{G}_1\} &= \mathbb{P}\left\{\gamma_{ij}^{(\text{d})}(t) \geq \tau\right\} \\ &= \mathbb{P}\left\{\rho_j^{(\text{su})}|h_{ij}^{(\text{su})}(t)|^2 \geq \tau\right\} \\ &= \exp\left(-\frac{\tau}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2}\right).\end{aligned}\tag{4.170}$$

On the other side, for the derivation of  $\mathbb{P}\{\mathcal{G}_2\}$ , recall that

$$\check{\gamma}_{ij}^{(\text{r})}(t) = \gamma_{ij}^{(\text{su})}(t) + \check{\gamma}_{ij}^{(\text{sru})}(t),\tag{4.171}$$

where  $\gamma_{ij}^{(\text{su})}(t) = \gamma_{ij}^{(\text{d})}(t)$  and where  $\check{\gamma}_{ij}^{(\text{sru})}(t)$  is given as the following

$$\check{\gamma}_{ij}^{(\text{sru})}(t) = \frac{\gamma_{ij}^{(\text{ru})}(t) \delta}{\gamma_{ij}^{(\text{ru})}(t) + \delta + 1}.\tag{4.172}$$

Recalling that  $\gamma_{ij}^{(\text{ru})}(t)$  and  $\gamma_j^{(\text{sr})}(t)$  are defined as follows

$$\gamma_{ij}^{(\text{ru})}(t) = \rho_{ij}^{(\text{ru})}|h_{ij}^{(\text{ru})}(t)|^2, \text{ with } \rho_{ij}^{(\text{ru})} = \frac{P^{(\text{r})}}{(\sigma_{z,ij}^{(\text{u})})^2},\tag{4.173}$$

$$\gamma_j^{(\text{sr})}(t) = \rho_j^{(\text{sr})}|h_j^{(\text{sr})}(t)|^2, \text{ with } \rho_j^{(\text{sr})} = \frac{P^{(\text{s})}}{(\sigma_{z,j}^{(\text{r})})^2},\tag{4.174}$$

and that  $\sigma_z^{(\text{r})} = \sigma_{z,j}^{(\text{r})}$  and  $\sigma_{z,i}^{(\text{u})} = \sigma_{z,ij}^{(\text{u})}$ , we can rewrite  $\gamma_{ij}^{(\text{ru})}(t)$  and  $\gamma_j^{(\text{sr})}(t)$  as

$$\gamma_{ij}^{(\text{ru})}(t) = \rho_i^{(\text{ru})}|h_{ij}^{(\text{ru})}(t)|^2, \text{ with } \rho_i^{(\text{ru})} = \frac{P^{(\text{r})}}{(\sigma_{z,i}^{(\text{u})})^2},\tag{4.175}$$

$$\gamma_j^{(\text{sr})}(t) = \rho^{(\text{sr})}|h_j^{(\text{sr})}(t)|^2, \text{ with } \rho^{(\text{sr})} = \frac{P^{(\text{s})}}{(\sigma_z^{(\text{r})})^2}.\tag{4.176}$$

Using the fact that  $\gamma_j^{(\text{sr})}(t)$  is independent of both  $\gamma_{ij}^{(\text{d})}(t)$  and  $\check{\gamma}_{ij}^{(\text{r})}(t)$ , we have

$$\begin{aligned}\mathbb{P}\{\mathcal{G}_2\} &= \mathbb{P}\left\{\gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(\text{r})}(t) \geq \tau, \gamma_j^{(\text{sr})}(t) \geq \delta\right\} \\ &= \mathbb{P}\left\{\gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(\text{r})}(t) \geq \tau\right\} \mathbb{P}\left\{\gamma_j^{(\text{sr})}(t) \geq \delta\right\}.\end{aligned}\tag{4.177}$$



## 4.7. Appendix

---

By noticing that  $\check{\gamma}_{ij}^{(r)}(t) \geq \tau$  can be rewritten as  $\check{\gamma}_{ij}^{(\text{sru})}(t) \geq \tau - \gamma_{ij}^{(\text{su})}(t)$  and since  $\gamma_{ij}^{(\text{d})}(t) = \gamma_{ij}^{(\text{su})}(t)$ , the following holds

$$\begin{aligned} \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(r)}(t) \geq \tau \right\} &= \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(\text{sru})}(t) \geq \tau - \gamma_{ij}^{(\text{d})}(t) \right\} \\ &= \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \frac{\gamma_{ij}^{(\text{ru})}(t) \delta}{\gamma_{ij}^{(\text{ru})}(t) + \delta + 1} \geq \tau - \gamma_{ij}^{(\text{d})}(t) \right\}. \end{aligned} \quad (4.178)$$

Depending on the value of  $\delta$ , two cases are to consider here: the first case is when  $\delta < \tau$  and in the second case we have  $\delta \geq \tau$ .

- For  $\delta < \tau$ : In this case, if  $\gamma_{ij}^{(\text{d})}(t) < \tau - \delta$ , or equivalently  $\delta - (\tau - \gamma_{ij}^{(\text{d})}(t)) < 0$ , we get

$$\begin{aligned} &\mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \frac{\gamma_{ij}^{(\text{ru})}(t) \delta}{\gamma_{ij}^{(\text{ru})}(t) + \delta + 1} \geq \tau - \gamma_{ij}^{(\text{d})}(t) \right\} = \\ &\mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \gamma_{ij}^{(\text{ru})}(t)(\delta - \tau + \gamma_{ij}^{(\text{d})}(t)) \geq (\delta + 1)(\tau - \gamma_{ij}^{(\text{d})}(t)) \right\} = \\ &\mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \gamma_{ij}^{(\text{ru})}(t) \leq \frac{(\delta + 1)(\tau - \gamma_{ij}^{(\text{d})}(t))}{\delta - (\tau - \gamma_{ij}^{(\text{d})}(t))} \right\} = 0, \end{aligned} \quad (4.179)$$

where the last equality holds since we know that  $\gamma_{ij}^{(\text{ru})}(t) \geq 0$ , while here we have the probability of getting  $\gamma_{ij}^{(\text{ru})}(t) < 0$  because

$$\frac{(\delta + 1)(\tau - \gamma_{ij}^{(\text{d})}(t))}{\delta - (\tau - \gamma_{ij}^{(\text{d})}(t))} < 0. \quad (4.180)$$

On the other hand, if  $\gamma_{ij}^{(\text{d})}(t) \geq \tau - \delta$ , or equivalently  $\delta - (\tau - \gamma_{ij}^{(\text{d})}(t)) \geq 0$ , it yields

$$\mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \frac{\gamma_{ij}^{(\text{ru})}(t) \delta}{\gamma_{ij}^{(\text{ru})}(t) + \delta + 1} \geq \tau - \gamma_{ij}^{(\text{d})}(t) \right\} = \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \gamma_{ij}^{(\text{ru})}(t) \geq \frac{(\tau - \gamma_{ij}^{(\text{d})}(t))(\delta + 1)}{\delta - (\tau - \gamma_{ij}^{(\text{d})}(t))} \right\}.$$

Based on the above and on equation (4.178), we can write

$$\mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(r)}(t) \geq \tau \right\} = \int_{\tau - \delta}^{\tau} \mathbb{P} \left\{ \gamma_{ij}^{(\text{ru})}(t) \geq \frac{(\tau - \gamma)(\delta + 1)}{\delta - (\tau - \gamma)} \right\} f_{\gamma_{ij}^{(\text{d})}(t)}(\gamma) d\gamma, \quad (4.181)$$

where  $f_{\gamma_{ij}^{(\text{d})}(t)}(\gamma)$  represents the PDF of variable  $\gamma_{ij}^{(\text{d})}(t)$ . Since  $h_{ij}^{(\text{ru})}(t)$  is assumed to follow a (complex) Gaussian distribution with zero mean and variance  $(\sigma_i^{(\text{ru})})^2$ , it follows that

$$\mathbb{P} \left\{ \gamma_{ij}^{(\text{ru})}(t) \geq \frac{(\tau - \gamma)(\delta + 1)}{\delta - (\tau - \gamma)} \right\} = \exp \left( -\frac{(\tau - \gamma)(\delta + 1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta - (\tau - \gamma))} \right). \quad (4.182)$$

In addition, we have

$$f_{\gamma_{ij}^{(\text{d})}(t)}(\gamma) = \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp \left( -\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) d\gamma. \quad (4.183)$$

## 4.7. Appendix

---

We thus obtain

$$\begin{aligned} & \int_{\tau-\delta}^{\tau} \mathbb{P} \left\{ \gamma_{ij}^{(\text{ru})}(t) \geq \frac{(\tau-\gamma)(\delta+1)}{\delta-(\tau-\gamma)} \right\} f_{\gamma_{ij}^{(\text{d})}(t)}(\gamma) d\gamma = \\ & \int_{\tau-\delta}^{\tau} \exp \left( -\frac{(\tau-\gamma)(\delta+1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta-(\tau-\gamma))} \right) \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp \left( -\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) d\gamma. \end{aligned} \quad (4.184)$$

It is worth noting that the above integral does not have a closed-form solution. Using the assumption that  $h_j^{(\text{sr})}(t)$  follows a complex Gaussian distribution with zero mean and variance  $(\sigma^{(\text{sr})})^2$  yields

$$\mathbb{P} \left\{ \gamma_j^{(\text{sr})}(t) \geq \delta \right\} = \exp \left( -\frac{\delta}{\rho^{(\text{sr})}(\sigma^{(\text{sr})})^2} \right). \quad (4.185)$$

Combining equations (4.177), (4.181), (4.184), and (4.185) yields

$$\begin{aligned} \mathbb{P} \{ \mathcal{G}_2 \} = & \left[ \int_{\tau-\delta}^{\tau} \exp \left( -\frac{(\tau-\gamma)(\delta+1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta-(\tau-\gamma))} \right) \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp \left( -\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) d\gamma \right] \exp \left( -\frac{\delta}{\rho^{(\text{sr})}(\sigma^{(\text{sr})})^2} \right). \end{aligned} \quad (4.186)$$

Recalling that  $\mathbb{P} \{ \mathcal{G} \} = \mathbb{P} \{ \mathcal{G}_1 \} + \mathbb{P} \{ \mathcal{G}_2 \}$ , we eventually get

$$\begin{aligned} \mathbb{P} \{ \mathcal{G} \} = & \exp \left( -\frac{\tau}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) + \\ & \left[ \int_{\tau-\delta}^{\tau} \exp \left( -\frac{(\tau-\gamma)(\delta+1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta-(\tau-\gamma))} \right) \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp \left( -\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2} \right) d\gamma \right] \exp \left( -\frac{\delta}{\rho^{(\text{sr})}(\sigma^{(\text{sr})})^2} \right). \end{aligned} \quad (4.187)$$

Since  $\mathbb{P} \{ \check{C}_{ij}(t) > 0 \} = \mathbb{P} \{ \mathcal{G} \}$ , the first desired result regarding  $\mathbb{P} \{ \check{C}_{ij}(t) > 0 \}$  follows.

- For  $\delta \geq \tau$ : The derivations in this case are similar to those given before, thus only the outline will be provided here. Using equations (4.178) and (4.181), we can write

$$\begin{aligned} \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(\text{r})}(t) \geq \tau \right\} &= \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \check{\gamma}_{ij}^{(\text{ru})}(t) \geq \tau - \gamma_{ij}^{(\text{d})}(t) \right\} \\ &= \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \frac{\gamma_{ij}^{(\text{ru})}(t) \delta}{\gamma_{ij}^{(\text{ru})}(t) + \delta + 1} \geq \tau - \gamma_{ij}^{(\text{d})}(t) \right\} \\ &= \mathbb{P} \left\{ \gamma_{ij}^{(\text{d})}(t) < \tau, \gamma_{ij}^{(\text{ru})}(t) \geq \frac{(\tau - \gamma_{ij}^{(\text{d})}(t))(\delta + 1)}{\delta - (\tau - \gamma_{ij}^{(\text{d})}(t))} \right\} \\ &= \int_0^{\tau} \mathbb{P} \left\{ \gamma_{ij}^{(\text{ru})}(t) \geq \frac{(\tau - \gamma)(\delta + 1)}{\delta - (\tau - \gamma)} \right\} f_{\gamma_{ij}^{(\text{d})}(t)}(\gamma) d\gamma, \end{aligned} \quad (4.188)$$

where in the above equalities we use the fact that here  $\delta - (\tau - \gamma_{ij}^{(\text{d})}(t)) \geq 0$  for  $0 \leq \gamma_{ij}^{(\text{d})}(t) \leq \tau$ .

## 4.7. Appendix

---

Hence,  $\mathbb{P}\{\mathcal{G}\}$  can be expressed as

$$\begin{aligned} \mathbb{P}\{\mathcal{G}\} = & \exp\left(-\frac{\tau}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2}\right) + \\ & \left[ \int_0^\tau \exp\left(-\frac{(\tau-\gamma)(\delta+1)}{\rho_i^{(\text{ru})}(\sigma_i^{(\text{ru})})^2(\delta-(\tau-\gamma))}\right) \frac{1}{\rho_i^{(\text{su})}(\sigma_i^{(\text{su})})^2} \exp\left(-\frac{\gamma}{\rho^{(\text{su})}(\sigma^{(\text{su})})^2}\right) d\gamma \right] \exp\left(-\frac{\delta}{\rho^{(\text{sr})}(\sigma^{(\text{sr})})^2}\right). \end{aligned} \quad (4.189)$$

Since  $\mathbb{P}\{\check{C}_{ij}(t) > 0\} = \mathbb{P}\{\mathcal{G}\}$ , the second desired result concerning  $\mathbb{P}\{\check{C}_{ij}(t) > 0\}$  follows.

From the above expressions of  $\mathbb{P}\{\check{C}_{ij}(t) > 0\}$ , it is plain to see that this probability is the same independently of channel index  $j$ , meaning that for any two channels  $j_1$  and  $j_2$  the following holds

$$\mathbb{P}\{\check{C}_{ij_1}(t) > 0\} = \mathbb{P}\{\check{C}_{ij_2}(t) > 0\}. \quad (4.190)$$

Therefore, the expression of  $\beta_r$  given in Theorem 15 reduces to

$$\beta_r = \left(1 - \frac{1}{T_b}\right) \min_{i \in \{1, \dots, N\}} \left\{ \mathbb{P}\{\check{C}_{ij}(t) > 0\} \right\}. \quad (4.191)$$

This completes the proof.



# Chapter 5

## Conclusions and Perspectives

### 5.1 Conclusions

In this thesis, we have addressed problems of feedback allocation and user scheduling in wireless systems, taking into consideration the dynamic traffic arrivals for the users and accounting for realistic limitations regarding the feedback process. The main performance measure we adopt is the queueing stability, which is mainly captured by the stability region the proposed algorithm is capable to achieve.

In Chapter 3, we have considered a multi-point to multi-point MIMO system under limited backhaul capacity and where the probing cost is taken into account. For this system, SVD technique is used if only one pair is active, whereas IA technique is applied if more than one pair are active. Two systems are considered, namely the symmetric system, where all the path loss coefficients are equal to each other, and the general system. One of the important questions that this chapter answers is: Given the limited backhaul conditions and the probing cost, does scheduling more than one pair (and thus using IA) is better than scheduling only one pair (and thus using SVD)? We have answered this question by (i) proposing throughput-optimal centralized scheduling policies that each time-slot select the subset of pairs that should be active, and (ii) characterizing the conditions under which the IA technique can deliver queueing stability gains w.r.t. the SVD technique. For the general system, we have also provided a low-complexity scheduling policy since the throughput-optimal one is of a high computational complexity under this system, where clearly the stability region of the former policy is a subset of that of the latter one.

The main conclusion that can be drawn from Chapter 4 is that the users must be more involved in the feedback and scheduling decisions in future wireless networks, particularly in cooperative relaying systems. Specifically, in this chapter two multichannel multiuser systems are considered, namely one without relaying and the other with relaying, where both limited and delayed feedback is accounted for. Note that for the system with relaying the users have imperfect knowledge of the states of the links between the relay and the BS. By exploiting the fact that the users can know their instantaneous CSIs (even if this knowledge is sometimes imperfect), and assuming idealized contention procedure, for each system we have developed a decentralized joint feedback and scheduling algorithm where the

users contend among each other to make the decisions. In addition, by leveraging the same fact, we have proposed a second algorithm that greatly imitates the former one and that does not use the idealized contention assumption. In this algorithm, the users make the feedback decisions using a threshold-based scheme and then the BS performs user scheduling based on the subset of reported feedback. The trade-off between the amount of feedback resources and the delay in the feedback process was also exploited via this algorithm. Furthermore, for the system with relaying, in addition to the two algorithms mentioned before, a third joint feedback and scheduling algorithm is developed in the special case where the delay is not accounted for. All these algorithms and their good stability performances have led us to the statement given at the beginning of this paragraph.

In this thesis, we highlight the value of network state information, particularly CSI, for scheduling in wireless systems. A central idea punctuating the entire work is that the information structure available for the scheduling mechanism influences the system performance. With its investigations and results, this thesis serves as a step forward towards a more comprehensive theory that tackles scheduling problems under various kinds of incomplete information. As wireless systems experience steep growth in performance demands, with no signs of slowing down, we believe that such a general framework to investigate wireless system scheduling with incomplete knowledge of system state will prove to be an invaluable asset in designing high-performance systems.

## 5.2 Perspectives

In Chapter 3, a centralized policy was adopted to select the subset of active pairs at each time-slot, based on the queue lengths and the statistics of the channels. Important extensions of this work can be addressing the stability analysis when we adopt decentralized or even mixed (i.e. combining centralized and decentralized schemes) policies for feedback and scheduling. These schemes can overcome the shortcomings of the centralized one by exploiting the fact that each receiver can know its actual channel realization, however they may require very cumbersome calculations. Another interesting but challenging extension would be to propose and analyze a joint scheduling and power control scheme, which may help enlarging the achievable stability region. One approach to tackle this problem would be to consider time average power expenditure constraints, in which case the concept of *virtual queues* can be used to transform these constraints into queue stability problems [16]. The idea here is that scheduling and power control should be performed in such a way as to ensure the stability of both data and virtual queues.

In Chapter 4, one of the two systems we studied is the multiuser multichannel system with cooperative relaying, where we recall that delayed and limited feedback is accounted for, and where we assume that each user has a respective queue at the BS. This study can be extended to consider that each user has a queue also at the relay. In such a system, it can be noticed that the departure process of the BS queues is the arrival process of the relay queues. From a stability performance point of view, the obvious candidate for scheduling in this scenario is the *Back-Pressure* algorithm [19], which routes and schedules packets based on differential backlogs (i.e. queue length differences between the BS and the relay). In this regard, it would be interesting to investigate the impact of the different imperfections and limitations in the system on the performance of such an algorithm.

## 5.2. Perspectives

---

The systems we have considered in this work adopt advanced and complicated transmission techniques, such as IA and cooperative relaying, which have made the stability analysis particularly challenging. Other advanced transmission techniques exist, such as coordinated multi-point (CoMP) and network coding. CoMP is essentially a range of different techniques that enable the dynamic coordination of transmission and reception over a variety of different BSs. Network coding allows mixing of messages from different nodes before sending this mixture on shared links, instead of separate links for every messages. It would be intriguing to address scheduling problems for systems applying such techniques and where dynamic traffic arrivals are taken into consideration. Clearly, the more realistic limitations are considered for these systems, the more challenging the stability analysis will be.

In this thesis, stability of the queues is the main performance metric we adopted to analyze and evaluate the scheduling policies. While stability is an important metric, in most practical systems quality-of-service (QoS) is more important. A key parameter in the QoS requirements is the queueing delay a packet experiences as it waits to be transmitted. Proper scheduling mechanism is vital for guaranteeing delay-sensitive applications to run smoothly. Note that the scheduling policies proposed in this thesis ensure finite delays in the system since, using Little's Theorem, we know that the average delay is proportional to the average queue length. However, this is not the same as providing an exact analysis where hard packet-delay constraints need to be respected. Therefore, under such constraints, it would be interesting to design delay-based scheduling policies and to analyze their performances; such an analysis can be done using, for example, large deviations techniques [129].

Finally, another interesting extension of the work in this thesis would be to consider a scenario with *flow-level dynamics*, meaning that the collection of active queues (or equivalently, the set of users) in the system is not fixed. Indeed, in most practical systems the collection of active queues dynamically varies, as sessions eventually end while new sessions occasionally start. Under such a scenario, developing efficient scheduling algorithms is of particular interest, especially since Max-Weight policies were shown to no longer be optimal [130].





# Bibliography

- [1] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update 2016–2021,” *White paper*, 2017.
- [2] G. Foschini and M. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998. [Online]. Available: <http://dx.doi.org/10.1023/A:1008889222784>
- [3] D. Gesbert, M. Kountouris, R. W. H. Jr., C. b. Chae, and T. Salzer, “Shifting the MIMO paradigm,” *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, Sept 2007.
- [4] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.
- [5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufveson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [6] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of MIMO channels,” *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.
- [7] V. R. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the  $K$ -user interference channel,” *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008.
- [8] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication-part i: channel inversion and regularization,” *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan 2005.
- [9] M. B. Stefania Sesia, Issam Toufik, “LTE, the UMTS long term evolution: From theory to practice.” Wiley, 2009.
- [10] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband radio,” *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, Sept 2004.
- [11] W. Zhuang and M. Ismail, “Cooperation in wireless communication networks,” *IEEE Wireless Communications*, vol. 19, no. 2, pp. 10–20, April 2012.

## Bibliography

---

- [12] 3GPP, “LTE-Advanced,” 2013. [Online]. Available: <http://www.3gpp.org/lte-advanced>
- [13] O. S. Shin, A. M. Chan, H. T. Kung, and V. Tarokh, “Design of an OFDM cooperative space-time diversity system,” *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 2203–2215, July 2007.
- [14] L. Dai, B. Gui, and L. J. C. Jr., “Selective relaying in OFDM multihop cooperative networks,” in *2007 IEEE Wireless Communications and Networking Conference*, March 2007, pp. 963–968.
- [15] L. Georgiadis, M. J. Neely, and L. Tassiulas, “Resource allocation and cross-layer control in wireless networks,” *Found. Trends Netw.*, vol. 1, no. 1, pp. 1–144, Apr. 2006.
- [16] M. J. Neely, “Stochastic network optimization with application to communication and queueing systems,” *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.
- [17] A. Ephremides and B. Hajek, “Information theory and communication networks: an unconsumed union,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2416–2434, Oct 1998.
- [18] X. Liu, E. K. P. Chong, and N. B. Shroff, “A framework for opportunistic scheduling in wireless networks,” *Comput. Netw.*, vol. 41, no. 4, pp. 451–474, Mar. 2003. [Online]. Available: [http://dx.doi.org/10.1016/S1389-1286\(02\)00401-2](http://dx.doi.org/10.1016/S1389-1286(02)00401-2)
- [19] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks,” *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.
- [20] —, “Dynamic server allocation to parallel queues with randomly varying connectivity,” *IEEE Trans. Inform. Theory*, vol. 39, no. 2, pp. 466–478, Mar 1993.
- [21] L. Tassiulas, “Scheduling and performance limits of networks with constantly changing topology,” *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1067–1073, May 1997.
- [22] M. J. Neely, “Dynamic power allocation and routing for time-varying wireless networks,” *IEEE J. Sel. Areas Commun*, pp. 89–103, 2005.
- [23] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, “Achieving 100 percent throughput in an input-queued switch,” *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1260–1267, Aug 1999.
- [24] P. Kumar and S. Meyn, “Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies,” *IEEE Transactions on Automatic Control*, vol. 41, no. 1, pp. 4–17, Jan 1996.
- [25] E. Leonardi, M. Mellia, F. Neri, and M. Ajmone Marsan, “Bounds on average delays and queue size averages and variances in input-queued cell-based switches,” in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2001, pp. 1095–1103 vol.2.
- [26] W. Szpankowski, “Stability conditions for some distributed systems: Buffered random access systems,” *Buffered Random Access Systems, Adv. Appl. Probab*, vol. 26, pp. 498–515, 1993.

## Bibliography

---

- [27] M. Neely, E. Modiano, and C. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 138–152, Feb 2003.
- [28] M. J. Neely, "Dynamic power allocation and routing for satellite and wireless networks with time varying channels," Ph.D. dissertation, Massachusetts Institute of Technology, 2003.
- [29] V. Tsibonis, L. Georgiadis, and L. Tassiulas, "Exploiting wireless channel state information for throughput maximization," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, vol. 1, March 2003, pp. 301–310 vol.1.
- [30] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2053–2064, Oct 2001.
- [31] A. Eryilmaz, R. Srikant, and J. R. Perkins, "Throughput-optimal scheduling for broadcast channels," *Proc. SPIE 4531*, Jul 2001.
- [32] S. Shakkottai, R. Srikant, and A. L. Stolyar, "Pathwise optimality of the exponential scheduling rule for wireless channels," *Adv. in Appl. Probab.*, vol. 36, no. 4, pp. 1021–1045, 12 2004. [Online]. Available: <http://dx.doi.org/10.1239/aap/1103662957>
- [33] A. L. Stolyar, "Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 1–53, 02 2004. [Online]. Available: <http://dx.doi.org/10.1214/aoap/1075828046>
- [34] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 104–116, Jan 2005.
- [35] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1452–1463, Aug 2006.
- [36] K. Kar, X. Luo, and S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," in *Proceedings of the IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 1640–1648. [Online]. Available: <http://dx.doi.org/10.1109/INFCOM.2007.192>
- [37] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queuing system with asynchronously varying service rates," *Probab. Eng. Inf. Sci.*, vol. 18, no. 2, pp. 191–217, Apr. 2004. [Online]. Available: <http://dx.doi.org/10.1017/S0269964804182041>
- [38] Z. Ji, Y. Yang, J. Zhou, M. Takai, and R. Bagrodia, "Exploiting medium access diversity in rate adaptive wireless LANs," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '04. New York, NY, USA: ACM, 2004, pp. 345–359. [Online]. Available: <http://doi.acm.org/10.1145/1023720.1023754>

## Bibliography

---

- [39] N. B. Chang and M. Liu, “Optimal channel probing and transmission scheduling for opportunistic spectrum access,” in *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, ser. MobiCom ’07. New York, NY, USA: ACM, 2007, pp. 27–38. [Online]. Available: <http://doi.acm.org/10.1145/1287853.1287858>
- [40] P. Chaporkar, A. Proutiere, H. Asnani, and A. Karandikar, “Scheduling with limited information in wireless systems,” in *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, ser. MobiHoc ’09. New York, NY, USA: ACM, 2009, pp. 75–84.
- [41] A. Gopalan, C. Caramanis, and S. Shakkottai, “On wireless scheduling with partial channel-state information,” *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 403–420, Jan 2012.
- [42] M. Karaca, Y. Sarikaya, O. Ercetin, T. Alpcan, and H. Boche, “Joint opportunistic scheduling and selective channel feedback,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 3024–3034, June 2013.
- [43] C. Swannack, E. Uysal-biyikoglu, and G. Wornell, “Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel,” in *Proc. Allerton Conf. Commun., Contr., and Computing*, 2004.
- [44] M. Ouyang and L. Ying, “Approaching throughput optimality with limited feedback in multi-channel wireless downlink networks,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1827–1838, Dec 2013.
- [45] A. Destounis, M. Assaad, M. Debbah, and B. Sayadi, “A traffic aware joint CQI feedback and scheduling scheme for multichannel downlink systems in TDD feedback mode,” in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 1523–1527.
- [46] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, “Low-complexity scheduling algorithms for multi-channel downlink wireless networks,” in *2010 Proceedings IEEE INFOCOM*, March 2010, pp. 1–9.
- [47] —, “Scheduling for small delay in multi-rate multi-channel wireless networks,” in *2011 Proceedings IEEE INFOCOM*, April 2011, pp. 1251–1259.
- [48] B. Ji, G. R. Gupta, X. Lin, and N. B. Shroff, “Performance of low-complexity greedy scheduling policies in multi-channel wireless networks: Optimal throughput and near-optimal delay,” in *2013 Proceedings IEEE INFOCOM*, April 2013, pp. 2589–2597.
- [49] —, “Low-complexity scheduling policies for achieving throughput and asymptotic delay optimality in multichannel wireless networks,” *IEEE/ACM Transactions on Networking*, vol. 22, no. 6, pp. 1911–1924, Dec 2014.
- [50] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, “Scheduling in multi-channel wireless networks: Rate function optimality in the small-buffer regime,” *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1101–1125, Feb 2014.
- [51] M. Assaad, “Frequency-time scheduling for streaming services in OFDMA systems,” in *2008 1st IFIP Wireless Days*, Nov 2008, pp. 1–5.

## Bibliography

---

- [52] N. U. Hassan and M. Assaad, "Low complexity margin adaptive resource allocation in downlink MIMO-OFDMA system," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3365–3371, July 2009.
- [53] A. Ahmad and M. Assaad, "Margin adaptive resource allocation in downlink OFDMA system with outdated channel state information," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2009, pp. 1868–1872.
- [54] N. U. Hassan and M. Assaad, "Resource allocation in multiuser OFDMA system: Feasibility and optimization study," in *2009 IEEE Wireless Communications and Networking Conference*, April 2009, pp. 1–6.
- [55] A. Ahmad and M. Assaad, "Optimal resource allocation framework for downlink OFDMA system with channel estimation error," in *2010 IEEE Wireless Communication and Networking Conference*, April 2010, pp. 1–5.
- [56] L. Ying and S. Shakkottai, "On throughput optimality with delayed network-state information," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5116–5132, Aug 2011.
- [57] —, "Scheduling in mobile ad hoc networks with topology and channel-state uncertainty," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2504–2517, Oct 2012.
- [58] M. Johnston and E. Modiano, "A new look at wireless scheduling with delayed information," in *Information Theory (ISIT), 2015 IEEE International Symposium on*, June 2015, pp. 1407–1411.
- [59] C. Wang and R. D. Murch, "Optimal downlink multi-user MIMO cross-layer scheduling using hol packet waiting time," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2856–2862, Oct 2006.
- [60] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1468–1477, September 2007.
- [61] M. Kobayashi, G. Caire, and D. Gesbert, "Transmit diversity versus opportunistic beamforming in data packet mobile downlink transmission," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 151–157, Jan 2007.
- [62] S. Lakshminarayana, M. Assaad, and M. Debbah, "H-infinity control based scheduler for the deployment of small cell networks," in *2011 International Symposium of Modeling and Optimization of Mobile, Ad Hoc, and Wireless Networks*, May 2011, pp. 9–16.
- [63] K. Huang and V. Lau, "Stability and delay of zero-forcing SDMA with limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6499–6514, Oct 2012.
- [64] A. Destounis, M. Assaad, M. Debbah, and B. Sayadi, "Traffic-aware training and scheduling for the 2-user MISO broadcast channel," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 1376–1380.
- [65] —, "A threshold-based approach for joint active user selection and feedback in MISO downlink systems," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 2869–2875.

## Bibliography

---

- [66] A. Destounis, M. Assaad, M. Debbah, B. Sayadi, and A. Feki, "On queue-aware power control in interfering wireless links: Heavy traffic asymptotic modelling and application in QoS provisioning," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2345–2356, Oct 2014.
- [67] S. Lakshminarayana, M. Assaad, and M. Debbah, "Transmit power minimization in small cell networks under time average QoS constraints," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2087–2103, Oct 2015.
- [68] A. Destounis, M. Assaad, M. Debbah, and B. Sayadi, "Traffic-aware training and scheduling for MISO wireless downlink systems," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2574–2599, May 2015.
- [69] E. M. Yeh and R. A. Berry, "Throughput optimal control of cooperative relay networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3827–3833, Oct 2007.
- [70] V. J. Venkataramanan, X. Lin, L. Ying, and S. Shakkottai, "On scheduling for minimizing end-to-end buffer usage over multihop wireless networks," in *2010 Proceedings IEEE INFOCOM*, March 2010, pp. 1–9.
- [71] B. Ji, C. Joo, and N. Shroff, "Throughput-optimal scheduling in multihop wireless networks without per-flow information," *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 634–647, April 2013.
- [72] —, "Throughput-optimal scheduling in multihop wireless networks without per-flow information," *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 634–647, April 2013.
- [73] S. Liu, E. Ekici, and L. Ying, "Scheduling in multihop wireless networks without back-pressure," *IEEE/ACM Transactions on Networking*, vol. 22, no. 5, pp. 1477–1488, Oct 2014.
- [74] S. Moharir and S. Shakkottai, "Maxweight versus backpressure: Routing and scheduling in multichannel relay networks," *IEEE/ACM Transactions on Networking*, vol. 23, no. 5, pp. 1584–1598, Oct 2015.
- [75] C. Joo, X. Lin, and N. B. Shroff, "Understanding the capacity region of the greedy maximal scheduling algorithm in multihop wireless networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 4, pp. 1132–1145, Aug 2009.
- [76] E. Modiano, D. Shah, and G. Zussman, "Maximizing throughput in wireless networks via gossiping," *SIGMETRICS Perform. Eval. Rev.*, vol. 34, no. 1, pp. 27–38, Jun. 2006. [Online]. Available: <http://doi.acm.org/10.1145/1140103.1140283>
- [77] S. Sanghavi, L. Bui, and R. Srikant, "Distributed link scheduling with constant overhead," *SIGMETRICS Perform. Eval. Rev.*, vol. 35, no. 1, pp. 313–324, Jun. 2007. [Online]. Available: <http://doi.acm.org/10.1145/1269899.1254920>
- [78] S. Lakshminarayana, B. Li, M. Assaad, A. Eryilmaz, and M. Debbah, "A fast-CSMA based distributed scheduling algorithm under SINR model," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 1598–1602.

- [79] B. Li and A. Eryilmaz, "Optimal distributed scheduling under time-varying conditions: A fast-CSMA algorithm with applications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3278–3288, July 2013.
- [80] M. Deghel, M. Assaad, M. Debbah, and A. Ephremides, "Queueing stability and CSI probing of a TDD wireless network with interference alignment," *IEEE Transactions on Information Theory*, vol. PP, no. 99, pp. 1–1, 2017.
- [81] —, "Traffic-aware scheduling and feedback allocation in multichannel wireless networks," (*submitted to*) *IEEE Transactions on Wireless Communications*, 2017.
- [82] —, "Traffic-aware scheduling and feedback reporting in multichannel wireless networks with relaying," (*submitted to*) *IEEE Transactions on Information Theory*, 2017.
- [83] M. Deghel, M. Assaad, and M. Debbah, "Opportunistic feedback reporting and scheduling scheme for multichannel wireless networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–7.
- [84] —, "Queueing stability and CSI probing of a TDD wireless network with interference alignment," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 794–798.
- [85] —, "System performance of interference alignment under TDD mode with limited backhaul capacity," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1673–1678.
- [86] M. Deghel, E. Baştuğ, M. Assaad, and M. Debbah, "On the benefits of edge caching for MIMO interference alignment," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 655–659.
- [87] J. Thukral and H. Bölcskei, "Interference alignment with limited feedback," *CoRR*, vol. abs/0905.0374, 2009. [Online]. Available: <http://arxiv.org/abs/0905.0374>
- [88] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*, June 2010, pp. 619–623.
- [89] M. Rezaee and M. Guillaud, "Limited feedback for interference alignment in the  $K$ -user MIMO interference channel," in *Information Theory Workshop (ITW), 2012 IEEE*, Sept 2012, pp. 667–671.
- [90] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, February 2012.
- [91] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," in *IEEE International Conference on Communications Workshops, 2009. ICC Workshops 2009.*, June 2009, pp. 1–5.
- [92] S. Park, O. Simeone, O. Sahin, and S. Shamai, "Performance evaluation of multiterminal backhaul compression for cloud radio access networks," *CoRR*, vol. abs/1311.6492, 2013. [Online]. Available: <http://arxiv.org/abs/1311.6492>

## Bibliography

---

- [93] M. Rezaee, M. Guillaud, and F. Lindqvist, "CSIT sharing over finite capacity backhaul for spatial interference alignment," in *IEEE International Symposium on Information Theory Proceedings (ISIT'13)*. IEEE, July 2013, pp. 569–573.
- [94] E. T. Ar and I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.
- [95] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2008–2026, April 2011.
- [96] H. Boche and M. Wiczanowski, "The interplay of link layer and physical layer under MIMO enhancement: benefits and challenges," *Wireless Communications, IEEE*, vol. 13, no. 4, pp. 48–55, Aug 2006.
- [97] —, "Optimization-theoretic analysis of stability-optimal transmission policy for multiple-antenna multiple-access channel," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2688–2702, June 2007.
- [98] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of gaussian MIMO-BC," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1640–1646, Aug 2006.
- [99] P. Chaporkar, A. Proutiere, H. Asnani, and A. Karandikar, "Scheduling with limited information in wireless systems," in *Proceedings of the Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '09. New York, NY, USA: ACM, 2009, pp. 75–84. [Online]. Available: <http://doi.acm.org/10.1145/1530748.1530759>
- [100] O. El Ayach, A. Lozano, and R. Heath, "On the overhead of interference alignment: Training, feedback, and cooperation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 4192–4203, November 2012.
- [101] S. W. Peters and R. W. Heath, "Interference alignment via alternating minimization," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 2445–2448.
- [102] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, Nov 2006.
- [103] L. Georgiadis, M. J., and R. Tassiulas, "Resource allocation and cross-layer control in wireless networks," in *Foundations and Trends in Networking*, 2006, pp. 1–149.
- [104] X. Chen and C. Yuen, "Performance analysis and optimization for interference alignment over MIMO interference channels with limited feedback," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1785–1795, April 2014.
- [105] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4771–4782, September 2010.



## Bibliography

---

- [106] B. Jóhannesson and N. Giri, “On approximations involving the beta distribution,” *Communications in Statistics-Simulation and Computation*, vol. 24, no. 2, pp. 489–503, 1995.
- [107] X. Chen and C. Yuen, “Performance analysis and optimization for interference alignment over MIMO interference channels with limited feedback,” *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1785–1795, April 2014.
- [108] S. Nadarajah and S. Kotz, “On the product and ratio of gamma and beta random variables,” *Allgemeines Statistisches Archiv*, vol. 89, no. 4, pp. 435–449, 2005. [Online]. Available: <http://dx.doi.org/10.1007/s10182-005-0214-9>
- [109] H. Bateman, A. Erdélyi, H. van Haeringen, and L. Kok, *Tables of integral transforms*. McGraw-Hill New York, 1954, vol. 1.
- [110] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.
- [111] D. Gesbert and M. S. Alouini, “How much feedback is multi-user diversity really worth?” in *2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, vol. 1, June 2004, pp. 234–238.
- [112] T. Tang and R. W. Heath, “Opportunistic feedback for downlink multiuser diversity,” *IEEE Communications Letters*, vol. 9, no. 10, pp. 948–950, Oct 2005.
- [113] P. Svedman, S. K. Wilson, L. J. Cimini, and B. Ottersten, “Opportunistic beamforming and scheduling for OFDMA systems,” *IEEE Transactions on Communications*, vol. 55, no. 5, pp. 941–952, May 2007.
- [114] V. Hassel, D. Gesbert, M.-S. Alouini, and G. E. Oien, “A threshold-based channel state feedback algorithm for modern cellular systems,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2422–2426, July 2007.
- [115] J. Chen, R. A. Berry, and M. L. Honig, “Limited feedback schemes for downlink OFDMA based on sub-channel groups,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1451–1461, October 2008.
- [116] S. N. Donthi and N. B. Mehta, “Joint performance analysis of channel quality indicator feedback schemes and frequency-domain scheduling for LTE,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3096–3109, Sept 2011.
- [117] M. Karaca, Y. Sarikaya, O. Ercetin, T. Alpcan, and H. Boche, “Efficient wireless scheduling with limited channel feedback and performance guarantees,” in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, Sept 2012, pp. 106–111.
- [118] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. B. Shroff, “Scheduling with rate adaptation under incomplete knowledge of channel/estimator statistics,” in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, Sept 2010, pp. 670–677.
- [119] A. T. Gamage, M. S. Alam, X. Shen, and J. W. Mark, “Joint relay, subcarrier and power allocation for OFDMA-based femtocell networks,” in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 679–684.

## Bibliography

---

- [120] C. N. Hsu, H. J. Su, and P. H. Lin, "Joint subcarrier pairing and power allocation for OFDM transmission with decode-and-forward relaying," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 399–414, Jan 2011.
- [121] D. W. K. Ng and R. Schober, "Cross-layer scheduling for OFDMA amplify-and-forward relay networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1443–1458, March 2010.
- [122] M. Hajiaghayi, M. Dong, and B. Liang, "Optimal channel assignment and power allocation for dual-hop multi-channel multi-user relaying," in *2011 Proceedings IEEE INFOCOM*, April 2011, pp. 76–80.
- [123] X. Zhang, S. Chen, and W. Wang, "Multiuser radio resource allocation for multiservice transmission in OFDMA-based cooperative relay networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, p. 940518, 2008. [Online]. Available: <http://dx.doi.org/10.1155/2009/940518>
- [124] Y. Li, W. Wang, J. Kong, W. Hong, X. Zhang, and M. Peng, "Power allocation and subcarrier pairing in OFDM-based relaying networks," in *2008 IEEE International Conference on Communications*, May 2008, pp. 2602–2606.
- [125] M. S. Alam, J. W. Mark, and X. S. Shen, "Relay selection and resource allocation for multi-user cooperative OFDMA networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2193–2205, May 2013.
- [126] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec 2004.
- [127] A. A. Reddy, S. Banerjee, A. Gopalan, S. Shakkottai, and L. Ying, "Wireless scheduling with heterogeneously delayed network-state information," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, Sept 2010, pp. 1577–1584.
- [128] S. Haykin, *Adaptive Filter Theory (3rd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.
- [129] L. Ying, R. Srikant, A. Eryilmaz, and G. E. Dullerud, "A large deviations analysis of scheduling in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5088–5098, Nov 2006.
- [130] P. van de Ven, S. Borst, and S. Shneer, "Instability of maxweight scheduling algorithms," in *IEEE INFOCOM 2009*, April 2009, pp. 1701–1709.

# Ordonnancement et feedback dans les réseaux sans fil avec prise en compte du trafic

## Résumé

La demande des systèmes de communication sans fil pour des débits élevés continue d'augmenter, et il n'y a pas de signes que cette tendance va se ralentir. Trois des techniques les plus importantes qui ont émergé pour répondre à de telles demandes sont l'OFDMA, le relais coopératif et le MIMO. Afin d'utiliser pleinement les capacités des systèmes appliquant de telles techniques, il est essentiel de développer des algorithmes efficaces d'ordonnancement et, plus généralement, des algorithmes efficaces d'allocation de ressources. Les études classiques sur ce sujet examinent des systèmes où les demandes de données des utilisateurs ne sont pas prises en considération et/ou un CSI parfait et complet est supposé être disponible pour le mécanisme d'ordonnancement. Cependant, dans la pratique, différentes limitations peuvent entraîner l'absence d'une connaissance parfaite et/ou complète du CSI, telles que les ressources limitées pour le feedback, le coût de sondage et le retard dans le processus de feedback.

Par conséquent, dans cette thèse nous examinons les problèmes d'ordonnancement et de feedback sous des considérations réalistes concernant la connaissance du CSI. L'analyse est effectuée au niveau des paquets et considère la dynamique des files d'attente avec des processus d'arrivée arbitraires, et où la mesure de performance principale que nous adoptons est la stabilité des files d'attente. La première partie de la thèse considère un système MIMO multipoint à multipoint utilisant le mode TDD, tout en supposant un backhaul à capacité limitée et en tenant compte du coût du feedback. En ce qui concerne la technique de gestion de l'interférence, nous appliquons l'alignement d'interférence (IA) si plus d'une paire sont actives et SVD si une seule paire est active. La deuxième partie de la thèse considère un système OFDMA avec plusieurs utilisateurs et canaux, où un feedback retardé et limité est pris en compte. Deux scénarios sont étudiés, à savoir le système sans relais et le système avec relais. Pour ce dernier, nous considérons une imperfection supplémentaire supposant que les utilisateurs ont une connaissance incomplète des coefficients du fading entre la station de base et le relais.

Nous supposons que les données destinées à chaque utilisateur sont stockées dans une file d'attente respective qui se trouve au niveau du contrôleur (par exemple, station de base). La dynamique de ces files d'attente est considérée dans l'analyse du système. En effet, il est démontré que l'information sur la longueur de la file d'attente est très importante dans la conception de politiques d'ordonnancement robustes et plus généralement des algorithmes d'allocation de ressources robustes; telles politiques assurent des débits de données élevés et des faibles retards de transmission des paquets en présence de canaux qui varient dans le temps et pour diverses demandes de données par les utilisateurs. L'objectif principal du mécanisme d'ordonnancement dans cette thèse est de stabiliser le système et d'atteindre ainsi un débit maximal et de maintenir un faible retard de paquets.

Cette thèse met en valeur l'information concernant l'état du réseau, en particulier l'état des canaux, pour l'ordonnancement des utilisateurs dans les réseaux sans fil. Une idée centrale ponctuant l'ensemble du travail est que la structure d'information disponible pour le mécanisme d'ordonnancement influence la performance du système. Avec ses résultats, cette thèse constitue un pas en avant vers une

théorie plus complète abordant les problèmes d'ordonnancement et d'allocation de ressources sous différents types d'informations incomplètes. Comme les systèmes de communication sans fil connaissent une forte augmentation dans les demandes de performance, sans aucun signe de ralentissement, nous estimons qu'un tel cadre général pour enquêter sur l'ordonnancement dans un système sans fil avec une connaissance incomplète de l'état du système se révélera un atout inestimable dans la conception de systèmes performants.

**Titre:** Ordonnancement et feedback dans les réseaux sans fil avec prise en compte du trafic

**Mots clés:** stabilité des files d'attente, ordonnancement, allocation de feedback, réseaux sans fil, 5G

**Résumé:** La demande des systèmes de communication sans fil pour des débits élevés continue d'augmenter, et il n'y a pas de signes que cette tendance va se ralentir. Trois des techniques les plus importantes qui ont émergé pour répondre à de telles demandes sont l'OFDMA, le relais coopératif et le MIMO. Afin d'utiliser pleinement les capacités des systèmes appliquant de telles techniques, il est essentiel de développer des algorithmes efficaces d'ordonnancement et, plus généralement, des algorithmes efficaces d'allocation de ressources. Les études classiques sur ce sujet examinent des systèmes où les demandes de données des utilisateurs ne sont pas prises en considération et/ou un CSI parfait et complet est supposé être disponible pour le mécanisme d'ordonnancement. Cependant, dans la pratique, différentes limitations peuvent entraîner l'absence d'une connaissance parfaite et/ou complète du CSI, telles que les ressources limitées pour le feedback, le coût de sondage et le retard dans le processus de feedback.

Par conséquent, dans cette thèse nous examinons les prob-

lèmes d'ordonnancement et de feedback sous des considérations réalistes concernant la connaissance du CSI. L'analyse est effectuée au niveau des paquets et considère la dynamique des files d'attente avec des processus d'arrivée arbitraires, et où la mesure de performance principale que nous adoptons est la stabilité des files d'attente. La première partie de la thèse considère un système MIMO multipoint à multipoint utilisant le mode TDD, tout en supposant un backhaul à capacité limitée et en tenant compte du coût du feedback. En ce qui concerne la technique de gestion de l'interférence, nous appliquons l'alignement d'interférence (IA) si plus d'une paire sont actives et SVD si une seule paire est active. La deuxième partie de la thèse considère un système OFDMA avec plusieurs utilisateurs et canaux, où un feedback retardé et limité est pris en compte. Deux scénarios sont étudiés, à savoir le système sans relais et le système avec relais. Pour ce dernier, nous considérons une imperfection supplémentaire supposant que les utilisateurs ont une connaissance incomplète des coefficients du fading entre la station de base et le relais.

**Title:** Traffic-aware scheduling and feedback reporting in wireless networks

**Keywords:** Queueing stability, scheduling, feedback allocation, wireless networks, 5G

**Abstract:** Demand of wireless communication systems for high throughputs continues to increase, and there are no signs this trend is slowing down. Three of the most prominent techniques that have emerged to meet such demands are OFDMA, cooperative relaying and MIMO. To fully utilize the capabilities of systems applying such techniques, it is essential to develop efficient scheduling algorithms and, more generally, efficient resource allocation algorithms. Classical studies on this subject investigate in much detail settings where the data requests of the users are not taken into consideration or where the perfect and full CSI is assumed to be available for the scheduling mechanism. In practice, however, different limitations may result in not having perfect or full CSI knowledge, such as limited feedback resources, probing cost and delay in the feedback process.

Accordingly, in this thesis we examine the problems of scheduling and feedback allocations under realistic con-

siderations concerning the CSI knowledge. Analysis is performed at the packet level and considers the queueing dynamics in the systems with arbitrary arrival processes, where the main performance metric we adopt is the stability of the queues. The first part of the thesis considers a multi-point to multi-point MIMO system with TDD mode under limited backhaul capacity and taking into account the feedback probing cost. Regarding the interference management technique, we apply interference alignment (IA) if more than one pair are active and SVD if only one pair is active. The second part of the thesis considers a multiuser multichannel OFDMA-like system where delayed and limited feedback is accounted for. Two scenarios are investigated, namely the system without relaying and the system with relaying. For the latter one, an additional imperfection we account for is that the users have incomplete knowledge of the fading coefficients between the base-station and the relay.