Practical interference management strategies in Gaussian networks

by

Seyed Ali Hesammohseni

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2016

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Increasing demand for bandwidth intensive activities on high-penetration wireless hand-held personal devices, combined with their processing power and advanced radio features, has necessitated a new look at the problems of resource provisioning and distributed management of coexistence in wireless networks. Information theory, as the science of studying the ultimate limits of communication efficiency, plays an important role in outlining guiding principles in the design and analysis of such communication schemes. Network information theory, the branch of information theory that investigates problems of multiuser and distributed nature in information transmission is ideally poised to answer questions about the design and analysis of multiuser communication systems. In the past few years, there have been major advances in network information theory, in particular in the generalized degrees of freedom framework for asymptotic analysis and interference alignment which have led to constant gap to capacity results for Gaussian interference channels. Unfortunately, practical adoption of these results has been slowed by their reliance on unrealistic assumptions like perfect channel state information at the transmitter and intricate constructions based on alignment over transcendental dimensions of real numbers. It is therefore necessary to devise transmission methods and coexistence schemes that fall under the umbrella of existing interference management and cognitive radio toolbox and deliver close to optimal performance.

In this thesis we work on the theme of designing and characterizing the performance of conceptually simple transmission schemes that are robust and achieve performance that is close to optimal. In particular, our work is broadly divided into two parts. In the first part, looking at cognitive radio networks, we seek to relax the assumption of non-causal knowledge of primary user's message at the secondary user's transmitter. We study a cognitive channel

model based on Gaussian interference channel that does not assume anything about users other than primary user's priority over secondary user in reaching its desired quality of service. We characterize this quality of service requirement as a minimum rate that the primary user should be able to achieve. Studying the achievable performance of simple encoding and decoding schemes in this scenario, we propose a few different simple encoding schemes and explore different decoder designs. We show that surprisingly, all these schemes achieve the same rate region. Next, we study the problem of rate maximization faced by the secondary user subject to primary's QoS constraint. We show that this problem is not convex or smooth in general. We then use the symmetry properties of the problem to reduce its solution to a feasibly implementable line search. We also provide numerical results to demonstrate the performance of the scheme.

Continuing on the theme of simple yet well-performing schemes for wireless networks, in the second part of the thesis, we direct our attention from two-user cognitive networks to the problem of smart interference management in large wireless networks. Here, we study the problem of interference-aware wireless link scheduling. Link scheduling is the problem of allocating a set of transmission requests into as small a set of time slots as possible such that all transmissions satisfy some condition of feasibility. The feasibility criterion has traditionally been lack of pair of links that interfere too much. This makes the problem amenable to solution using graph theoretical tools. Inspired by the recent results that the simple approach of treating interference as noise achieves maximal Generalized Degrees of Freedom (which is a measure that roughly captures how many equivalent single-user channels are contained in a given multi-user channel) and the generalization that it can attain rates within a constant gap of the capacity for a large class of Gaussian interference networks, we study the problem of scheduling links under a set Signal to Interference plus Noise Ratio (SINR) constraint. We show that for nodes distributed in a metric space and obeying path loss channel model, a refined framework based on combining geometric and graph theoretic results can be devised to analyze the problem of finding the feasible sets of transmissions for a given level of desired SINR. We use this general framework to give a link scheduling algorithm that is provably within a logarithmic factor of the best possible schedule. Numerical simulations confirm that this approach outperforms other recently proposed SINR-based approaches. Finally, we

conclude by identifying open problems and possible directions for extending these results.

Acknowledgments

I would like to thank all members of my PhD committee for taking time out of their busy schedules to be part of the committee and to give their comments. Their suggestions have been extremely valuable. Special thanks goes to Professor Catherine Rosenberg for very fruitful discussions and her crucial guidance on formulating the mixed integer program for link scheduling and in generously giving me permission to use her group's server to perform numerical simulations. Professors Stephen Smith and Wei Yu immensely helped with their suggestions about clarifying the description of the algorithm and expanding the discussion on the complexity of the scheduling problem in Chapter 4. Remarks by professors Richard Trefler and Stephen Smith helped me in clarifying some aspects of the presentation of channel models in chapters 2 and 3 that were unclear.

Doing a PhD is a long and arduous undertaking that is impossible without the help and support of many people.

First and foremost, I have to thank my supervisors professors Catherine C. Gebotys and Mohamed O. Damen. Professor Damen has always been there to offer technical insights, pointers to the relevant literature and has overall been an excellent listener and outstanding technical critique. Professor Gebotys has been nothing but helpful in in all matters whether scientific, navigating departmental paperwork or general life advice. I also acknowledge the opportunity to work with Professor Amir K. Khandani in coding and signal transmission laboratory and especially of collaborating with Dr. Kamyar Moshksar.

I would also like to thank my parents and my sister for their support. They have always believed in my potential and been supporting of my varied endeavours. Words are not able to express the debt of gratitude I owe my parents, Forouzandeh and Jafar, for instilling in me the qualities that shape my personality to this day. My sister, Maryam, has always been a constant source of moral support, an excellent conversational companion over matters mundane or profound and over distances long or short, and an overall great source of and hope and energy when goings got tough.

Finally, I have been blessed to have experienced the great friendship of many people with whom I have spent great moments during my time Waterloo. I like to thank, in no particular order Reza, Daniel, Sina, Ershad, Sandy, Mina, Soroosh, Sarah, Nasser, Patty, Behnoush,

Mahyar, Tirdad, Sepideh and many others that I have undoubtedly inadvertently forgot.

Dedication

Dedicated to my family, with love and admiration.

Table of Contents

List of Figures **xi**

List of Symbols **xiv**

1 Introduction **1**

2 Background and preliminaries **9**

2.1 Network information theory 10

2.1.1 An information-theoretic view of cognitive radio and cooperative communications 14

2.2 Wireless link scheduling 19

2.3 Review of results on interference channel 25

2.3.1 Formal definition of discrete memoryless single user and interference channel and their associated capacity regions 25

2.3.2 Formal definition of the Gaussian interference channel 28

2.3.3 Capacity region of the Gaussian interference channel 31

2.3.4 Achievability schemes for the interference channel 34

2.3.5 Generalized Degrees of Freedom (GDoF) 36

2.4 Interference alignment 38

2.5 Optimality of Treating Interference as Noise (TIN) 39

3 Two-user cognitive GIC **41**

3.1 The model and problem 42

3.1.1 The channel model 42

3.1.2 The problem statement . . . 43
3.1.3 Example of a practical application . . . 43
3.2 The achievable rate region $\mathcal{R}$ . . . 44
3.3 Remarks on encoder and decoder structure . . . 46
3.3.1 Non-unique joint typicality decoding . . . 46
3.3.2 Multi-layer encoding and successive cancellation . . . 48
3.4 Analysis of the rate-optimization problem . . . 51
3.5 Relative magnitude of f and g . . . 53
3.5.1 Characterizing the boundaries of $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ . . . 54
3.6 Possible extreme cases of the problem . . . 57
3.6.1 $W_c = 0$ . . . 57
3.6.2 $W_c = W_0$ . . . 57
3.7 Numerical Examples . . . 57
3.7.1 Maximum can be attained in different parts of the boundary . . . 59
3.7.2 Sensitivity of achievable rates to changes in parameter values . . . 61
3.8 Shape of the rate curve . . . 61
3.9 Conclusion . . . 62

4 Approximate link scheduling in large networks 65
4.1 Model and assumptions . . . 67
4.2 Formal definition of scheduling problem . . . 69
4.2.1 Example of a practical application scenario . . . 71
4.2.2 Complexity of exact scheduling . . . 72
4.3 Mixed integer programming formulation . . . 73
4.3.1 First approach . . . 76
4.3.2 Adding ordering constraints to reduce symmetry . . . 79
4.4 Proposed approximate algorithm . . . 81
4.4.1 Notation and preliminaries . . . 81
4.4.2 Description of the algorithm, its correctness and performance . . . 85
4.5 Simulations and conclusion . . . 96

4.5.1 Setup and choice of parameters 96
4.5.2 Comparison with exact solution algorithms 97
4.5.3 Throughput performance in large-network scenario 99

5 Summary of contributions and future work 103
5.1 Summary of contributions 103
5.2 Limitations 104
5.3 Comparison 104
5.4 Future Work 105

Bibliography 106

Appendix A Proofs from chapter 3 121
A.1 Proof of Claim 3.1 121
A.2 Proof of Claim 3.2 123
A.3 Proof of Lemma 3.1 124
A.4 Proof of claim 3.3 125
A.5 Proof of Claim 3.4 126
A.6 Proof of Lemma 3.2 126
A.7 Proof of proposition 3.1 128

Appendix B Proofs from chapter 4 129
B.1 Proof of Lemma 4.1 129
B.2 Proof of Lemma 4.2 132
B.3 Proof of Lemma 4.3 134
B.4 Proof of Lemma 4.4 135
B.5 Proof of Theorem 4.1 136
B.6 Proof of Theorem 4.2 137

List of Figures

2.1 Two-user multiple-access channel 10
2.2 Two-user broadcast channel 11
2.3 Relay channel 12
2.4 Interference channel when used n times with encoders and decoders 13
2.5 Two-user discrete memoryless interference channel 26
2.6 Coding and decoding setup for the two-user discrete memoryless interference channel 27
2.7 Gaussian Interference Channel (GIC) 28
2.8 Band-limited Gaussian interference channel 30

3.1 Band-limited Gaussian interference channel, $BL(W_1, W_2)$ denotes the class of signals limited to the (W_1, W_2) band and $S_z(f)$ is the spectral density of noise 42
3.2 The chimney rate region $\mathcal{R}$ described in 3.7. 46
3.3 The region $\mathcal{D}_1$. 58
3.4 The region $\mathcal{D}_2$. 59
3.5 Plot of s_2 in terms of W_c on the part of $\partial_{\text{rm}}\mathcal{D}_1$ represented by $Q = \min\{f(W_c), g(W_c)\}$. 59
3.6 Plot of s_2 in terms of Q on the part of $\partial_{\text{rm}}\mathcal{D}_1$ that is a vertical line segment 60
3.7 Plot of $r_2 + s_1 - R_{th}$ in terms of W_c on $\partial_{\text{rm}}\mathcal{D}_2$ represented by $Q = g(W_c)$. . 60
3.8 Plot of maximum achievable R_2 as a function of changing R^*. 61
3.9 Plot of maximum achievable R_2 as a function of changing a_{12}. 62
3.10 Achievable rate region for some specific parameter values 63
3.11 Achievable rate region for some specific parameter values 64

4.1 Flowchart for Algorithm 1. 87
4.2 Flowchart for Algorithm 3. 90
4.3 Output schedule length of proposed algorithm compared to the bounds obtained by mixed integer programming 99
4.4 Sum-rate comparison of our algorithm with FlashLinQ, ITLinQ and no scheduling 101

List of Symbols

α Path loss exponent

β SINR threshold

$\chi(G)$ Chromatic number of graph G

Δ Maximum to minimum link length in network

$\ell_l, D(l, l')$ Length of link l, distance of origin of l to destination of l'

$[N]$ The set $\{1, ..., N\}$

$\mathbb{1}_A$ Indicator function of the set A

$\mathcal{R}$ Rate regions

$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ Channel input and output alphabets

$C(x)$ $\frac{1}{2}\log(1+x)$, capacity of a Gaussian channel with signal power to noise variance ratio of x

$F(P, N)$ $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$

i, j, m, n Integers

$I_A(B)$ Normalized ISR (affectance) of set of links B on set of links A

l Network link

$o(l), d(l)$ Origin and destination of link l

P, Q Transmit power values

r, s, t Rate expressions (with appropriate subscripts)

R_i Achievable rate for user i

s Scheduling-independent set or ISet

$x(t), f, g$ Real-valued functions of a real variable

X, Y, Z Random Variables

Chapter 1

Introduction

Cognitive radio, as first proposed by Mitola [1] is an important research direction in engineering next generation wireless systems [2, 3]. The premise of cognitive radio is based on the observations by regulatory authorities in many countries that despite the heavy increase in demand for wireless spectrum, most traditional band licensees are not using their allocated spectrum efficiently at all. In particular, there exists a hierarchy of legacy radio users who often use decades old technology and vintage band licenses and whose efficiency of spectrum use is far from optimal, and another class of highly agile and capable radios that can potentially tap into the unused portion of these users' spectrum at the same time as being cognizant of the very strict quality of service requirements that these incumbent users demand. In effect, a cognitive radio is any radio system that is simultaneously configurable in its parameters and aware of the wireless environment it is operating in. Using this knowledge, the cognitive radio tries to opportunistically adapt its transmission parameters in such a way as to maximize its resource usage efficiency and minimize undesirable interference on the user that has the primary priority to the spectrum it is using.

Also, despite the fact that having a network of extremely capable and context-aware but non-cooperating cognitive radios is a huge improvement over the current architecture, there are potential performance gains to be made by making these intelligent nodes able to

cooperate with one another. Research into the gains from cooperation is mostly inspired by the very promising theoretical results on the gains to diversity and multiplexing possible through the use of multiple-antenna systems and generalized beam-forming, and their successful implementation in such standards as IEEE 802.11n WiFi [4].

The so called cooperative cognitive radio schemes have attracted interest for the design of next generation wireless systems. These systems try to replicate the gains obtained from multiple antenna transmit and receive strategies using a heterogeneous and distributed network of (possibly single-antenna) transmitters and receivers. This is done through forming a distributed virtual antenna array across multiple nodes. It is obvious that when no centralized coordination is involved, some overhead and therefore loss of efficiency is to be expected but the hope is that in cases of interest, this loss of efficiency is more than compensated by the gains achieved through these distributed beam-forming schemes.

These developments, against the backdrop of the huge increase in the number of wireless-capable personal mobile devices over the past few years, have rekindled research interest into the use of ever-more complex and adaptive transmit and receive strategies that exploit the specific properties of these types of decentralized heterogeneous networks at the same time as being aware of the very real limitations in channel quality, delay tolerance, transmit power and channel estimation accuracy that these platforms inherently suffer from. Communication over wireless radio channels has to contend with many problems that are not of a serious concern for guided media like wire-line and fibre optics. The most challenging among these problems is the fact that the free space is a shared resource and that radio channels, by virtue of the flexibility of their setup, often present much less favourable conditions for transmitting data than their wire-line counterparts and suffer from the effects of multi-path fading and time-dependent shadowing. Also, the interactive nature of some of the communication services offered by these devices often means that delay constraints are very strict, whereas for some other usage scenarios, long delays can be tolerated. Until very recently, information-theoretic results had almost no bearing on engineering approaches to the design

of these systems in the real world [5, 6, 7]. As a result of this, there was often a lack of deep knowledge of the effects of these non-idealities and limitations and perhaps opportunities and advantages offered by them and as such, there was an unfulfilled demand for a much deeper understanding of the fundamental trade-offs underlying communication in the presence of a wide range of interferers and uncertainty in the specification of the transmission medium. This has spurred interest in new research directions and design paradigms that try to incorporate the specific properties and challenges that are faced by the designer of a decentralized wireless ad-hoc network and in particular to propose new classes of system designs that are specifically tailored to such limitations [8, 9].

This requires gaining a broad insight into the applicability of any proposed scheme along these ideas, and a thorough understanding of the theoretical possibilities and limitations of communication over radio networks. This type of analysis, of what is fundamentally achievable and what is not, becomes especially important when trying to decide on a benchmark or figure of merit against which to evaluate the performance of different classes of real-world systems. This is because such ultimate performance limits can never be achieved by any real world system but can typically be approached very closely by highly optimized and clever designs. Therefore, they serve as a single point of reference against which different systems with different underlying architectures can be compared without any bias toward any particular approach to the problem. This is where the role of network information theory becomes apparent. Since the seminal work of Shannon [10], the probabilistic framework offered by information theory has shown to be an invaluable tool in analyzing the ultimate limits on the transmission of information under various adverse scenarios and in evaluating the improvement headroom available to any real-world communication system. Network information theory is a natural extension of this point-to-point formalism that tries to quantify and study the effects of competition, cooperation and distributed operation on the fundamental possibilities in transfer of information and hence is naturally suited as a firm theoretical ground for analyzing and gaining deep insights into the broad design problems facing the

next generation's network engineers.

Therefore, it seems important to propose and analyze idealized models of communication scenarios that might arise in the context of such channel-aware, heterogeneously capable and multi-tiered communication systems as proposed under the banner of cognitive, cooperative and device-to-device communications and to use the tools of multi-user information theory to study and analyze these problems.

In the first part of this thesis, we propose one such model of a cognitive channel in a two-user setting. The main characteristic of our problem setup is that the users are not symmetric in their priority of access to the channel and in their capability to adapt themselves to its particular realization. Specifically, we have a primary or legacy user, who is not expected to accommodate the bandwidth needs of the other user, nor is it expected to use advanced detection and interference management techniques in decoding its desired signal. The secondary user on the other hand, should guarantee that its presence in the band does not cause the attainable rate of the first user to fall below a certain threshold. It has a range of adaptive tools and strategies at its disposal to asses and minimize its effect on the primary user's quality of service and to squeeze the maximum possible performance out of the available spectral resources for transferring its own data. We characterize an achievable rate region for primary and secondary user of the channel. We then show that a number of alternate encoder and decoder architectures give rise to the same rate region as achieved by our first encoding scheme. We also derive a weak converse result, showing that our rate region cannot be improved by adding multilayer random coding to the cognitive transmitter's codeword. Because our problem setup involves a rate-optimizing cognitive secondary user, we next state and analyze the optimization problem that this secondary user has to solve in order to attain maximum transmission rate. We use the properties of the rate-expressions involved and the symmetries of the problem to reduce this rate-optimization problem to a number of simpler constituent problems. We also analyze and derive sufficient conditions on the channel coefficients under which some of these subproblems will dominate the others.

Next, to gain insights into the performance of the proposed transmission schemes and our decomposition of the rate-optimization problem, we provide illustrative numerical examples and simulations and interpret the plotted results.

The second part of the thesis concerns the problem of link scheduling in larger wireless networks. Link scheduling in a broadcast propagation medium is the problem of partitioning a set of network transfer requests across the smallest possible set of timeslots. There is a trade-off between utilization of the common medium and quality of individual links in broadcast networks and too many simultaneously transferring links leads to transmission failures. As such, some metric of link quality should be maintained while trying to satisfy different requests simultaneously. Traditionally, in designing algorithms for dynamic link scheduling, interference is looked at as an all or nothing phenomenon. In this view, each pair of links either conflict or not. This has the advantage of making the problem simpler to conceptualize and gives rise to notions such as radius of interference and guard intervals around transmitting nodes that preclude other transmissions. Although it leads to straightforward scheduling methods, the pairwise conflict model of transmission feasibility can be very far from a realistic representation with respect to the underlying physical layer. The failure or success of network links at clearing transmission demands directly depends on their rate which itself depends on the signal to interference and noise ratio (SINR) seen at the receivers. As will be discussed next, signal to interference and noise ratio has also been shown to be fundamental to characterizing channel capacity for large networks.

The connection between SINR and channel capacity is established using degrees of freedom analysis [11]. The framework of Degrees-of-Freedom (DoF) has emerged over the past decade as a powerful tool in analyzing and understanding the asymptotic behaviour of wireless channel capacity in the limit of high SNR. Degrees-of-Freedom (DoF) of a multiple-input multiple-output channel is the multiple of the capacity of a single-input single-output channel it is capable of transferring at high SNR values. A channel with degree of freedom N behaves like N parallel SISO channels at high SNR values. Each of these equivalent SISO channels is

known as one degree of freedom of the larger channel. The adoption of DoF framework has also paved the way for the introduction of interference alignment. It was through the use of interference alignment that the $N/2$ degrees of freedom of an N-user interference channel was established. This showed that in many cases, judicious design of signals at the transmit side and simple treating of interference as noise at the receive-side can achieve rates within a constant gap to the capacity. Unfortunately, interference alignment results are not thought to be robust enough to be applicable in many real-world scenarios [12, 13], but they point towards the power of simple schemes in interference management. More recently, it has been shown that simply treating interference as noise, even without alignment at the transmitters, can achieve the same performance for large classes of interference channel[1].

This opens up the potential for scheduling algorithms that directly target Signal to Interference and Noise Ratio (SINR) constraints, as it is a metric that captures the achievable rate under these conditions. This is the problem we tackle in the second part of this thesis. Specifically, we look at an ad-hoc network of wireless nodes and adopt a path loss model of channel coefficients. We show that unlike the previous approaches that mostly looked at the problem of link scheduling in terms of pairwise conflicts between different links, which are straightforwardly modeled by a conflict graph, additional subtleties are involved when the problem is studied under signal plus interference and noise ratio constraints. In particular, because of the accumulative nature of interference on the noise floor, it seems hard to pick up feasible subsets of links without incurring the costs of a combinatorial search. We show that under quite general assumptions on the distribution of nodes, a pairwise relaxation of the notion of SINR-feasibility can be obtained. This approach allows us to still use the graph-based model for link scheduling, while remaining faithful to the SINR model of radio operation. In particular, we use this refined graph-based analysis of the scheduling conflict to derive an algorithm for SINR-feasible link scheduling that has provable approximation

[1]The next chapter, after going over the required background, gives a comprehensive review of Degrees-of-Freedom framework and its generalization in Generalized Degrees-of-Freedom analysis for multi-user channels and how they have paved the way for most of the recent advances in network information theory

guarantees. Moreover, we use simulations to show that this algorithm compares favourably with state of the art scheduling algorithms that have been proposed for scenarios similar to ours.

The rest of this document is organized as follows: Chapter 2 gives a background of a few of the canonical problems in network information theory, presenting a brief review of information-theoretic work done on analyzing cognitive radio and cooperative communications. We then review the problem of link scheduling in wireless networks and discuss the prior work that mostly concerns pairwise notions of scheduling conflict. Finally, we go into depth on the formal definition of interference channel upon which our models are based and a selection of results on capacity, achievable rates and outer bounds are reviewed. This includes a look at the generalized degrees of freedom work and results on the optimality of treating interference as noise. Chapter 3 contains the first part of the thesis. In this part we analyze a model of cognitive Gaussian interference channels that does not presume non-causal knowledge of primary user's message by the secondary user. After formally defining the model, we analyze several transmission strategies and derive their achievable regions. We also show that our achievable rates cannot be improved upon by random multi-layer coding of the type used in the vast majority of achievability results in network information theory. Having characterized an achievable region for this channel, we formulate the rate optimization problem for our setup and use the structure of this optimization problem to simplify and categorize its different working regimes. This results in a breakdown of the problem into a family of one-dimensional optimization problems with solutions corresponding to these different regimes. Next, we give a number of demonstrative numerical examples to gain insight into the available performance.

Chapter 4 contains the second part of the thesis where we define the scheduling problem that we are trying to solve and argue its importance. We then formally establish our model and assumptions. We show that this problem can be exactly solved by formulating as a mixed integer program, but exact solution is not tractable for larger networks. Then, after

defining relevant notation and terminology, we show that the SINR-feasibility criterion can be cast into the language of graph-theoretic independent set scheduling. We do this through a pairwise relaxation of the notion of SINR-feasibility that allows for a graph-theoretical analysis facilitating the use of existing graph-theoretic tools to bear on the problem, but is still refined enough to be related to the optimal solution with a provable approximation ratio. We then state the algorithm and its approximation ratio. Numerical results about its performance are also provided. Finally, the last chapter discusses some possible directions for extending the model and some related problems for future work.

Chapter 2

Background and preliminaries

In this chapter, we start by giving a background of some of the canonical problems of network information theory and the state of their resolution in various special cases and their variations and generalizations. We then briefly review the literature on information-theoretic approaches to cognitive radio networks and in particular review a few works whose model is similar to our model of the cognitive channel. We then review the problem of link scheduling in large wireless networks and review the existing work in this area. This will serve as a brief overview on the state of progress, both in the broader field of information theory and in the special case of information-theoretic investigations of cognitive communication problems and network link scheduling. Next, we will have a whirlwind tour of the interference channel, arguably the most important channel model in multi-user information theory, as this is the model that underlies the work of this thesis. We then review both classic and very recent results on the capacity of interference channels. This includes a review of the celebrated work on interference alignment for interference channels with N users where $N > 2$, and the more recent results on the optimality of treating interference as noise for large classes of interference channels.

2.1 Network information theory

Network information theory owes its starting to the work of Shannon on two-way channels [14]. This was the first time that Shannon's own approach to the mathematical theory of information transmission [10] from a decade earlier was extended to a communication scenario in which more than a one transmitter-receiver pair are involved and there is a trade-off between the users' utilization of the channel. Perhaps not coincidentally, this is also the first time that a skeleton of a model that would later become the interference channel was discussed in the literature. Shannon only succeeded in proving a capacity result for the special Gaussian case of the two-way Channel. In all other cases, problem has proven difficult to solve. This demonstrates that there are subtleties involved in solving the problem of reliable communications when more than one user and terminal is involved and that these difficulties are substantially different from those faced in single-user information theory.

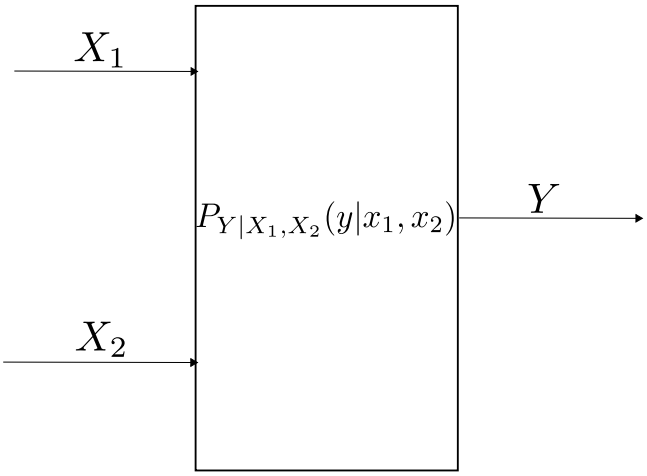

Figure 2.1: Two-user multiple-access channel

During the couple of decades after this paper a multitude of different multi-user communication channels were introduced. Here, we will briefly go over a few of the canonical channel models in network information theory.

2.1. Network information theory

The two-user Multiple Access Channel (MAC) shown in figure 2.1, was introduced and solved simultaneously by Ahlswede [15] and Liao [16, 17]. It is intended to model the case where a single receiver tries to decode two messages sent simultaneously and independently by two separate transmitters, as for example might be faced by the BTS[1] in the up-link of a cellular network. Among others, this model can be trivially extended to the N-user case and also the case where synchronization between the transmitters and the receiver is not perfect. This problem is to date, the only one of the canonical problems in information theory to solved satisfactorily in the general case.

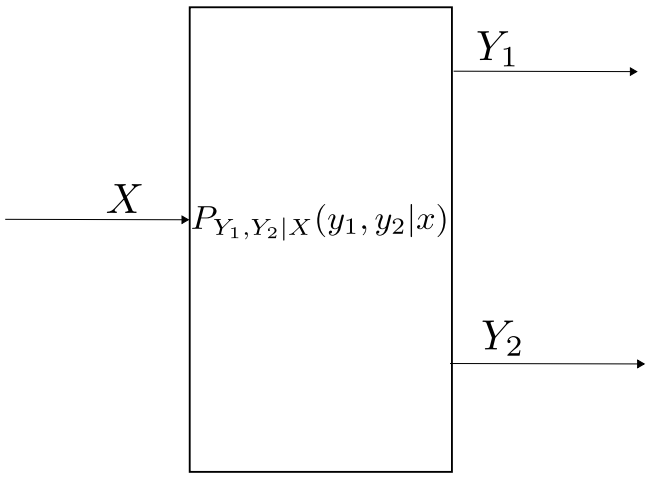

Figure 2.2: Two-user broadcast channel

The two-user Broadcast Channel, introduced by Cover [18] and shown in figure 2.2, attempts to model an operational dual to the Multiple-access channel by modeling the scenario in which a single transmitter is trying to send two separate messages to two independent receivers, as for example would be the case in the down-link of a cellular network. Some of the ways in which this model can be generalized are by extending to N receivers and to the case in which there is a common message as well as each receiver's private message in which all receivers are interested. In the special case of degraded broadcast channel, in which one

[1]Base transceiver station.

of the receivers has a degraded version of the signal from the other receiver, namely that its signal is independent of the transmitted signal given what the superior receiver has received, Cover conjectured in [18] and Bergmans obtained in [19], using what was termed superposition coding, an achievable rate region. Proof of the converse coding theorem for this region was given by Gallager and Bergmans [20, 21]. Fortunately, in the Gaussian case of the two-user broadcast channel this condition always holds, that is the receiver with the lower SNR receives a stochastically degraded version of the signal from the receiver with the higher SNR. Otherwise, for most other scenarios, the calculation of the capacity region[2] of the broadcast channel remains an open problem. More recently, using ideas from dirty-paper coding [22] and Gel'fand-Pinsker problem[23], the capacity region of the MIMO[3] Gaussian version of the broadcast channel was derived by Shamai and Caire [24] and almost concurrently by a few other groups [25, 26, 27]. Other than these classes and some other special cases the problem of characterization of the capacity region of the broadcast channel remains open.

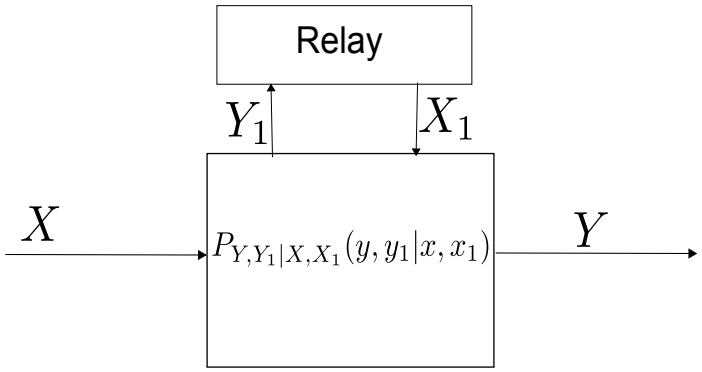

Figure 2.3: Relay channel

[2]The capacity region of a channel, which will be formally defined in the sequel, denotes the set of transfer rates (or rate tuples for multi-user channels) which can be supported with asymptotically vanishing error probability on that channel.

[3]Multiple-input Multiple Output.

2.1. Network information theory

In the Relay Channel, first proposed by van der Muelen [28, 29, 30] and shown in figure 2.3, a transmitter-receiver pair are trying to communicate a message with the help of a third node, a so called relay, that perhaps has a better line of sight to the transmitter. The relay can listen to what is transmitted by the transmitter (represented by Y_1) and uses these observations causally to help the receiver in decoding the codeword by sending the input X_1 over the network. Variations of this problem exist in which there is more than one relay node, connected either serially in a multi-hop topology or in a single-hop parallel topology (the so called diamond relay network). Cover and El-Gamal [31] proposed two different coding schemes for the classical relay channel, namely decode-forward and compress-forward. They also derived the cut-set outer bound on relay channel capacity. Using these inner and outer bounds they proved the capacity for the two special cases of degraded and reversely degraded relay channels. In both these cases, the cut-set bound coincided with the achievable rate but otherwise no converse proof for the capacity of general relay channels is known.

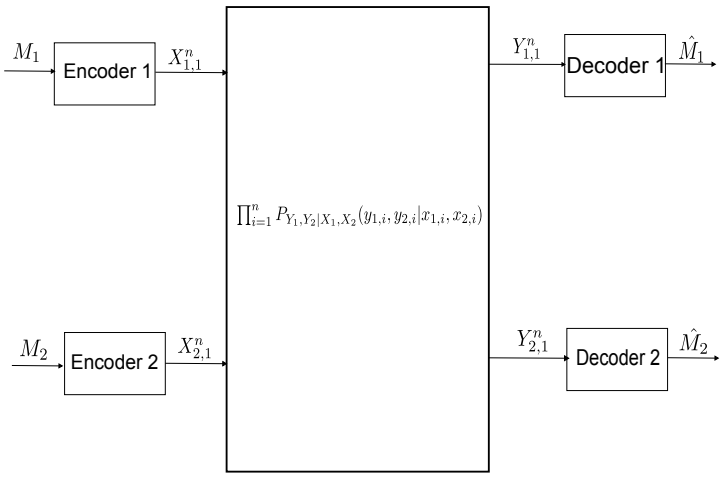

Figure 2.4: Interference channel when used n times with encoders and decoders

The Interference Channel (IC) (figure 2.4) was introduced by Shannon [14] and expounded on by Ahlswede [32]. In this channel two sender transmitter pairs are trying to exchange messages through some shared information transfer medium. One important point about

the interference channel is that its definition subsumes multiple-access and broadcast channels as special cases (by setting the outputs or inputs to be probabilistically equivalent). The determination of the exact capacity region of the interference channel has remained an open problem for nearly four decades and is probably the most challenging of the canonical problems of network information theory. The original inspiration for this model was to analyze the effect of crosstalk on the performance of communication systems using adjacent twisted copper wire pairs. With the advent of ubiquitous wireless terminals and the inherent broadcast nature of wireless networks, namely that every receiver in the range can hear every transmission whether intended for it or not, this problem has gained new-found importance as a model central to the characterization of wireless network performance limits. Since the model of interference channel is central to our work its formal definition and a thorough review of relevant results on its capacity and achievable rate regions under different scenarios will be given in the sequel.

2.1.1 An information-theoretic view of cognitive radio and cooperative communications

As we saw previously, cognitive radio and cooperative communication are believed by many to be an important research direction in the field of wireless communications and as such they are subject to heavy research activity focused on proposing, analyzing and implementing practical and theoretical systems and protocols in order to identify the most promising approaches and the possible gains from using these technologies.

An important facet of this research effort is building measures for quantifying the performance of any given scheme. The performance offered by an idealized version of the problem we are trying to solve in which computational and delay constraints are done away with, is a good benchmark to compare practical approaches against, since it is a fundamental limit of the problem unaffected by and not favouring any specific implementation choice, and it cannot be surpassed by any realizable system. It is here that network information

theory has proven to be the tool of choice when theoretical analysis of the performance limits of a network of interconnected nodes is concerned. Even though many of the canonical problems of network information theory have not yet been solved in their full generality, an information-theoretic analysis of a communication problem can often shed light on the fundamental trade-offs that are involved and the nature and magnitude of improvements one can expect.

Most information theoretical treatments of cognitive radio and cooperative communications, hereafter shortened to cognitive radio for brevity, have divided their setup into three broad families: overlay, underlay and interweave [33]. The difference between these classes of cognitive models relates to how the primary user and secondary users interact and is briefly reviewed here.

In the overlay setup, the cognitive users in the network have access to not only the channel parameters that they have estimated from the RF environment, but also to the codebooks and messages of the primary users as well and as such can actively aid the primary users by using a portion of their power to relaying their signal. They can also help their own intended receivers by treating the known message of the primary as a channel state parameter and using techniques like dirty paper coding to mitigate the effect of interference from the primary on their intended receiver. Using this knowledge, they can basically use the channel in a completely unobtrusive manner by compensating for any signal to noise ratio degradation they have caused on the primary's link with an equal amount of signal to noise ratio improvement through proper division of their power between relaying for the primary user and sending to the secondary receiver. The idealized assumption of access to the messages and codebooks of the primary user is justified in practical terms based on the fact that many higher-level network protocols use relatively static-in-time codebooks and modulation schemes in their physical layer and that the built-in automatic repeat request[4] mechanism for retransmission and acknowledgment of messages means that if the cognitive

[4] ARQ

transmitter happens to be strategically located in a way that it gets better reception of primary's signals than the primary receiver, it will have access to primary's message just in time to help it relay the message to the secondary at the same time as using part of the band for its own message transmission. The authors in [34] investigate a model where this kind of non-causal message information from the primary transmitter is available to the secondary transmitter. They derive an achievable region for this channel based on time-sharing between selfish dirty paper coding of secondary user and relaying the first user's message with part of the secondary transmitter power. They also try to relax the non-causal message knowledge condition and consider the case where the secondary has to wait a fraction of the codeword length of the primary before it gains access to the message and then begin cooperating with primary. In [35], a similar model is studied whereby the extra so called coexistence conditions which prohibit the secondary from changing the effective signal-to-noise ratio of the primary receiver and the receiver not tailoring its coding and decoding scheme to the presence of the secondary are added. With these additional constraints, the authors derive the capacity region of their model and prove that a combination of relaying the primary's message and compensating for its known interference at the secondary receiver is optimal. Their assumptions limit their result to case where interference is weak, that is the channel coefficient between each transmitter and its corresponding receiver is higher than the channel coefficient between that transmitter and the other receiver. The same observations were made in [36] for weak interference that considered this problem as a special case of the problem of interference channels with degraded message sets. The problem of interference channel with degraded message sets was studied in [37, 38] in the case of strong interference.

In the underlay setup, the secondary has information about the channel coefficients of its own and primary's channel, but not its message or codebook, yet it wants to multiplex its own signal into the same band without causing any undue loss of quality of service for the primary. In this situation, what the secondary can do is to tailor its transmit power and direction in a way that it achieves the maximum possible performance for its own transmission at the

same time as keeping the primary's link within the acceptable quality of service envelope. This model has the advantage that it does not rely on the strong assumption of prior non-causal knowledge of primary's message. It is typically much easier to use electromagnetic reciprocity to estimate channel coefficients from one's own transmitter to other network terminals and as for channel coefficients between other pairs of users, carefully listening to the primary users' communications when they are first setting up a transmission session and doing power adjustment and channel learning, it is possible to get a good estimate of their channel coefficients. Also, if the primary users are in geographically fixed locations, as is the case for many legacy users such as television repeaters, a location-aware secondary user can conceivably model the channel quality between other pairs of users.

The interweave setup, which is perhaps more true to the original conception of cognitive radio networks, is where the secondary transmitter opportunistically tries to find holes in the spectral usage of the primary user and fill these holes with its own data. The main challenge in this arrangement is robust and efficient detection of the presence of absence of primary user activity at any given time, frequency and place. This poses a signal processing challenge for the secondary user since its non-obtrusive use of the channel directly depends on how likely it's whitespace-detection procedure is to make false positive and false negative detections of the primary user activity. This is where an information-theoretic analysis may come into play that tries to characterize the limitations of any estimation procedure subject to random disturbances and how, if at all, can cooperation between geographically separate nodes help in resolving false positives and false negatives when detecting primary user activity.

The first problem that we consider has elements of both underlay and interweave setups, since it both opportunistically senses white-spaces in primary user's band usage and at the same time, to achieve higher rates, underlays part of its signal into the same band as the primary user without unduly affecting its utilization of its bandwidth under use and tries to opportunistically either cancel or treat as noise the interference coming from the primary user without relying on any cooperation from it. A number of works in the literature have

studied similar models which are reviewed here.

The authors in [39] consider a fading interference channel shared between a non-CSI[5] aware primary user and a cognitive secondary user. The primary user is using a constant-power and constant-rate coding approach and the quality of service metric being imposed on the secondary user is on primary user's outage probability[6] for its chosen transmission rate not going above some ϵ. They derive the optimum power allocation strategy for the cases where the secondary user has either a peak or an average power constraint and is trying to maximize either its own ergodic capacity [7] or its outage capacity[8] for some ϵ'. The difference with our model is in the fading channel setup and the quality of service metric used.

In the paper [40], a model is considered in which a number of users are trying to share a number of sub-bands in a multi-carrier communication system. They propose an iterative setup where the users try to update their power allocation over the spectrum by adopting, at each stage, the power allocation strategy that maximizes their rate given the interference that they see at that stage. The receivers act opportunistically and try to use multi-user detection whenever possible to maximize their achievable rates. They propose coding schemes based on joint and separate coding over sub-carriers and solve the maximization problem of each user at each iterative step for these coding schemes. The model in this work is quite similar to our setup except that it is solving a distributed spectrum sharing problem in which all the users are trying to cooperatively converge to a stationary point of their stepwise objective functions and no user is given any quality of service guarantee. In our work there is a hierarchy of priority in access to the channel and the primary user is guaranteed a minimum transmission rate without having to cooperate with the secondary.

In [41], the authors study a fading network of one primary and many secondary cognitive users that are aware of the number of other secondary users. In this work, the quality of

[5]Channel State Information.

[6]The probability over all fading states that the instantaneous maximum achievable rate on the link falls below the user's chosen transmission rate.

[7]the expectation of maximum achievable instantaneous rate under the fading distribution.

[8]The maximum rate that can be achieved in with probability greater than $1-\epsilon$ over all fading states.

service metric for the primary user is the attainability of a certain fraction ν of the outage capacity of its interference-free link for some outage probability ϵ even in the presence of secondary users. The secondary users have the option of canceling interference from the primary user or treating it as noise but treat interference from their peers as noise. A number of approaches are considered in this work. These include the secondary users having a so-called activity factor which denotes the probability of them turning on their transmitters in each time slot and of which they try to maximize the expected value, also the strategy of users continuously modulating their transmit power up and down so as to maximize their expected transmission rate is used. They also offer a combination of these methods and compare the performance of all three approaches numerically. The difference with our model again is in the fading model and the quality of service metric adopted.

2.2 Wireless link scheduling

In this section we give a background of relevant work on the problem wireless link scheduling, which is the subject of the second part of thesis.

As described in the introduction, typically, wireless scheduling is approached through declaring conflicts between pairs of links that are in some sense "too close" to transmit simultaneously. Concretely, this approach maps the problem to a graph-based one where links form the vertices of the graph and there is an edge between every pair of vertices if the corresponding pair of links are not able to be active simultaneously. This is called the conflict graph of the link-set. The problem of link scheduling in this setting reduces to a colouring of this conflict graph. These algorithms are generally named independent-set scheduling algorithms as each monochromatic set of vertices is an independent set of the conflict graph. Wireless networks are dynamic entities where transmission demands are best represented by stochastic arrival processes and tools like queuing theory can be used to characterize the dynamic stability conditions of the network, namely conditions under which queue lengths

remain finite at all nodes. In a ground breaking work, the authors in [42] showed that the dynamics of a network under such a model can be stabilized by any scheduling algorithm that selects an independent set[9] that has the maximum aggregated queue length at each time instant. They called these algorithms maximum weight independent set scheduling algorithms. This work established the connection between the dynamic problem of network stability (congestion-avoidance) and the static problem of finding maximal independent sets in the conflict graph. Subsequent work ([43, 44, 45, 46]) has generalized this dynamic to static framework by adopting more fine-grained criteria than stability such as total utility maximization and delay minimization for both general and special (1-hop, 2-hop or disk graph) conflict graphs and in particular, by establishing ([47]) the connection between minimizing routing delay and graph colouring[10] through relating achievable average delay to the chromatic number[11] of the conflict graph. Unfortunately, as noticed in these works, graph colouring and maximum independent set finding are NP-complete problems in the general case. The hope is that the straightforward mapping of the geometric arrangement of links to connectivity properties of vertices in the graph can be used to ensure that the derived conflict graph belongs to a family that is amenable to more efficient colouring and independent set finding. Even setting aside the issue of algorithmic efficiency, since the actual radio interference is not modeled well by any pairwise representable notion of conflict, namely because of the accumulative property of interference, it is tricky to tune these algorithms to real-world deployments without sacrificing either efficiency or reliability. Devising algorithms that directly tackle the broadcast nature of the medium therefore becomes necessary, which requires looking at the wireless network as large interference channel and trying to adapt techniques that have worked in achieving higher rates in that context to this problem. Getting a more faithful model of the physical channel requires studying the problem of scheduling according

[9] An independent set of graph is a set of vertices no two of which are connected by an edge.

[10] A vertex colouring of a graph is an assignment of colours to its vertices such that each monochromatic set is also an independent set.

[11] Chromatic number of a graph is the minimum number of colours for which a vertex colouring of the graph exists.

to observed SNR at the receiver, namely the received Signal to Interference and Noise Ratio (SINR). In fact, in one of the first works considering scaling laws of large wireless networks [48], the authors had studied a "physical" SINR model in addition to their main guard-disk based "protocol model" and had shown that for large wireless networks throughput scales like $\frac{1}{\sqrt{n}}$ on average with increasing network size in a given fixed area. Later, The authors in [49] considered solving the joint scheduling and power control under SINR constraints for a given instance of the network and conjectured it to require exponential enumeration of active subsets in the general case. They provided a simplex-like basis exchange algorithm to solve this problem and discussed some relaxations. In [50], the authors conjectured the same exponential complexity even under a geometric path loss model, where the network nodes have the extra structure of being distributed in a metric space and having the channel coefficients obeying a path loss formula. They also showed that any scheduling method that only uses local information could be worse than optimum by an order of $\log \Delta$ where Δ is the largest to smallest link length ratio. They therefore focused on the case where links don't vary lengthwise by more than a constant factor. Later, the authors of [51] showed that this problem is indeed NP-complete to solve optimally and obtained an approximation algorithm when nodes are located in the Euclidean plane $\mathbb{R}^2$ that uses the 4-colourability of planar graphs as an ingredient. Briefly, they partition the set of links into different classes based on length such that the link length within any single class vary by at most a factor of two. For any of these classes, they divide the Euclidean plane into square cells (with side lengths related to link length scale of the class) and 4-colour the adjacency graph of this cell decomposition. Their algorithm assigns different time-slots to different colour classes and to links of differing lengths. Later results ([52, 53, 54, 55]) also looked at the complexity of exact and approximate method for this algorithm and its variations for the related problem of one-shot scheduling (selecting a maximal SINR-feasible set). In [56], a review of these works has been given which along more recent work [57] come to the conclusion that large constant factors might hinder the practicality of these family of methods.

Our method is most similar to that used by these works, in particular in that we try to adapt graph colouring to SINR-based scheduling. An important difference is that, following the observation of [50], we adopt a power control scheme based on link lengths which has important theoretical advantages. Also, to make sure that the constant factors do not hinder the practical applicability of our algorithm, we refrain from doing a cell-based decomposition of the plane of nodes as performed in [51] and follow-up works. Instead, we go to great lengths to devise an alternate graphic representation for the set of links that, despite representing a binary vertex connectivity criterion, is close enough to the set-based SINR constraint of transmission feasibility to be gainfully used in producing an efficient link schedule. We also have to show that this graph, while not being a planar or disk graph, is still efficiently colourable. Our approach also requires the use of an elaborate set of techniques to carefully bound interference powers in each slot of the resulting schedule in order to show correctness and asymptotically good approximation factors without sacrificing constant factors. As a result, our method does not suffer from drawbacks pointed to by [57].

On the practical side, the next generation cellular network standards (5G) currently under development call for inclusion of Device-to-Device (D2D) modes of operation ([58, 59]). This is in addition to coordinated multi-point, already part of the standard, that enables distributed processing of signals to and from users near the border of cell coverage areas across network-operator controlled base-stations. The new recommendations, rather also call for a two-tier mode of operation involving UE's[12] communicating directly without any base station involvement. This will alter the design space of feasible signaling methods and network management schemes as the performance of current approaches will be limited by the validity of assumptions they implicitly make. In particular, assumptions about the existence of a hierarchical structure in the network and relative homogeneity of nodes power and performance characteristics for traditional cellular and ad-hoc networks respectively, might prove to be inadequate in dealing with D2D networks. In addition, the advent of Internet of

[12]User Equipment is cellular technology parlance for mobile handsets.

Things (IoT) enabled devices and networks of autonomous vehicles and drones, where direct Machine-To-Machine (M2M) communication without any human involvement is envisioned to be much more common, presents new challenges as it increases the number of nodes deployed in a small local area from tens to hundreds and perhaps even thousands. This puts a strain on the scalability properties of current scheduling algorithms and requires approaches that are more attuned to the nature of the wireless medium and do not catastrophically fail under larger density and number of nodes.

For tackling the scheduling challenges in these large networks, a group of scheduling algorithms has been proposed in the literature that does not target schedule length optimality, but rather try to perform SINR-based link scheduling in a way that achieves reasonably high throughput with low time-complexity. In [60], the FlashLinQ algorithm has been proposed through collaboration between an academic group and an industrial team within Qualcomm. The authors use the multi-tone structure of 802.11 spectrum access and dedicate a certain fraction of tones to control signaling and users contend by showing their interest in transmission using these control tones and continue in rounds until all requisite SNR conditions are met. In each round, links are assigned a priority order that changes pseudo-randomly over different timeslots to respect fairness between the links. The algorithm has two global parameters γ_{tx} and γ_{rx}. Each intended receiver sends pilot tones which allows its corresponding transmitter to estimate the channel by electromagnetic reciprocity. During each round, the highest priority link is set to be active and other links are investigated in the order of priority. They are activated for this round only if they cause less than γ_{tx} Interference to Signal Ratio (ISR) on receive less than γ_{rx} ISR from all, necessarily higher priority, links that have earlier been declared active. Fairness is achieved by pseudo-randomly cycling link priorities.

ITlinQ [61], is another algorithm that uses results from the work of the same authors in [62]. That work shows that treating interference as noise achieves optimum Generalized Degrees of Freedom (GDoF) under certain conditions on the coefficients of the channel. The

main idea is splitting the set of links into subsets for which treating interference as noise is optimal under their condition. Straightforward implementation of this set partitioning is a hard combinatorial problem. A pairwise-testable simplified version of this TIN-optimality condition is therefore used to have a tractable algorithm. Links are given a priority ordering similar to FlashLinQ above, and a link i is only added if $INR_{ij} < \sqrt{SNR_i}$ and $INR_{ji} < \sqrt{SNR_i}$ hold for all links j previously activated. By a pseudo-random cycling of link priorities, fairness can be guaranteed and all links eventually scheduled. Enforcing this condition does not necessarily lead to a short schedule as the constraint of achieving GDoF on active links at each timeslot is too stringent. Nevertheless, they show that if the nodes are generated from a random process in such that a very specific scaling relationship holds between the statistics of the distance from a transmitter to its designated receiver and the statistics of the distance between unrelated receivers-transmitter pairs, their schedules are only logarithmically longer than optimal. They also simulate their algorithm and show it compares favourably with FlashLinQ. We will briefly mention ITLinQ again when we have reviewed the basics of interference channel and GDoF analysis and put the paper [62], on which it is based, into context. In Chapter 4, we show that our algorithm compares favourably with both these algorithms in achieving high throughputs in large networks.

Since our models are based on the two-user and K-user Gaussian interference channel, we next give a brief introduction to the interference channel, its formal definition and to various results derived in the literature for its capacity in special cases, and for bounds on its rate region in more generalized scenarios. We also briefly review the Generalized Degrees of Freedom (GDoF) framework and Interference Alignment (IA) which have aided in the understanding of the limits to interference management in large wireless network. This will set the stage for a discussion of our work in the rest of the thesis.

2.3 Review of results on interference channel

This section tries to give the formal mathematical definitions of the concepts involved in the problem of Gaussian interference channel and a brief review of the relevant results from the literature. Definitions given here are for the most part standard and can be found in any textbook on information theory. Results are stated as theorems but the proofs have been omitted for brevity.

2.3.1 Formal definition of discrete memoryless single user and interference channel and their associated capacity regions

The classical work of Shannon on determining the capacity of single-user channels, introduces the discrete memoryless channel as a probabilistic system specifying the conditional probability $P_{Y|X}(y|x)$ of receiving any letter of the finite output alphabet set $\mathcal{Y}$ given that any letter of the finite input alphabet set $\mathcal{X}$ has been sent. A code of rate R and length n is an encoder-decoder function pair (E, D) such that the transmitter of the channel uses $E : \mathcal{M} \to \mathcal{X}^n$ to map any message from a message set $\mathcal{M}$ of cardinality $\lceil 2^{nR} \rceil$ to the n-fold Cartesian product of $\mathcal{X}$ with itself and to use n transmissions to send them over the channel. The receiver in turn, uses $D : \mathcal{Y}^n \to \mathcal{M}$ to map any received n-sequence of channel output symbols to its estimate of the sent message. The probability of error for such a code assuming an equiprobable distribution over the set of input messages is defined as:

$$\lambda_e^n = \frac{1}{|\mathcal{M}|} \sum_{i=1}^{|\mathcal{M}|} P(D(y^n) \neq m_i | E(m_i) \text{ was sent and } y^n \text{ received}). \tag{2.1}$$

The capacity of the channel, C is then defined as the supremum of the rates R so defined for which there exists a sequence of codes of rate R for all n such that the sequence of error probabilities of these codes converges to zero. In a less formal way, the supremum of the rates for which reliable transmission (with zero asymptotic probability of error) is possible.

The discrete memoryless interference channel, as we saw previously and as shown in

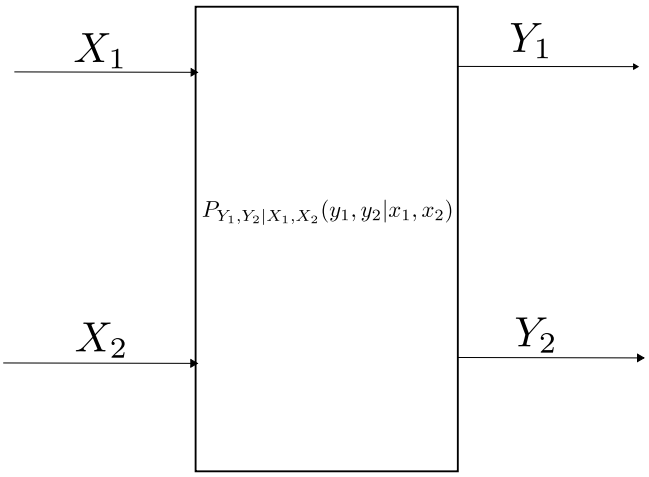

Figure 2.5: Two-user discrete memoryless interference channel

figure 2.5, is characterized by the presence of two pairs of transmitters and receivers using a shared medium to exchange messages. The formal probabilistic set-up of the discrete memoryless interference channel problem is very similar to the single-user channel except that we now have to account for more users and more rates. Formally, an Interference channel is characterized by two pairs of finite input and output alphabets $(\mathcal{X}_1, \mathcal{Y}_1)$ and $(\mathcal{X}_2, \mathcal{Y}_2)$ and the conditional probability specification $P_{Y_1,Y_2|X_1,X_2}(y_1, y_2|x_1, x_2)$ that specifies the probability of receiving any pair of letters from the two output alphabets given that any pair of letters from the two input alphabets are sent. This generally non-factorizable specification is meant to model the cross-channel effects of the two users of the channel having to share the communication resources. A code with rate pair (R_1, R_2) for the discrete memoryless interference channel is a pair of message sets $\mathcal{M}_1, \mathcal{M}_2$ of cardinality $\lceil 2^{nR_1} \rceil$ and $\lceil 2^{nR_2} \rceil$ respectively and two encoder-decoder function pairs (E_1, D_1) and (E_2, D_2) where $E_1 : \mathcal{M}_1 :\to \mathcal{X}_1^n$ and $E_2 : \mathcal{M}_2 :\to \mathcal{X}_2^n$ are the encoders that map the message of each user to a length-n sequence of the letters in the input alphabet of the corresponding user. On the receive side, each receiver uses the corresponding decoder function $D_1 : \mathcal{Y}_1^n \to \mathcal{M}_1$ and $D_2 : \mathcal{Y}_2^n \to \mathcal{M}_2$ that estimate the message from their corresponding transmitter based on

2.3. Review of results on interference channel

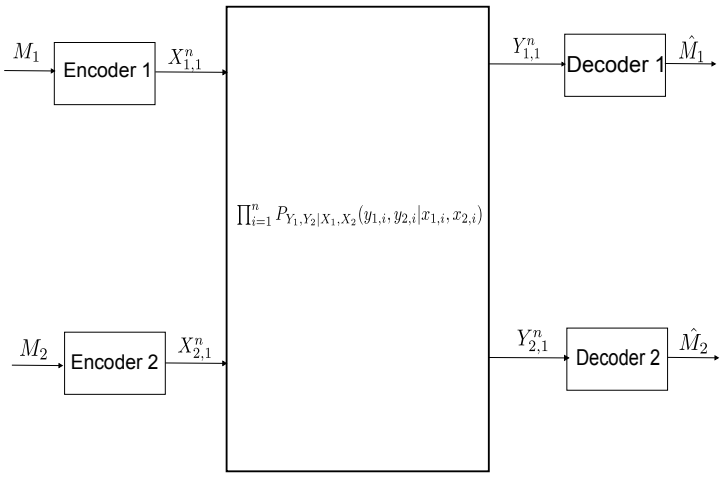

Figure 2.6: Coding and decoding setup for the two-user discrete memoryless interference channel

what they have received on their output for n channel use intervals. Each user's probability of error in such a coding scheme is defined as shown in Figure 2.6 by

$$\lambda_{1,e}^{(n)} = \frac{1}{|\mathcal{M}_1|\times|\mathcal{M}_2|} \sum_{i=1}^{|\mathcal{M}_1|} \sum_{j=1}^{|\mathcal{M}_2|} P(D_1(y_{1,1}^n) \neq m_i | E(m_i) \text{ and } E(m_j) \text{ were sent and } y_1^n, y_2^n \text{ received}) \tag{2.2}$$

$$\lambda_{2,e}^{(n)} = \frac{1}{|\mathcal{M}_1|\times|\mathcal{M}_2|} \sum_{i=1}^{|\mathcal{M}_1|} \sum_{j=1}^{|\mathcal{M}_2|} P(D_2(y_{2,1}^n) \neq m_j | E(m_i) \text{ and } E(m_j) \text{ were sent and } y_1^n, y_2^n \text{ received}) \tag{2.3}$$

$$\lambda_e^{(n)} = \max(\lambda_{1,e}^{(n)}, \lambda_{2,e}^{(n)}). \tag{2.4}$$

A rate pair (R_1, R_2) is said to be achievable for the discrete memoryless interference channel if there exists a sequence of codes of rate (R_1, R_2), one for each n, such that the sequence $\{\lambda_e^{(n)}\}$ converges to zero. The capacity region of the discrete memoryless interference channel is defined as the closure of the set of rate tuples (R_1, R_2) for which there exists a sequence of codes such that $(\mathcal{M}_1, \mathcal{M}_2) = (2^{\lceil nR_1 \rceil}, 2^{\lceil nR_2 \rceil})$ for the n'th code in the sequence

and that the associated sequence $\{\lambda_e^{(n)}\}$ converges to zero.

2.3.2 Formal definition of the Gaussian interference channel

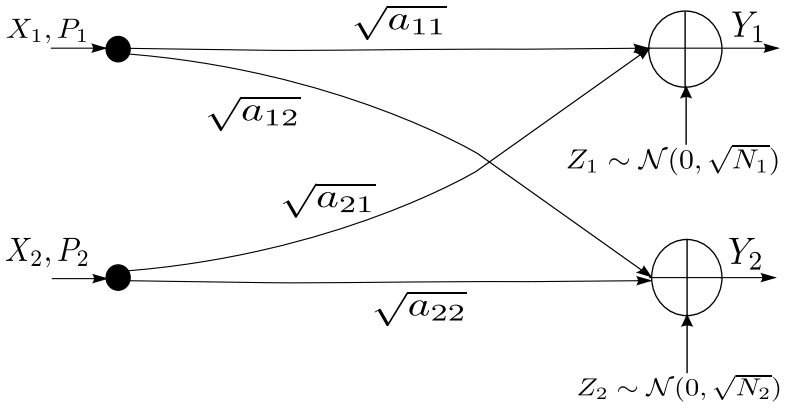

Figure 2.7: Gaussian Interference Channel (GIC)

The discrete-time Gaussian Interference Channel (GIC), shown in figure 2.7, is characterized by real-valued input and output alphabets and the relationship between inputs and outputs is given by

$$Y_1[n] = a_{11}X_1[n] + a_{21}X_2[n] + Z_1[n] \tag{2.5}$$

$$Y_2[n] = a_{12}X_1[n] + a_{22}X_2[n] + Z_2[n], \tag{2.6}$$

where $Z_1[n]$ and $Z_2[n]$ are stationary zero-mean i.i.d Gaussian random processes independent of each other with sample variance of N_1 and N_2 (sometimes termed noise power) respectively and for any code of length n we have

$$\frac{1}{n}\sum_{i=1} n|X_1[i]|^2 \leq P_1 \tag{2.7}$$

$$\frac{1}{n}\sum_{i=1} n|X_2[i]|^2 \leq P_2, \tag{2.8}$$

where P_1 and P_2 are the power constraints of user 1 and 2.

This channel can be normalized so that the direct gains a_{11} and a_{22} and noise powers N_1 and N_2 become equal to 1. This is done by rescaling inputs and outputs as

$$X_1^{'} = \frac{a_{11}}{\sqrt{N_1}} X_1, Y_1^{'} = \frac{1}{\sqrt{N_1}} Y_1, Z_1^{'} = \frac{1}{\sqrt{N_1}} Z_1 \tag{2.9}$$

$$X_2^{'} = \frac{a_{22}}{\sqrt{N_2}} X_2, Y_2^{'} = \frac{1}{\sqrt{N_2}} Y_2, Z_2^{'} = \frac{1}{\sqrt{N_2}} Z_2, \tag{2.10}$$

and changing the power constraints to

$$P_1^{'} = \frac{a_{11}^2}{N_1} P_1, P_2^{'} = \frac{a_{22}^2}{N_2} P_2, \tag{2.11}$$

whereby the cross-channel gains become $\sqrt{a_{12}^{'}}$ and $\sqrt{a_{21}^{'}}$ defined as:

$$a_{12}^{'} = \frac{a_{12}^2 N_1}{a_{22}^2 N_2} \tag{2.12}$$

$$a_{21}^{'} = \frac{a_{21}^2 N_2}{a_{11}^2 N_1}. \tag{2.13}$$

We will assume that all two-user interference channels are normalized from now on and the cross gains are represented in square root notation.

The definition of a coding scheme for the discrete-time Gaussian interference Channel is similar to the discrete memoryless case. A very important difference is that the preceding definitions of discrete memoryless channels concern only the case where the channels is discrete in time and the input and output alphabets are finite. The generalization from this setup to the setup of continuous-alphabet channel where both time and signal amplitudes are continuous requires generalizing all the definitions involved in the problem set-up to the continuous-alphabet case and is basically a continuity argument. In short, in this argument a sequence of ever finer quantizations of the input and output alphabet spaces is considered and it is shown via a sequential completeness argument on the space of such quantizations

that the limit of such increasingly accurate models of the underlying continuous-alphabet channel exists and that indeed the problem of coding for the underlying channel is well-defined. Of note is that for continuous alphabet channels, a restriction of the power of input signals, or equivalently the variance of input random variables is necessary because otherwise the capacity of the channel would be infinite. Most direct and converse coding arguments carry through without change for this continuous alphabet setting by appropriate inclusion of the power constraint and therefore many results for discrete channels have almost analogous counterparts for the continuous-alphabet case. The rigorous details of this argument for extension to continuous alphabets can be found in more mathematically-oriented books on information theory (c.f. [63]).

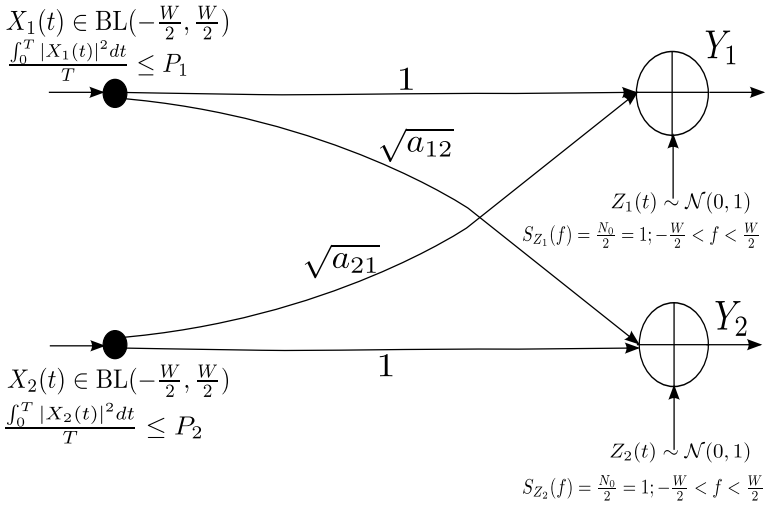

Figure 2.8: Band-limited Gaussian interference channel

The definition of discrete-time Gaussian interference channel sets the stage for the definition of continuous-time Gaussian interference channel that is shown in figure 2.8. in this channel, the signals and noises involved are in general continuous random processes in time and are limited to be in the class of band-limited random processes of bandwidth W (referred to as $\mathrm{BL}(-\frac{W}{2}, \frac{W}{2})$ in the figure). The spectral density of the noise, denoted by $S_z(f)$ is assumed to be flat and normalized to 1 over the bandwidth. The same power constraints

as the discrete-time case apply here with the suitable redefinition of signal powers. Again, using Shannon-Nyquist sampling theorem in this case, the problem of continuous-time channels can be reduced to a discrete-time continuous vector analogue where the components of the vectors are given by the $2W$ samples per second required per Shannon sampling theory. The power constraint now becomes a trace constraint on the covariance matrix of the input random vectors of this equivalent channel and, for the case of AWGN[13] channels, using Hadamard inequality for the maximum determinant of a matrix with a trace condition it is shown that this vector channel is equivalent to $2WT$ separate discrete-time channels over which the total energy constraint of the signal is equally divided. This argument is also standard (c.f. [64, 65]) and will not concern us anymore here.

2.3.3 Capacity region of the Gaussian interference channel

In general, the problem of calculating the capacity of the Gaussian interference channel is still open. There are a few special cases where the capacity has been found which we will review in this section. The next section is concerned with the other cases and reviews the major inner bounds (achievability schemes) that have been discussed in the literature.

Channel without interference

If both cross-gains are zero in an interference channel it is easy to see how the channel can be decomposed to two parallel point-to-point AWGN channels for each of which the capacity is known.

Very strong interference

This is the first major result on interference channels and seems counterintuitive since it shows that having very strong interference is as good as having no interference at all. The intuitive reason for this is that interference, unlike noise, is structured and hence if it is

[13] Additive White Gaussian Noise

strong enough it can be decoded and compensated for. This result is from Carleial [66] and his coding theorems state:

Theorem 2.1. *In a discrete memoryless channel for which both of the inequalities below hold for any separable joint probability distribution on channel inputs, $p(x_1, x_2) \sim p(x_1)p(x_2)$,*

$$I(X_1; Y_1|X_2) \leq I(X_1; Y_2) \tag{2.14}$$

$$I(X_2; Y_2|X_1) \leq I(X_2; Y_1), \tag{2.15}$$

the following rate region is achievable

$$R_1 \leq I(X_1; Y_1|X_2) \tag{2.16}$$

$$R_2 \leq I(X_2; Y_2|X_1). \tag{2.17}$$

For the Gaussian continuous-alphabet discrete-time channel, the following equivalent theorem holds.

Theorem 2.2. *If the power constraints and cross coefficients of a Gaussian interference channel satisfy*

$$1 + P_1 < a_{2,1} \tag{2.18}$$

$$1 + P_2 < a_{1,2}, \tag{2.19}$$

the capacity region of the channel is attained for Gaussian input distributions and is given by

$$R_1 \leq \frac{1}{2}\log(1 + P_1) \tag{2.20}$$

$$R_2 \leq \frac{1}{2}\log(1 + P_2). \tag{2.21}$$

Intuitively, in the very strong interference case, the cross-receiver's reception of the signal,

even considering interference from its own transmitter is better than the intended receiver. This means that under any feasible rate pair, the cross receivers can completely eliminate the effect of crosstalk and decode their intended signals as if there was no interference.

Strong interference

The case of very strong interference that we saw previously can be considered a special case of strong interference, to be defined here. In this case, the cross paths have a better reception than the direct path, but only if they can eliminate the effect of their intended signal. In this case, the interference channel can be decomposed into two multiple-access channels and the capacity region is shown to be the intersection of these capacity regions. In this case we have the following coding theorem which is due to Costa and El-Gamal ([67]):

Theorem 2.3. *In a discrete memoryless channel, if, for any separable probability distribution on channel inputs $p(x_1, x_2) \sim p(x_1)p(x_2)$, the inequalities*

$$I(X_1; Y_1|X_2) \leq I(X_1; Y_2|X_2) \tag{2.22}$$

$$I(X_2; Y_2|X_1) \leq I(X_2; Y_1|X_1), \tag{2.23}$$

hold, then the capacity region of the channel is given by

$$R_1 \leq I(X_1; Y_1|X_2, Q) \tag{2.24}$$

$$R_2 \leq I(X_2; Y_2|X_1, Q) \tag{2.25}$$

$$R_1 + R_2 \leq \min\{I(X_1, X_2; Y_1|Q), I(X_1, X_2; Y_1|Q)\}, \tag{2.26}$$

for every time sharing random variable Q such that the probability distribution $P_{QX_1X_2}$ factorizes as $p(q)p(x_1|q)p(x_2|q)$.

The conditions of this theorem for the Gaussian case imply that

$$1 < a_{2,1} \tag{2.27}$$

$$1 < a_{1,2}, \tag{2.28}$$

and the capacity region reduces to

$$R_1 \leq \frac{1}{2}\log_2(1+\frac{P_1}{N_1}) \tag{2.29}$$

$$R_2 \leq \frac{1}{2}\log_2(1+\frac{P_2}{N_2}) \tag{2.30}$$

$$R_1 + R_2 \leq \min\{\frac{1}{2}\log_2(1+\frac{P_1+a_{21}P_2}{N_1}), \frac{1}{2}\log_2(1+\frac{P_2+a_{12}P_2}{N_2})\}. \tag{2.31}$$

Except for these two cases, the capacity of the interference channel in mixed and weak interference regime is only known for special cases and restricted ranges of parameters. This has motivated the introduction of inner bounds that characterize the achievable rates and outer bounds that characterize the maximal rates that could possibly, but not necessarily be achieved. We will review these bounds in the upcoming sections.

2.3.4 Achievability schemes for the interference channel

Han-Kobayashi Method

The intuitive idea of the Han-Kobayashi scheme [68] is for each transmitter to divide its message into a public and a private part and use superposition coding to send them together. The public parts of the messages are decoded by both receivers, but each user only decodes its own private part and treats private part of the message of the other user as noise.

Formally, for every distribution $P_{QU_1V_1U_2V_2X_1X_2Y_1Y_2}$ that can be factorized as

$$P_Q P_{U_1|Q} P_{V_1|Q} P_{X_1|U_1,V_1} P_{U_2|Q} P_{V_2|Q} P_{X_2|U_2,V_2} P_{Y_1,Y_2|X_1,X_2}, \tag{2.32}$$

where the $P_{Y_1,Y_2|X_1,X_2}$ term denotes the transition probabilities that characterize the channel and U_i and V_i, $i \in \{1,2\}$, are auxiliary random variables representing the public and private parts of the message respectively. Then, any rate pair (R_1, R_2) such that R_1 and R_2 can be decomposed as $R_{1v} + R_{1u}$ and $R_{2v} + R_{2u}$ respectively that satisfy the following conditions, both in their original form and with indices 1 and 2 swapped, is achievable. The conditions are:

$$R_{1v} \leq I(V_1; Y_1|U_1, U_2, Q) \tag{2.33}$$

$$R_{1u} \leq I(U_1; Y_1|V_1, U_2, Q) \tag{2.34}$$

$$R_{2u} \leq I(U_2; Y_1|V_1, U_1, Q) \tag{2.35}$$

$$R_{1v} + R_{1u} \leq I(V_1, U_1; Y_1|U_2, Q) \tag{2.36}$$

$$R1v + R_{2u} \leq I(V_1, U_2; Y_1|U_1, Q) \tag{2.37}$$

$$R_{1u} + R_{2u} \leq I(U_1, U_2; Y_1|V_1, Q) \tag{2.38}$$

$$R_{1v} + R_{1u} + R_{2u} \leq I(V_1, U_1, U_2; Y_1|Q). \tag{2.39}$$

Here, R_{iv} and R_{iu}, $i \in \{1,2\}$ denote the private and public message rates respectively. A much simpler set of inequalities describing this rate region that was provided by Chong et. al in [69] is that for any $P_{QU_1X_1U_2X_2Y_1Y_2}$ that can be factorized as $P_Q P_{U_1|Q} P_{U_2|Q} P_{X_1|U_1} P_{X_2|U_2} P_{Y_1,Y_2|X_1,X_2}$,

the rates

$$R_1 \leq I(X_1; Y_1|U_2, Q) \tag{2.40}$$

$$R_1 + R_2 \leq I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|Q) \tag{2.41}$$

$$2R_1 + R_2 \leq I(X_1; Y_1|U_1, U_2, Q) + I(X_1, U_2; Y_1|Q) + I(X_2, U_1; Y_2|U_2, Q) \tag{2.42}$$

$$R_2 \leq I(X_2; Y_2|U_1, Q) \tag{2.43}$$

$$R_1 + R_2 \leq I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|Q) \tag{2.44}$$

$$R_1 + 2R_2 \leq I(X_2; Y_2|U_1, U_2, Q) + I(X_2, U_1; Y_2|Q) + I(X_1, U_2; Y_1|U_1, Q) \tag{2.45}$$

$$R_1 + R_2 \leq I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q), \tag{2.46}$$

are achievable. As can be seen, this formulation removes the need for splitting rates into their public and private parts and also the auxiliary random variables V_1 and V_2 are not needed.

The Han-Kobayashi rate region is to date the most comprehensive achievability scheme proposed for the interference channel, as it subsumes many other achievability schemes as special cases. But despite conceptual simplicity, calculating and optimizing over the auxiliary variables to arrive at the entire achievable rate region is a notoriously difficult and non-convex optimization problem.

2.3.5 Generalized Degrees of Freedom (GDoF)

This section gives an overview of the degrees of freedom and generalized degrees of freedom notions and their role in clarifying the effects of moderate and weak interference on interference channel capacity. An attempt at deriving limiting expressions for capacity region is another alternative to inner and outer bounds that is used to tackle the complexity of exact characterization of capacity for interference channels and can aid understanding the role of structured interference as opposed to structureless noise. One way to accomplish this is by

looking at the high-SNR behaviour of channel capacity since it deemphasizes the relative importance of noise. This framework is known as the General Degrees of Freedom (GDoF) analysis, and is a generalization of the notion of Degrees of Freedom (DoF). To understand the terminology, we have to take a look at multiple-antenna channels. It is a well-known result for $N \times N$ multiple-input-multiple-output(MIMO) point-to-point AWGN channels with non-singular coefficient matrices, that the channel can be decomposed into N uncorrelated spatial components corresponding to the left and right eigenvectors of the channel matrix. This means that as the SNR goes to infinity, the asymptotically fastest growing term in channel capacity is $\frac{1}{2}N\log(1+SNR)$. In other words:

$$\lim_{SNR\to\infty} \frac{C_{MIMO-N\times N}(SNR)}{\log(1+SNR)} = N. \tag{2.47}$$

This scaling by N compared to SISO[14] capacity is known as the multiple-antenna multiplexing gain as it is essentially another way of stating the existence of N independent beamforming directions. Having N antennas at the transmitter and receiver is therefore said to give N spatial degrees of freedom.

Adopting a similar notion is more complicated for the interference and other multiuser channels as users are strategically competing for their rates and there is a capacity region rather than a single capacity. Despite the competitive nature of multiuser channels, the same insight of looking at the multiple of SISO point-to-point capacity at high-SNR proves useful when analyzing the minimax capacity of the symmetric interference channel (where cross and direct channel coefficients are equal for both users), since in this case, the rate region is symmetric by symmetry of the users. The generalized degree of freedom of a symmetric two-user interference channel is therefore defined as

$$d(\alpha) := \lim_{SNR\to\infty} \frac{\max_{(R_1,R_2)\in\mathcal{C}} \min\{R_1, R_2\}}{C_{awgn}}. \tag{2.48}$$

[14]Single-input-single-output

Here, C_{awgn} is the point to point capacity of the equivalent SISO channel. In this formulation, DoF is characterized by an interference strength parameter α that defines the INR-SNR scaling at high SNR's, namely $\alpha := \frac{\log INR}{\log SNR}$, where $\alpha > 1$ corresponds to strong and very strong interference regimes and $\alpha < 1$ represents weak interference. Generalized Degrees of Freedom (GDoF) refers to this curve of $d(\alpha)$ showing the change in DoF of the channel as a function of interference level. The work of Etkin, Tse and Wang [70] was the first to define GDoF and obtain the now famous "W" shape of $d(\alpha)$ curve. Their work showed that for values of $\alpha \in [0, 1/2]$ treating interference as noise is DoF-optimal. This was an observation that was also made almost simultaneously by [71, 72, 73]. These results went against the conventional view from the strong and very strong interference case that information-bearing interference signals should not be discarded as noise. This was a crucial observation and one that played a role in the application of interference alignment to interference channel which we review in the next section.

2.4 Interference alignment

As we discussed in the previous section, the GDoF framework led to the observation that for some operating regimes, not decoding interference might not be suboptimal. At the same time a generalization of this analysis to K-user channels was done in [74]. For MIMO and time-varying channels, where many independent dimensions are available, attention was turned towards encoding schemes that try to concentrate interference terms and desired signal into disjoint subspaces. This is the essential idea of interference alignment and was first used for capacity characterization in multiple-antenna X-channels (where each transmitter has messages for each receiver) in [75]. In [11], Cadambe and Jafar show that fully connected K-user interference channel with time-varying or frequency-selective coefficients can effectively have $1/2$ degrees of freedom per user for a total of $K/2$ as opposed to the $1/K$ previously thought for a total of 1. This was a major result that has led a large number

of follow-up publications investigating the applicability of interference alignment to other channel models and also finding other methods of performing interference alignment. These results include ergodic interference alignment [76], that can be applied to any channel at finite SNR where coefficients are time varying and ergodic-in-time with a symmetric around zero probability distribution, and real interference alignment [77] that uses the vector space dimensions of real numbers over rationals and their number-theoretic properties to provide dimensions over which to align signal and interference. The book [78] and survey article [79] offer a more complete exposition of the main ideas behind and variations of interference alignment. One common trait of these different variations is their perceived lack of practical relevance. Some of this perception stems from the GDoF framework itself, where it has been shown that the GDoF is an almost-everywhere discontinuous function of the channel coefficients [80] and that the GDoF of interference channels collapses when channel state information has finite precision [81], but also from the delay-rate trade-off of interference alignment [82, 83], to the required feedback and synchronization overhead and finite-SNR sub-optimality that results from ill-conditioned or incompletely specified channel matrices [8].

2.5 Optimality of Treating Interference as Noise (TIN)

In an even more promising direction, recently the authors in [62] have shown that for a large class of interference networks, treating interference as noise is GDoF optimal. This is done by removing the power control variables from the formulation of the GDoF region. Using the potential theory of network flow problems, they show that a polyhedral inner bound is obtained by a certain relaxation of this problem and that it matches an outer bound based on repeated application of the cyclic interference channel outer bounds of [84]. A more refined analysis then shows that the total gap to sum capacity is in fact constant. This work has been extended by [85, 86] both to gain insights into the relationship between combinatorial

problems and optimality of TIN and to extend it to other channel models. As mentioned, this surprising result has already been used in [61] as the basis of a link scheduling algorithm called ITLinQ which uses time sharing among subsets of links that satisfy the criteria of [62] to service all links in a network.

The next two chapters concern the problems addressed in this thesis. Chapter 3 analyzes a two-user cognitive channel and derives rate regions, decoder designs and rate optimization strategies of the cognitive user. In Chapter 4, we study a large network link scheduling problem and provide and analyze a scheduling method that has provable performance characteristics with respect to the globally optimum schedule.

Chapter 3

Two-user cognitive GIC without non-causal message information

In this chapter, we set up the model of Gaussian Interference Channel (GIC) with one cognitive user that is the subject of our analysis. We then make the key observation that since the legacy user cannot be realistically assumed to cooperate with the secondary user, a number of techniques including time-sharing will not be available and therefore, the achievable rate regions will be in general non-convex. This leads to a complicated rate optimization problems for the secondary (cognitive) user. Several rate-equivalent coding schemes are proposed for this setup. The secondary rate-optimization problem is analyzed and its behaviour under various regimes on parameter values classified. Part of the work in this chapter has been previously presented in the paper [87].

The model that we have proposed here has elements of both underlay and interweave setups. The secondary user both opportunistically senses whitespace in primary user's band and at the same time underlays part of its signal into the portion of the spectrum already under use by this user. The secondary user must guarantee that the quality of service of the primary does not fall below a certain threshold. The secondary receiver has the option to apply joint decoding or treat interference as noise without relying on any cooperation from

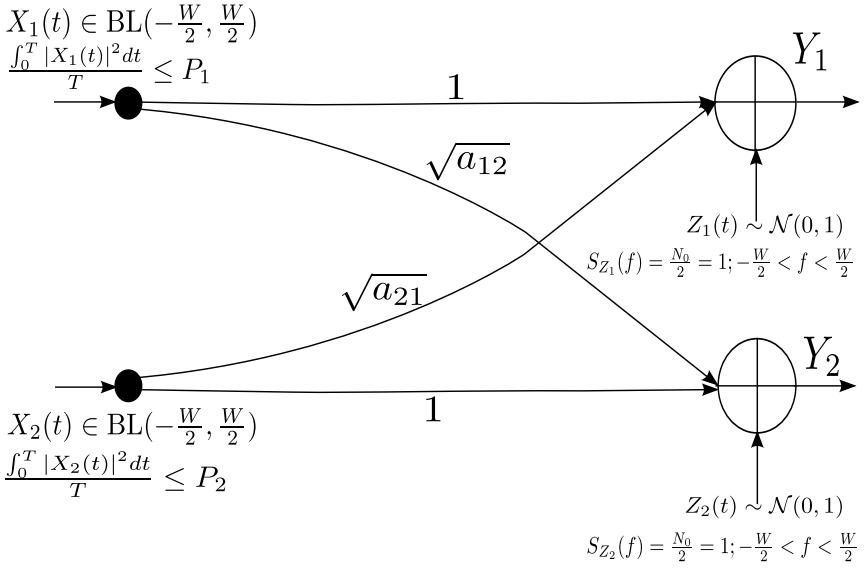

Figure 3.1: Band-limited Gaussian interference channel, $BL(W_1, W_2)$ denotes the class of signals limited to the (W_1, W_2) band and $S_z(f)$ is the spectral density of noise

the primary user. The problem that we solve on this model is determining the bandwidth usage and power division strategy of the secondary user so that its rate is maximized subject to the constraint on the rate of primary user.

3.1 The model and problem

In this section, the channel model and the problem we are trying to solved are described.

3.1.1 The channel model

The network topology in our model is the same as a Gaussian interference channel (shown in Fig. 3.1, see also [65] and references therein). The total bandwidth is W and the static and non-frequency selective channel coefficient from transmitter i to receiver j is denoted by $\sqrt{a_{ij}}$. We assume the usual normalization convention discussed in Chapter 2, so we have $a_{11} = a_{22} = 1$ without loss of generality. The noise is white, which means it has a flat spectral

density which is assumed equal to 1 by the same normalization convention. Transmitter i is limited to a total power of P_i. User 1 is the primary (non-cognitive) user which has the license to use the entire band. User 1 is unaware of the presence of user 2 which is the cognitive or secondary user. Moreover, there is no cooperation among the users. Based on its rate demand, user 1 only occupies a portion W_0 of the whole bandwidth W.

3.1.2 The problem statement

With the channel model as above, user 2 is seeking to multiplex its data on the band W. It occupies the whitespace $W_2 = W - W_0$ and to increase its transmission rate, it also underlays part of its power over a portion W_c of the band W_0 which is in use by user 1. This underlaying is constrained by the quality of service condition $R_1 > R_{\text{th}}$ where R_i is the transmission rate of user i and R_{th} is a fixed threshold on R_1. The average transmission power of user 2 over the bands W_c and W_2 is Q and $P_2 - Q$, respectively. Now the problem that user 2 is trying to solve is that of maximizing its achievable rate without violating user 1's rate constraint. The optimization problem is formulated as

$$\sup_{W_c, Q: W_c \leq W_0, Q \leq P_2} \quad \sup_{(R_1, R_2) \in \mathcal{R}, R_1 > R_{th}} R_2, \tag{3.1}$$

where sup denotes the supremum (least upper bound) operation. $\mathcal{R}$ denotes the achievable rate region for the two users where user 1 is using a legacy single-user encoder/decoder. In other words, we are seeking a characterization of $\mathcal{R}$ and the maximum transmission rate that user 2 can achieve when user 1 is achieving at least a rate R_{th} using simple single-layer encoder/decoder architecture.

3.1.3 Example of a practical application

This problem is intended to model the scenario where a primary user with legacy equipment, such as a VHF over-the-air TV broadcaster with licenses to a large number of analog VHF

channels, wants to coexist with an agile user that tries to use the bandwidth whitespace for its own data transmission. Since the broadcaster may be in the process of phasing out analog TV and some TV channels do not have constantly running programming, at any given instance only a fraction of the assigned bandwidth is carrying data. To maximally assist the secondary user in utilizing bandwidth the primary can announce its minimum requisite quality of service which is modeled here by a rate constraint. To incentivize the primary user to take part in sharing its bandwidth, the goal is to not impose any extra capital expenditure requirements on it. Therefore, it is not required to upgrade its analog and near end-of-life infrastructure to add the capability to dynamically announce which parts of the spectrum are in use, for example by simultaneous out-of-band signaling. The burden of whitespace detection, therefore falls on the cognitive user which is better-equipped to do it. The problem facing the secondary user, after detecting the unused band, will be to divide its transmission power over the empty and occupied parts of band in a way that maximizes its own rate without violating QoS specification of the primary user. The primary user is not required to change its equipment and transmission scheme, and is therefore has a higher incentive to accept new uses for its band.

3.2 The achievable rate region $\mathcal{R}$

Let both users utilize Gaussian codebooks. User 2 does not split its transmission rate by separate coding over the private and common parts of the spectrum, i.e., the signals transmitted over W_2 and W_c belong to a codeword in a codebook of rate R_2. At the receiver side, user 2 has the option to apply simultaneous decoding or treat interference as Gaussian noise, while user 1 only treats interference as Gaussian noise. Let us define

$$r_1 := (W_0 - W_c)\mathsf{C}\left(\frac{a_{1,2}P_1}{W_0}\right) + W_c\mathsf{C}\left(\frac{W_c a_{1,2}P_1}{W_0(W_c+Q)}\right), \tag{3.2}$$

3.2. The achievable rate region $\mathcal{R}$

$$s_1 := W_0 \mathsf{C}\left(\frac{a_{1,2}P_1}{W_0}\right), \tag{3.3}$$

$$r_2 := (W - W_0)\mathsf{C}\left(\frac{P_2 - Q}{W - W_0}\right) + W_c \mathsf{C}\left(\frac{W_0 Q}{W_c(W_0 + a_{1,2}P_1)}\right), \tag{3.4}$$

$$s_2 := (W - W_0)\mathsf{C}\left(\frac{P_2 - Q}{W - W_0}\right) + W_c \mathsf{C}\left(\frac{Q}{W_c}\right) \tag{3.5}$$

and

$$t := (W_0 - W_c)\mathsf{C}\left(\frac{P_1}{W_0}\right) + W_c \mathsf{C}\left(\frac{W_c P_1}{W_0(W_c + a_{2,1}Q)}\right), \tag{3.6}$$

where $\mathsf{C}(x) := \log(1 + x)$. Then the rate region $\mathcal{R}$ is given by

$$\begin{aligned} \mathcal{R} \;:=\; & \{(R_1, R_2) : 0 < R_1 < t\} \\ & \bigcap \Big(\{(R_1, R_2) : 0 < R_1 < s_1, 0 < R_2 < s_2, R_1 + R_2 < s_1 + r_2\} \\ & \qquad \bigcup \{(R_1, R_2) : 0 < R_2 < r_2\} \Big). \end{aligned} \tag{3.7}$$

Fig. 3.2 shows the outline of $\mathcal{R}$ which is non-convex and resembles a chimney.

As an alternative coding scheme, user 2 can employ separate Gaussian codebooks over the private and common parts of the spectrum, i.e., user 2 splits its rate R_2 among two independent codebooks transmitted over W_2 and W_c separately. As before, user 2 has the option to apply simultaneous decoding or treat interference as noise over the band W_c. Denoting the achievable rate region under this alternative coding scheme by $\mathcal{R}^{RS}$, where RS stands for rate-splitting, the following claim shows that $\mathcal{R}^{RS}$ is the same as $\mathcal{R}$ given in Equation 3.7.

Claim 3.1. *Rate splitting gives the same region as joint coding, or* $\mathcal{R}^{RS} = \mathcal{R}$.

Proof can be found in Appendix A.

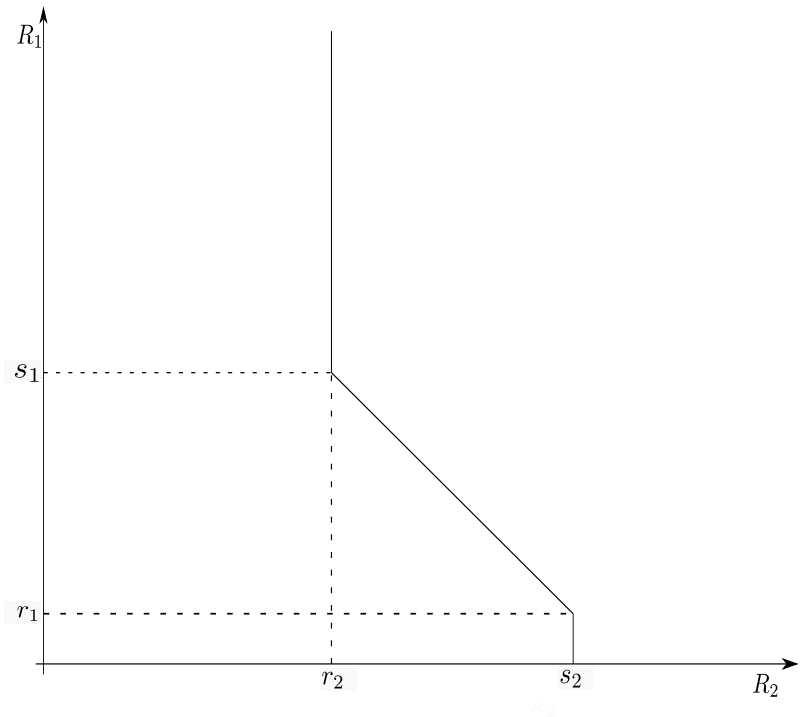

Figure 3.2: The chimney rate region $\mathcal{R}$ described in 3.7.

3.3 Remarks on encoder and decoder structure

In this section we describe two alternate achievable schemes that do not enlarge the rate region calculated in the previous section, but are of independent interest as they provide different encoder and decoder structures that achieve different trade-offs between simplicity of the decoding rule and memory requirements of their implementation.

3.3.1 Non-unique joint typicality decoding

In the previous schemes discussed, user 2 has to adaptively choose between treating interference as noise and simultaneous decoding. The authors of [69] and [88] have independently proposed an alternative decoder structure which is applicable to our scenario as well. To describe this decoder, we have to describe joint typicality and its use in information-theoretic

decoders. A deterministic n-sequence is called typical for probability distribution P_X of random variable X if the probability of it being observed in n independent draws of X is between $2^{-(nH(X)-\epsilon)}$ and $2^{-(nH(X)+\epsilon)}$, for a small positive ϵ, where $H(X)$ is the entropy of X. Intuitively, this means that the empirical frequency distribution of letters in the sequence is very close to P_X, so the the sequence is typical of what is seen in draws from X. This definition straightforwardly extends to joint typicality for two or more sequences with respect to a joint probability distribution, say $P_{X,Y}$, and is called joint typicality. This means that the empirical joint frequency distribution of the letters from these sequences is close to that joint distribution. Almost all achievability results in information theory rely on joint typicality decoding which means that the decoder tests all codewords for joint typicality with the observed channel output with respect to the input-output joint probability distribution. It declares success if and only if there is exactly one codeword jointly typical with the output and declares the index of that codeword as the message sent. When jointly decoding two codewords from a single channel output, the joint typicality of three sequences (two codewords and one channel output) has to be considered and a decoding error is raised even if the failure of unique joint typicality occurs for a message that is not intended for our receiver. The decoder proposed by [69] and [88], relaxes this by implementing the rule of joint typicality decoding with non-unique decoding for undesired message codewords[1]. It turns out that using this so-called indirect decoder in our setup will achieve the same rate region as derived before.

Claim 3.2. *Replacing the decoder of Section 3.2 with a non-unique joint typicality decoder will result in the same achievable region*

Proof can be found in Appendix A.

So, as we saw, replacing the rather complex decoder structure of the previous section achieves the same rates with considerably more elegance and operational simplicity, though

[1]This means that an error is not declared even if more than one codeword is jointly typical with the channel output if they all agree on the desired message index.

at the expense of higher space (memory) complexity at the decoder. It is operationally simpler because it achieves the same rate region with a unified one-step joint-typicality decoder. This is in contrast with the decoder structure of the previous section that involves implementing two joint-typicality decoders and selecting the output of the one giving the best rate. This simplicity comes at a price though, since in a simple lookup table implementation, this single joint-typicality joint decoder requires one lookup table of size $2^{R_1+R_2}$, rather than two successive lookup tables of size 2^{R_1} and 2^{R_2} respectively[2].

3.3.2 Multi-layer encoding and successive cancellation

It is well-known that the corner points of the capacity region of a Multiple-access channel can be achieved without resorting to simultaneous decoding. This is done by successive interference cancellation (first used for decoding superposition codes in the paper by Cover that introduced the broadcast channel [18]), where the message of one user is first decoded and its effect subtracted from the channel output before decoding the message of the other user. Normally, all other points on the boundary of the capacity region will be achieved by time-sharing between these corner points. Unfortunately our setup does not allow for time-sharing. Thus, it may seem that simultaneous decoding is inevitable in our scheme for achieving points other than the corner points. This is not true however, since as it turns out, with appropriate rate-splitting at the transmitter of user 2, successive cancellation of interference can achieve all the boundary points. Our construction follows that of [89], the details of which follow.

In the rate region achievable at receiver 2 in the previously discussed schemes, for the points on the line $R_1 + R_2 = s_1 + r_2$ that are not corner points, it is implicitly assumed in most practical designs that time-sharing between corner points can be used to achieve these points using a simple interference cancellation scheme instead of the more computationally demanding joint decoding at receiver 2. We recall that the specific setup of our problem

[2]This distinction might not be practically relevant since in both cases exponentially large lookup tables are required.

that precludes the cooperation by the primary, explicitly rules out this kind of time-sharing cooperation and that this is indeed the reason that our rate region is non-convex. But, while lack of cooperation leads to non-convexity of the achievable rate region it does not mean joint decoding is the only way to achieve the non-corner points on the boundary of the achievable region. Rimoldi et al. in [89] show that any point on the so-called dominant face of the capacity region of a multiple-access channel (which in the case of the two-user channel is exactly the set of rate pairs which attain the maximum sum-rate) can be achieved by a computationally simple successive interference cancellation scheme. If we define the function $F(p, n)$ as

$$F(P, n) := \log\left(1 + \frac{P}{n}\right), \tag{3.8}$$

which gives the twice the capacity of a discrete-time AWGN[3] channel with power constraint P and noise variance n, the function F satisfies

$$F(P_1, n + P_2) + F(P_2, n) = F(P_1 + P_2, n). \tag{3.9}$$

A consequence of this identity is that , for a point-to-point Gaussian channel, any division of rate by dividing the total power $P_1 + P_2$ between two superimposed codebooks with powers P_1, P_2, when combined with a successive decoding scheme does not incur any rate penalty or advantage. Likewise, in our own problem, if we focus on the two-user MAC seen by user 2 over the band W_c, which has power constraints $\frac{W_c}{W_0}P_1$ and Q and gains $\sqrt{a_{1,2}}$ and 1 for users 1 and 2 respectively, then for any rate pair (R_1, R_2) on the sum-rate boundary $R_1 + R_2 = W_c F\left(a_{1,2}\frac{W_c}{W_0}P_1 + Q, W_c\right)$, user 2 can divide its power into Q_1 and $Q_2 = Q - Q_1$ where Q_1 is the unique non-negative number smaller than Q such that $R_1 = W_c F\left(a_{1,2}\frac{W_c}{W_0}P_1, W_c + Q_1\right)$ and generate a codebook $\mathcal{C}_{2,1}$ with rate $R_{2,1} = W_c F(Q_2, W_c)$ with power Q_2 and another codebook $\mathcal{C}_{2,2}$ with rate $R_{2,2} = W_c F\left(Q_2, W_C + a_{1,2}\frac{W_c}{W_0}P_1 + Q_1\right)$

[3]Additive white Gaussian noise

and power Q_2. We have that

$$\begin{aligned} & F(Q_2, W_c) + F\left(a_{1,2}\frac{W_c}{W_0}P_1, W_c + Q_1\right) \\ +F\Big(Q_2\ , & W_c + a_{1,2}\frac{W_c}{W_0}P_1 + Q_1\Big) \\ = F\left(Q_1 + a_{1,2}\frac{W_c}{W_0}P_1 + Q_2, W_c\right) & = F(a_{1,2}\frac{W_c}{W_0}P_1 + Q, W_c). \end{aligned} \tag{3.10}$$

Since the term on the right side of (3.10) is the sum-rate of the channel divided by W_c, we can achieve this arbitrary rate pair by a two-layer encoding on the transmit side. On the receive side, we recover the messages in a three-step process. We first decode $\mathcal{C}_{2,2}$, treating the power contributed by $\mathcal{C}_1$ and $\mathcal{C}_{2,1}$ as noise. After that the message from $\mathcal{C}_1$ can be decoded treating the power contributed by $\mathcal{C}_{2,1}$ as noise. Finally, we decode $\mathcal{C}_{2,1}$ and reconstruct user 2's message (So the decoding order is $\mathcal{C}_{2,2} \to \mathcal{C}_1 \to \mathcal{C}_{2,1}$).

The coding approach referred to in the previous section leads to another result and that is a converse that multilayer random coding will not enlarge the rate region achievable by our scheme.

We begin with a lemma that generalizes Equation 3.9.

Lemma 3.1. *For the AWGN capacity function F defined by Equation 3.8, we have*

$$\sum_{i=1}^{n} F(P_i, n + \sum_{j=1}^{i-1} P_i) = F(\sum_{i=1}^{n} P_i, n). \tag{3.11}$$

Proof is by induction and can be found in Appendix A.

Claim 3.3. *Multilayer random coding at user 2's transmitter cannot enlarge the rate region calculated in Section 3.2.*

Proof can be found in Appendix A.

3.4 Analysis of the rate-optimization problem

In this section, we study the optimization problem in (3.1) and reduce it to a number of subproblems that can be solved using line search methods.

Using the definition of $\mathcal{R}$, it is easy to see that

$$\sup_{(R_1,R_2)\in\mathcal{R}:R_1\geq R_{\text{th}}} R_2 = \begin{cases} s_2 & R_{\text{th}} \leq \min\{s_1, u\} \\ r_2 + s_1 - R_{\text{th}} & \min\{r_1, t\} \leq R_{\text{th}} \leq \min\{s_1, t\} \\ r_2 & \min\{s_1, t\} \leq R_{\text{th}} \leq t. \end{cases} . \tag{3.12}$$

Defining

$$\mathcal{D} := (0, W_0] \times (0, P_2], \tag{3.13}$$

$$v_1 := \sup_{(W_c,Q):R_{\text{th}}\leq\min\{r_1,t\}} s_2, \tag{3.14}$$

$$v_2 := \sup_{(W_c,Q):\min\{r_1,t\}\leq R_{\text{th}}\leq\min\{s_1,t\}} r_2 + s_1 - R_{\text{th}}, \tag{3.15}$$

and

$$v_3 := \sup_{(W_c,Q):\min\{r_1,t\}\leq R_{\text{th}}\leq t} r_2, \tag{3.16}$$

it is evident that

$$\sup_{(W_c,Q)\in\mathcal{D}} \sup_{(R_1,R_2)\in\mathcal{R}:R_1\geq R_{\text{th}}} R_2 = \max\{v_1, v_2, v_3\}. \tag{3.17}$$

Since r_1 and t are decreasing functions of Q, we can define $f(W_c)$, $g(W_c)$ as the unique solution for Q in terms of W_c in the equations $r_1 = R_{\text{th}}$ and $t = R_{\text{th}}$, respectively. It is straightforward through algebraic manipulation to arrive at the following explicit expressions

for f and g:

$$f(W_c) = \begin{cases} \infty & W_c \leq W_0 \left(1 - \frac{R_{th}}{s_1}\right)^+ \\ W_c \left(\frac{2^{\frac{s_1}{W_0}} - 1}{2^{\frac{s_1}{W_0}} 2^{\frac{R_{th} - s_1}{W_c}} - 1} - 1 \right)^+ & W_c > W_0 \left(1 - \frac{R_{th}}{s_1}\right)^+, \end{cases} \tag{3.18}$$

and

$$g(W_c) = \begin{cases} \infty & W_c \leq W_0 \left(1 - \frac{R_{th}}{R^*}\right)^+ \\ \frac{W_c}{a_{2,1}} \left(\frac{2^{\frac{R^*}{W_0}} - 1}{2^{\frac{R^*}{W_0}} 2^{\frac{R_{th} - R^*}{W_c}} - 1} - 1 \right)^+ & W_c > W_0 \left(1 - \frac{R_{th}}{R^*}\right)^+, \end{cases} \tag{3.19}$$

where $x^+ := \max\{x, 0\}$ and $R^* := W_0 \mathsf{C}\left(\frac{P_1}{W_0}\right)$ is the maximum achievable rate for user 1 in the absence of user 2. Noting that s_1 is a constant, the conditions on R_{th} are seen to be equivalent to

$$R_{\text{th}} \leq \min\{r_1, \gamma\} \iff Q \leq \min\{f(W_c), g(W_c)\}, \tag{3.20}$$

$$\begin{aligned} &\min\{r_1, t\} \leq R_{\text{th}} \leq \min\{s_1, t\} \iff \\ &\quad \min\{f(W_c), g(W_c)\} \leq Q \leq g(W_c) \mathbb{1}_{R_{\text{th}} \leq s_1}, \end{aligned} \tag{3.21}$$

and

$$\min\{s_1, t\} \leq R_{\text{th}} \leq t \iff g(W_c) \mathbb{1}_{R_{\text{th}} \leq s_1} \leq Q \leq g(W_c), \tag{3.22}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function of a set.

Since r_2, s_2 and $r_2 + s_1 - R_{\text{th}}$ are increasing functions of W_c, the optimizers in (3.14), (3.15) and (3.16) are achieved on the boundary, i.e.

$$\begin{aligned} v_1 &= \sup_{(W_c, Q) \in \partial_{\text{rm}} \mathcal{D}_1} s_2, \\ \mathcal{D}_1 &:= \{(W_c, Q) \in \mathcal{D} : Q \leq \min\{f(W_c), g(W_c)\}\}, \end{aligned} \tag{3.23}$$

$$\begin{aligned} v_2 &:= \sup_{(W_c, Q) \in \partial_{\text{rm}} \mathcal{D}_2} r_2 + s_1 - R_{\text{th}}, \\ \mathcal{D}_2 &:= \{(W_c, Q) \in \mathcal{D} : \min\{f(W_c), g(W_c)\} \leq Q \leq \\ &\quad g(W_c) \mathbb{1}_{R_{\text{th}} \leq s_1}\}, \end{aligned} \tag{3.24}$$

and

$$\begin{aligned} v_3 = & \sup_{(W_c,Q)\in\partial_{\mathrm{rm}}\mathcal{D}_3} r_2, \\ \mathcal{D}_3 := & \{(W_c,Q)\in\mathcal{D}:\ g(W_c)\mathbb{1}_{R_{\mathrm{th}}\le s_1}\le Q\le g(W_c)\}, \end{aligned} \tag{3.25}$$

where for any region $\mathcal{C}$ in the x-y plane with boundary $\partial\mathcal{C}$, $\partial_{\mathrm{rm}}\mathcal{C}$ denotes the "rightmost" boundary of $\mathcal{C}$ defined by

$$\partial_{\mathrm{rm}}\mathcal{C} := \{(x,y)\in\partial\mathcal{C} : (x,y')\notin\partial\mathcal{C}, \forall\, y' > y\}. \tag{3.26}$$

Next, we discuss some properties of the functions f and g that will help us in solving the subproblems (3.23), (3.24) and (3.25).

3.5 Relative magnitude of f and g

The boundaries of $\mathcal{D}_1$ and $\mathcal{D}_2$ depend on the relative magnitudes of f and g. If f and g intersect, the boundaries $\partial_{\mathrm{rm}}\mathcal{D}_1$ and $\partial_{\mathrm{rm}}\mathcal{D}_2$ may not be smooth curves. In the cases of weak and strong interference, one of f and g dominates the other.

Claim 3.4. *For the case of weak interference $a_{1,2}, a_{2,1} < 1$, $g(W_c) > f(W_c)$ for all $W_c \in [0, W_0]$. For the case of strong interference $a_{1,2}, a_{2,1} > 1$, $g(W_c) < f(W_c)$ for all $W_c \in [0, W_0]$.*

Proof can be found in Appendix A.

When interference is mixed, i.e., $a_{1,2} > 1, a_{2,1} < 1$ or $a_{1,2} < 1, a_{2,1} > 1$, f and g can potentially intersect. It is easy to see that the equation $f(W_c) = g(W_c)$ can be simplified as

$$\begin{aligned} &(1-a_{2,1})2^{\frac{R^*+s_1}{W_0}}2^{(2R_{th}-R^*-s_1)T_c} \\ &+2^{\frac{R^*}{W_0}}(a_{2,1}2^{\frac{\beta_1}{W_0}}-1)2^{(R_{th}-R^*)T_c} \\ &+2^{\frac{s_1}{W_0}}(a_{2,1}-2^{\frac{R^*}{W_0}})2^{(R_{th}-s_1)T_c}+2^{\frac{R^*}{W_0}}-a_{2,1}2^{\frac{\beta_1}{W_0}}=0, \end{aligned} \tag{3.27}$$

where $T_c := \frac{1}{W_c}$. Descartes rule of signs [90] is useful for getting bounds on the number of roots of a polynomial equation by counting sign changes in its sequence of coefficients. The following lemma generalize Descartes rule of signs from polynomials to functions that are a linear combination of exponential terms:

Lemma 3.2. *The number of solutions of the equation* $a_1 e^{b_1 x} + a_2 e^{b_2 x} + \cdots + a_n e^{bnx} = 0$ *in the real variable* x *where* $b_1 < b_2 < \cdots < b_n$ *is at most the number of sign changes in the sequence of coefficients* $(a_1, a_2, \ldots, a_n)$ *and has the same even-odd parity. In particular, any such equation cannot have more than* $n - 1$ *solutions.*

Proof runs along the same line as the proof of Descartes rule of signs and can be found in Appendix A.

Using this Lemma, (3.27) can at most have three solutions in the real variable T_c. Note that $T_c = 0$ is always a solution, however, this corresponds to $W_c = \infty$ which is not acceptable. As such, if the number of sign changes in the (properly sorted) sequence of coefficients in equation (3.27) is two, there is exactly one finite value of W_c such that $f(W_c) = g(W_c)$. However, if the number of sign changes is three, the number of finite values W_c satisfying $f(W_c) = g(W_c)$ is either zero or two.

3.5.1 Characterizing the boundaries of $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$

We saw that the properties of the rate expressions characterizing our rate region are such that the extreme value is taken on the boundaries where different rate expressions dominate. These boundaries are given in terms of the functions f and g as defined above. In this section, we characterize these boundaries.

Note that s_1 is not dependent on the design parameters W_c and Q. If $R_{th} > s_1$, then $R_{th} > \min\{r_1, t\}$ due to the fact that $r_1 \leq s_1$. Hence, $\mathcal{D}_1$ is empty by (3.20) and (3.23). Also, $\mathcal{D}_2$ is empty by (3.24). As such, one only needs to compute v_3. If $R_{th} \leq s_1$, one needs to find all v_1, v_2 and v_3. In what follows, we discuss the cases $R_{th} > s_1$ and $R_{th} \leq s_1$ separately under various interference regimes:

3.5. Relative magnitude of f and g

Weak Interference ($a_{12}, a_{21} < 1$)

- $R_{th} > s_1$**:** In this case, only the rightmost boundary of $\mathcal{D}_3$ is of interest. This boundary is comprised of the curve $Q = g(W_c) : W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [0, g(W_0)]$.

- $R_{th} \leq s_1$**:** In this case, the rightmost boundaries of all $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are of interest. Since $g > f$, $\partial_{\text{rm}}\mathcal{D}_1$ consists of the curve $Q = f(W_c) : W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [0, f(W_0)]$, $\partial_{\text{rm}}\mathcal{D}_2$ consists of the curve $Q = g(W_c), W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [f(W_0), g(W_0)]$ and $\partial_{\text{rm}}\mathcal{D}_3$ consists of the curve $Q = g(W_c) : W_c \in [0, W_0]$. Therefore, $\partial_{\text{rm}}\mathcal{D}_3 \subset \partial_{\text{rm}}\mathcal{D}_2$. Since $R_{th} \leq s_1$, we have $r_2 + s_1 - R_{\text{th}} \leq r_2$ for given Q and W_c. Hence, $v_3 \leq v_2$ and computing v_3 is not necessary.

Strong Interference ($a_{12}, a_{21} > 1$)

- $R_{th} > s_1$**:** In this case, only the rightmost boundary of $\mathcal{D}_3$ is of interest. This boundary is comprised of the curve $Q = g(W_c) : W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [0, g(W_0)]$.

- $R_{th} \leq s_1$**:** In this case, the rightmost boundaries of all $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are of interest. Since $f > g$, $\partial_{\text{rm}}\mathcal{D}_1$ consists of the curve $Q = g(W_c) : W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [0, g(W_0)]$ and $\partial_{\text{rm}}\mathcal{D}_2$ and $\partial_{\text{rm}}\mathcal{D}_3$ coincide and are represented by the curve $Q = g(W_c), W_c \in [0, W_0]$. Since $R_{th} \leq s_1$, we have $r_2 + s_1 - R_{\text{th}} \leq r_2$ for given Q and W_c. Hence, $v_3 \leq v_2$ and computing v_3 is not necessary. Note that $\partial_{\text{rm}}\mathcal{D}_2 \subset \partial_{\text{rm}}\mathcal{D}_1$, however, $r_2 + s_1 - R_{\text{th}} = s_2 + r_1 - R_{\text{th}}$ may be smaller or larger than s_2. As such, one is required to compute both v_1 and v_2.

Mixed Interference ($a_{12} > 1, a_{21} < 1$ **or** $a_{12} < 1, a_{21} > 1$)

- $R_{th} > s_1$**:** In this case, only the rightmost boundary of $\mathcal{D}_3$ is of interest. This boundary is comprised of the curve $Q = g(W_c)\, W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [0, g(W_0)]$.

- $R_{th} \leq s_1$**:** In this case, the rightmost boundaries of all $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are of interest. As f and g may intersect, $\partial_{\mathrm{rm}}\mathcal{D}_1$ consists of the (possibly) non-smooth curve $Q = \min\{f(W_c), g(W_c)\}, W_c \in (0, W_0]$ and the vertical line segment $\{W_0\} \times [0, \min\{f(W_0), g(W_0)\}]$. As in the case of weak interference regime, $\partial_{\mathrm{rm}}\mathcal{D}_2$ consists of the curve $Q = g(W_c), W_c \in [0, W_0]$ and the vertical line segment $\{W_0\} \times [f(W_0), g(W_0)]$ and $\partial_{\mathrm{rm}}\mathcal{D}_3$ consists of the curve $Q = g(W_c) : W_c \in [0, W_0]$. This shows that $\partial_{\mathrm{rm}}\mathcal{D}_3 \subset \partial_{\mathrm{rm}}\mathcal{D}_2$. Following similar line of reasoning as in the case of weak interference regime, we conclude that $v_3 \leq v_2$ and computing v_3 is not necessary.

  Using Lemma 3.2, sufficient conditions can be given in terms of the relationship between channel coefficients and the primary user's R_{th} that constrain the number of intersections of f and g. The following proposition provides conditions such that $\partial_{\mathrm{rm}}\mathcal{D}_1$ consists of both f and g, i.e., f and g intersect over the interval $[0, W_0]$. The unnatural form of inequality on cross-channel gain $a_{2,1}$ is required for the sign-counting to work.

  Proposition 3.1. *Let* $R_{th} < s_1$, $a_{2,1} > \frac{2^{\frac{R^*}{W_0}} - 2^{\frac{R_{th}}{W_0}}}{2^{\frac{s_1}{W_0}} - 2^{\frac{R_{th}}{W_0}}}$ *and* $a_{1,2} < 1$. *Then* f *and* g *intersect at a unique point in the interval* $(0, W_0)$. *Denoting this unique solution of* $f(W_c) = g(W_c)$ *by* $W_c = w$, $\partial_{\mathrm{rm}}\mathcal{D}_1$ *is described by* $Q = f(W_c), W_c \in [0, w]$, $Q = g(W_c), W_c \in [w, W_0]$ *and the vertical line segment* $\{W_0\} \times [0, g(W_0)]$.

  Proof of this can be found in Appendix A.

In the sequel, we provide a numerical example where we demonstrate the situation happening in proposition 1.

3.6 Possible extreme cases of the problem

We have analyzed and broken down the rate optimization problem into its constituent sub-problems. In this section we discuss under which conditions the optimum W_c obtained will be equal to W_0 and 0 respectively. as these two cases correspond to when it's optimal to share the entire primary user's band and sharing exactly nothing

3.6.1 $W_c = 0$

In our formulation of the problem, if $R_{th} > s_1$, the optimal power-bandwidth relationship is given by g, likewise if $R_{th} < s_1$, it is given by $\min\{f, g\}$. Now, the functions f and g have both vertical asymptotes at a positive non-zero W_c at and to the left of which they both go to infinity. Since practically, Q is also limited by user 2's available power P_2, this means that for values of W_c smaller than the asymptote, including zero, the optimal Q equals P_2. On the other hand, for a fixed power budget clipped at P_2, rate is monotone increasing in W_c, hence $W_c = 0$ could not be optimal.

3.6.2 $W_c = W_0$

This case, that corresponds to the total band of primary user being shared can actually happen. One numerical example that this case is indeed possible is $W = 3, W_0 = 2, P_1 = P_2 = 100, a_{12} = 0.1, a_{2,1} = 2.2, R_{th} = \frac{s_1}{5}$ where the optimal W_c equals W_0 and gives $R_2 \approx 7.336$.

3.7 Numerical Examples

To give a better insight into the performance of the proposed scheme, this section provides a number of numerical examples.

First, we consider a scenario where $W = 8\,\text{Hz}$, $W_0 = 6\,\text{Hz}$, $P_1 = 7\,\text{dB}$, $P_2 = 9\,\text{dB}$,

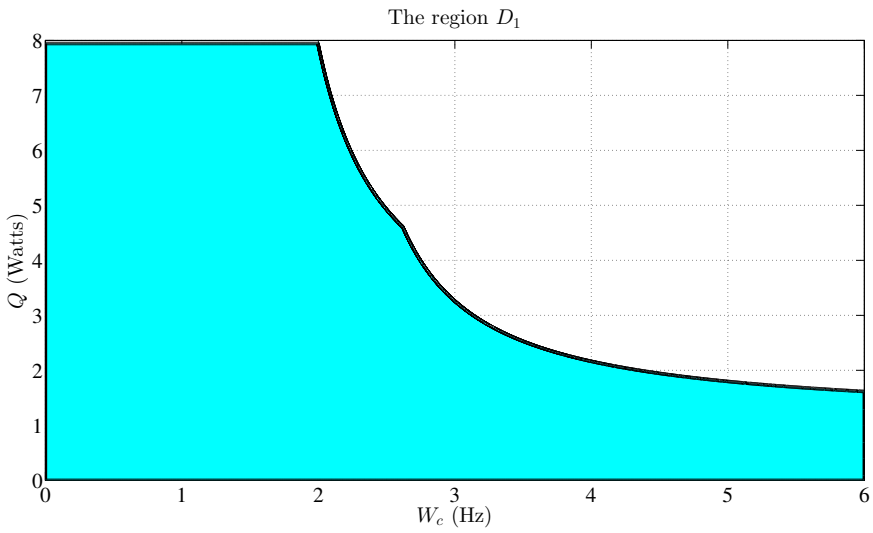

Figure 3.3: The region $\mathcal{D}_1$.

$a_{1,2} = \frac{6}{7}$ and $a_{2,1} = \frac{5}{2}$. In this case, $R^* = 5.6141$ bits/sec/Hz and $s_1 = 4.6742$ bits/sec/Hz. Assume $R_{th} < \beta_1$, say, $R_{th} = 3.5057$ bits/sec/Hz.[4] As discussed in the previous section, we only need to compute v_1 and v_2. The regions $\mathcal{D}_1$ and $\mathcal{D}_2$ are shown in Fig. 3.3 and Fig. (3.4), respectively. Note that the region $\mathcal{D}_2$ is the union of the shaded region and the tail-like curve extending from $W_c = 2.62$ Hz to $W_c = 6$ Hz. Note that $\partial_{\rm rm}\mathcal{D}_1$ has a breaking point at $(W_c, Q) = (2.62, 4.601)$. It turns out that $W_c = 2.62$ Hz is the solution for $f(W_c) = g(W_c)$. In fact, the part of $\partial_{\rm rm}\mathcal{D}_1$ extending from $W_c = 2$ Hz to $W_c = 2.62$ Hz is the curve $Q = f(W_c)$ and the part extending from $W_c = 2.62$ Hz to $W_c = 6$ Hz is the curve $Q = g(W_c)$. This observation is consistent with the result of proposition 1. In fact, $\frac{2^{\frac{R^*}{W_0}} - 2^{\frac{R_{th}}{W_0}}}{2^{\frac{\beta_1}{W_0}} - 2^{\frac{R_{th}}{W_0}}} = 1.5238$. Since $a_{2,1} > 1.5238$, $R_{th} < s_1$ and $a_{1,2} < 1$, the point $W_c = 2.62$ Hz[5] in fact the unique solution w to $f(W_c) = g(W_c)$ as verified by proposition 1. Fig. 3.5, Fig. 3.6 and Fig. 3.7 demonstrate the values of objective functions $\partial_{\rm rm}\mathcal{D}_1$ and $\partial_{\rm rm}\mathcal{D}_2$. It is seen that the optimum tuple (W_c, Q) is given by $(W_c, Q) = (5.36, 6.05)$ which occurs on the part of $\partial_{\rm rm}\mathcal{D}_1$ represented by $Q = \min\{f(W_c), g(W_c)\}$.

[4] We have $\frac{R_{th}}{R^*} = 0.6244$.

[5] While the frequencies used here are too small to be practically relevant, we note that this is an artifact of the chosen input parameters W and W_0 being small. As is customary in information theory literature, small parameter values are selected so that performance parameters can be demonstrated with relatively small numbers. Similar analysis carries over for multi-megahertz bandwidths as well.

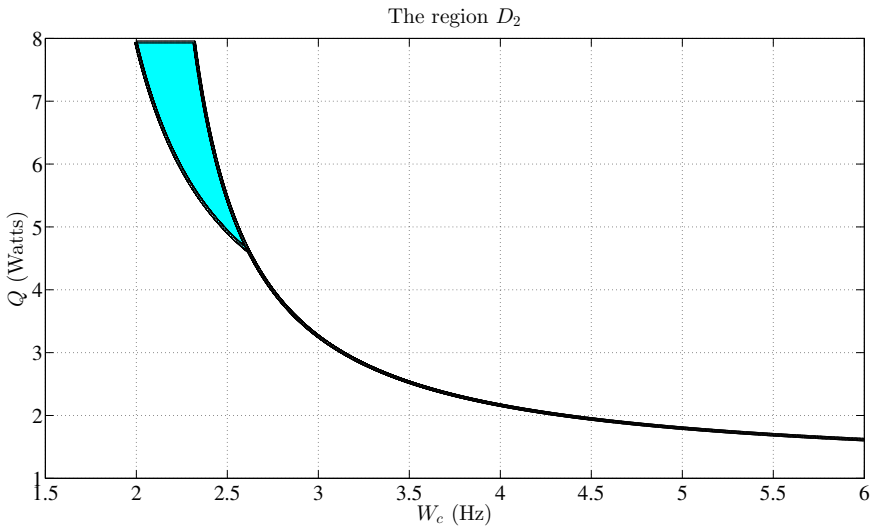

Figure 3.4: The region $\mathcal{D}_2$.

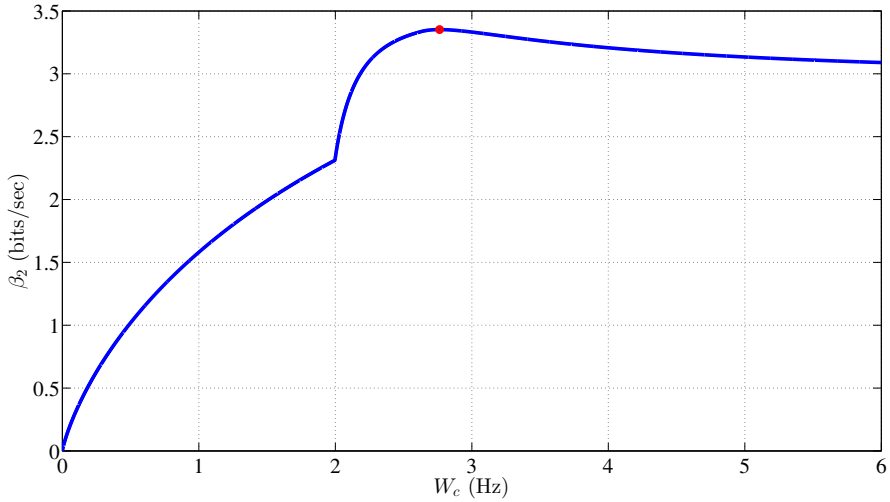

Figure 3.5: Plot of s_2 in terms of W_c on the part of $\partial_{\text{rm}}\mathcal{D}_1$ represented by $Q = \min\{f(W_c), g(W_c)\}$.

3.7.1 Maximum can be attained in different parts of the boundary

In this section, we provide a few concrete numerical examples to show how the point that achieves the maximum rate can be moved to different parts of the boundary by changes in value of the problem parameters. We assume that $W = 3\,\text{Hz}$, $W_0 = 2\,\text{Hz}$, $P_1 = P_2 = 20$ dB, $a_{12} = 0.1$ and $a_{21} = 2.2$. Now we can see that if we set $R_{\text{th}} = 1.5\,\text{bits/s}$, then the maximum will be taken on the rightmost boundary of the region of definition of v_2 and in particular the

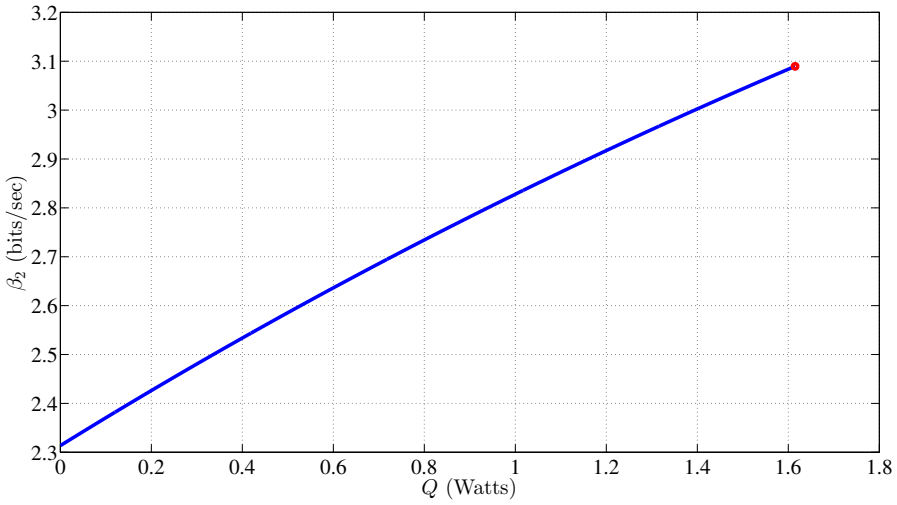

Figure 3.6: Plot of s_2 in terms of Q on the part of $\partial_{\mathrm{rm}}\mathcal{D}_1$ that is a vertical line segment

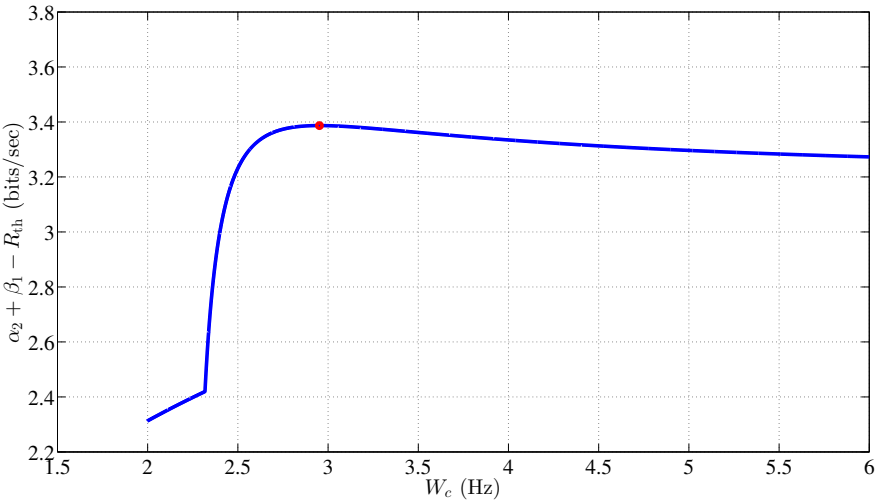

Figure 3.7: Plot of $r_2 + s_1 - R_{th}$ in terms of W_c on $\partial_{\mathrm{rm}}\mathcal{D}_2$ represented by $Q = g(W_c)$.

$W_c = 2\,\mathrm{Hz}$ and $Q = 63.33$ Watts. For the same problem, if we set $R_{\mathrm{th}} = 2.02\,\mathrm{bits/s}$, then the maximum will be taken on the topmost boundary of the region of v_2 for $W_c = 1.91\,\mathrm{Hz}$ and $Q = 57.76$ Watts. More interestingly for $R_{\mathrm{th}} = 1.63\,\mathrm{bits/s}$ the maximums at the topmost and rightmost boundaries coincide at $W_c = 2$ and $Q \approx 59.03$ Watts.

3.7.2 Sensitivity of achievable rates to changes in parameter values

To see how the behaviour of the maximum achievable rate changes with changes to parameter values, we demonstrate here how changes in a_{12} and R_{th} affect the achievable rates, holding all other parameters constant. For $W = 1\text{Hz}$, $W_0 = 0.7\text{Hz}$, $P_1 = P_2 = 100$, $a_{12} = 0.6$ and $a_{21} = 2^6$, Fig. 3.8 shows the effect of changing R_{th} on the maximum achievable R_2. Similarly for $W = 1\text{Hz}$, $W_0 = 0.7\text{Hz}$, $P_1 = P_2 = 100$ and $a_{21} = 2$, Fig 3.9 shows the effect of changing a_{12} on the maximum achievable R_2 for the case that R_{th} is chosen to be $0.33s_1$.

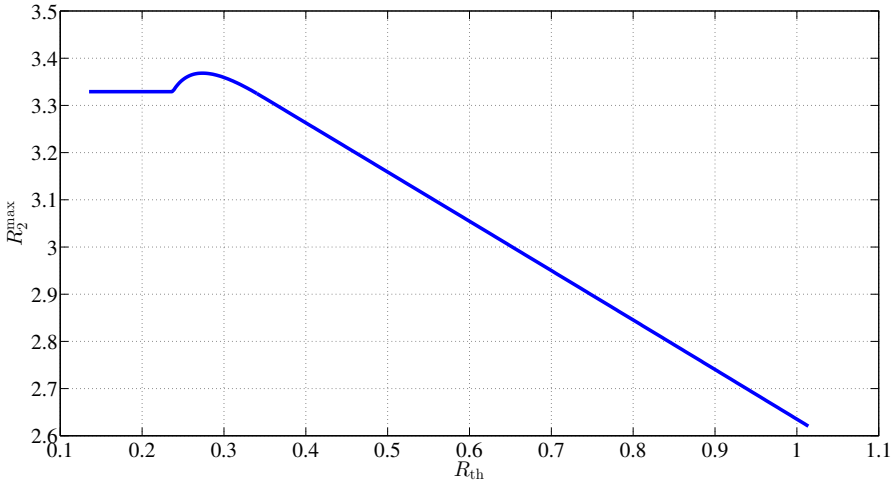

Figure 3.8: Plot of maximum achievable R_2 as a function of changing R^*.

3.8 Shape of the rate curve

To complete our numerical examples, in this section, we also provide examples of the overall shape of the rate region. In figure 3.10 you can see the maximum value of R_2 attainable by this scheme for $W = 5$, $W_0 = 2$, $a_{1,2} = 0.5$, $a_{2,1} = 5$ and $P_1 = P_2 = 10\text{dB}$ as a function of $R_1 = R_{th}$. The rate region is also calculated for $W = 5$, $W_0 = 2$, $a_{1,2} = 0.2$, $a_{2,1} = 0.2$ and

[6] As discussed previously, small round numbers are used for input parameters, as is customary in information theory literature, to be able to demonstrate the example without having to deal with many significant digits.

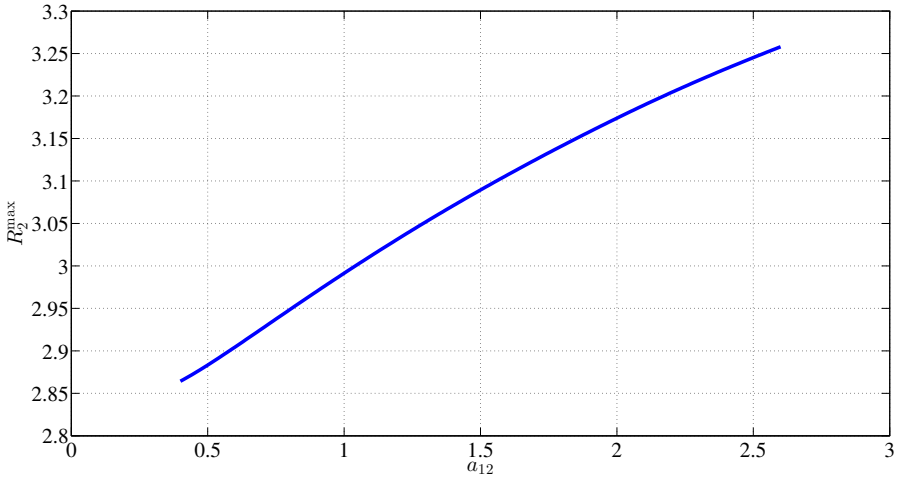

Figure 3.9: Plot of maximum achievable R_2 as a function of changing a_{12}.

$P_1 = 17$dB $P_2 = 13$dB as is shown in figure 3.11. As can be seen, this region is not in general convex.

3.9 Conclusion

Thus far, we proposed and analyzed the interference channel with one cognitive user. We proposed a series of achievability schemes for this channel that all achieve the same rate region, but do so by emphasizing different components of the trade-off between conceptual and computational simplicity of coding and decoding schemes. We then analyzed the rate optimization problem that the cognitive secondary user faces and noted its non-convexity. The structure of this problem was exploited to reduce its solution to a number of subproblems of smaller dimension. Rigorous results were obtained on the structure and interrelationship of these subproblems in different interference strength regimes. Numerical examples were provided to supplement these results and shed light on the behaviour of our problem and give an idea of its performance for some ranges of parameters.

3.9. Conclusion

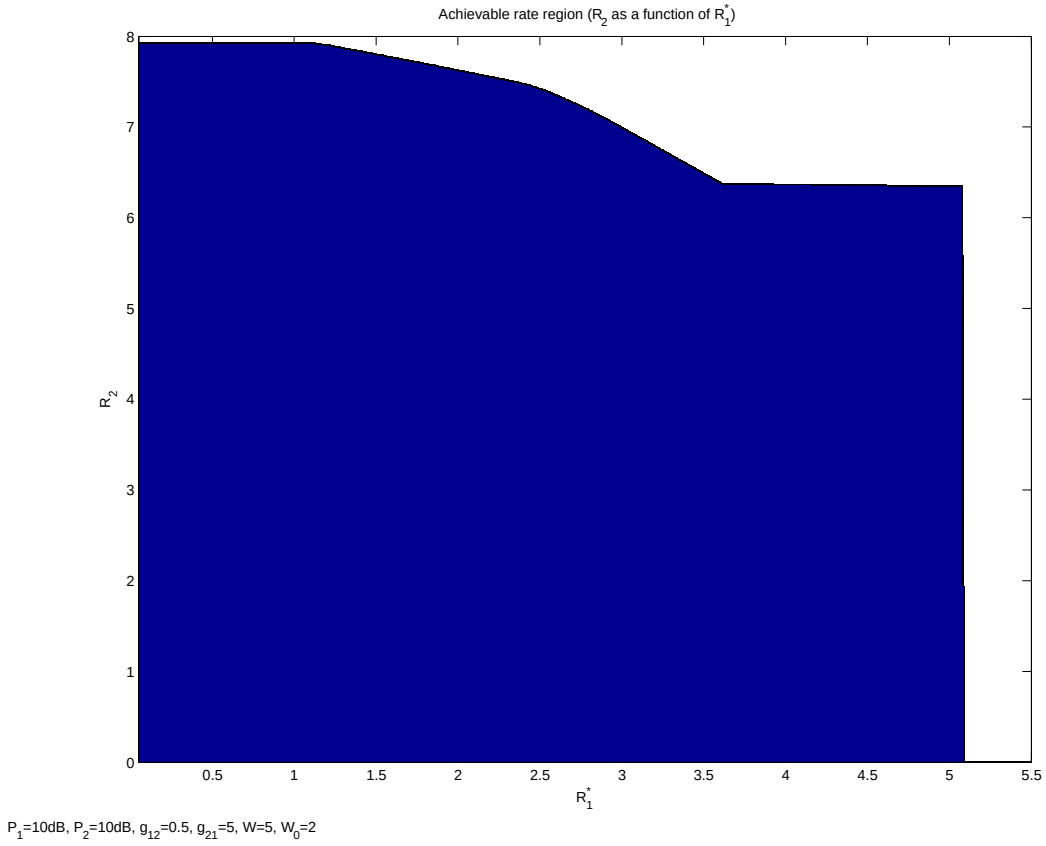

Figure 3.10: Achievable rate region for $W = 5$, $W_0 = 2$, $a_{12} = 0.5$, $a_{21} = 5$ and $P_1 = P_2 =$ 10dB

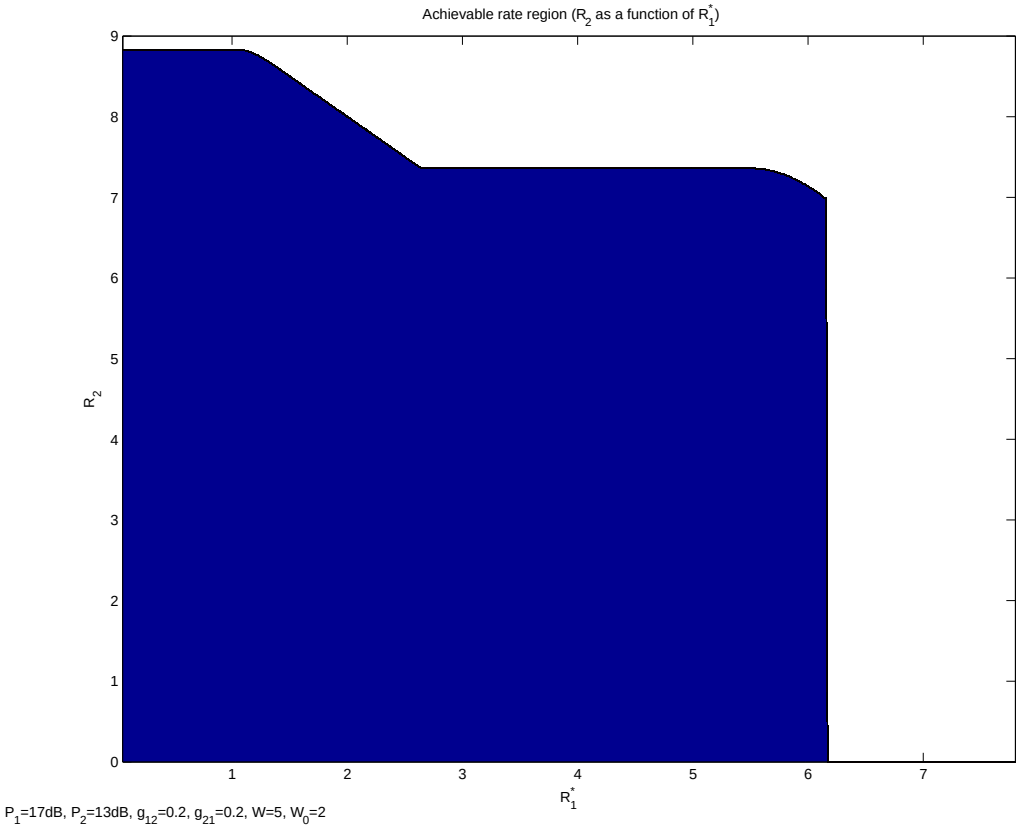

Figure 3.11: Achievable rate region for $W = 5$, $W_0 = 2$, $a_{1,2} = 0.2$, $a_{2,1} = 0.2$ and $P_1 = 17$dB $P_2 = 13$dB

Chapter 4

Approximate link scheduling in large networks

This chapter concerns the second part of the thesis that deals with the problem of SINR-feasible link scheduling in large wireless networks. Contrasting the approach taken in the previous chapter, we turn our attention from the two-user networks to the issue of interference management in the presence of large number of interfering users. The scale of these networks makes complex encoding and decoding schemes impractical and leaves only the simplest transmitter and receiver design strategies like treating interference from structured messages of other users as added noise. These simplified approaches are not without theoretical basis. As noted previously, the importance of interference management and interference control has been put on theoretically firm ground by the advent of degrees-of-freedom (DOF) analysis and interference alignment (IA) that put the burden of interference management on the transmitter and treat whatever remains at the receiver as noise. But these approaches, based mostly on channel-state dependent transmitter-side constellation designs, have offered limited prospects for large-scale practical applications due to excessive complexity of channel-code design and encoding operations[12, 13, 82]. Recently, a series of results have emerged that show near-optimality of treating-interference-as-noise (TIN) without any transmitter

side signal shaping for large classes of interference networks [62, 85]. This has paved the way towards providing a more direct mapping from theory to practice in interference management, and suggests an approach that is particularly suited to scenarios encountered in emerging large Machine-to-Machine (M2M) and Device-to-Device (D2D) networks. These developments stress the importance of treating Signal to Interference as Noise Ratio (SINR) itself as a key metric to optimize for attaining the best achievable rates in large networks, instead of seeing it just as an approximation to the *true* capacity metric whose only advantage is simplicity of analysis. Inspired by these results, new heuristic SINR-aware wireless scheduling algorithms have appeared in the literature including the previously reviewed FlashLinQ [60] and ITLinQ [61] that aim at producing higher-throughput link schedules that target SINR levels and are not computationally expensive. As discussed previously, this combination of theoretical and practical developments means that investigating approaches to link scheduling that are simple, scalable and have theoretically provable performance guarantees is an important research direction. This chapter, after defining a network model and stating the scheduling problem formally, describes an approximate scheduling algorithm that has provable performance guarantees and show that it performs well in practice, specially as the size of the network grows.

The outline of the rest of the chapter is as follows. Section 4.1 describes the model. Section 4.2 gives a formal definition of the minimum length scheduling problem. In section 4.3 the problem of finding the minimum-length schedule is formulated as a mixed integer program. Using this framework, mathematical programs are given that are amenable to solution using optimization tools such as CPLEX[91] and Gurobi[92]. Mixed integer programming gives exact solutions but the time and memory requirements grow very fast with the instance size, therefore an alternate formulation is given that can be repeatedly solved to obtain successively refined lower bounds[1] and feasible solutions and which provides an interval containing the exact minimum length and accommodates larger instance sizes.

[1] A lower bound is inferred when the MIP solver declares that the instance has no feasible integer solution with the given schedule length.

Next, an approximation algorithm for the solution of the link scheduling problem based on graph-theoretical ideas is proposed. This algorithm takes advantage of the geometric structure imposed by the metric space of the link endpoints to decompose the links into subsets having specific properties and uses a specialized scheduling algorithm as a subroutine to separately schedule subsets of the resulting decomposition. The proposed algorithm finally composes these sub-schedules into an approximate schedule for the general-form input instance. Section 4.4.1 defines the criteria based on which this decomposition process takes place. This section also gives notation and definitions for other properties that are used to further describe the scheduling algorithm and specify its approximation ratio. This algorithm and its building blocks are introduced in Section 4.4.2. This includes a flow-chart of the top-level method as well as the special-case approximate scheduling algorithm that is used as a subroutine. The description of the steps these algorithms take is accompanied with a series of lemmas establishing their correctness and approximation ratio. Proofs of these lemmas are outlined but the details are relegated to Appendix B. Finally, we report on numerical experiments and conclude this chapter in Section 4.5. Part of the work described in this chapter has been previously presented in the paper [93].

4.1 Model and assumptions

This section describes the underlying model of the wireless network. Our model considers a set of $\{1, ..., L\}$ links. Each link $l \in L$ is characterized by the tuple $(o(l), d(l), p(l), G(l))$.

The points $o(l), d(l)$ are the origin and destination points of link l. These points are only constrained to be in a doubling metric space M[2]. The use of doubling metric space assumption is technical as we make use of triangle inequality and doubling constraint in proving some lemmas. To make this technical assumption more intuitive, the reader can

[2]A doubling metric space is a metric space for which there are constants C and m such that every ball $B(p, r)$ of radius r can contain at most $C\left(\frac{r'}{r}\right)^m$ disjoint balls of radius r'. The number m is called the doubling dimension of the metric space.

assume that all nodes are in the Euclidean $\mathbb{R}^2$ plane[3]. All flows are single-link, so the input instance represents L simultaneous data transfer demands.

Link power $p(l)$ is given by the power-control scheme that we adopt. This is a local link-length dependent power control of the form $p(l) = p_0 \left(D_{o(l),d(l)}\right)^{\alpha/2}$ where D is the distance metric. We use the shorthand $\ell(l)$ for $D_{o(l),d(l)}$ for cleaner notation. This power control choice is halfway between full channel inversion (equal received power at all destinations) and uniform power (equal transmit power at all origins). The reason is that the aforementioned two power control schemes can produce schedules that are in the worst-case exponentially longer in some problem parameters than that achieved by $\ell(l)^{\alpha/2}$ power control [50]. In the case of uniform transmit power, this is caused by longer links being disadvantaged compared to shorter links. In the case of complete channel inversion, longer links will cause undue interference at other receivers.

Link gains $G(l)$ are characterized by a path loss model with a loss exponent of α. Therefore, for a link l, the direct gain is given by $|G(l)|^2 := |G_{o(l),d(l)}|^2 = g_0(D_{o(l),d(l)})^{-\alpha}$. Typical values of α are in the range of 2 to 6 [94].

The interference model we are considering is based on an additive white Gaussian noise (AWGN) interference channel, hence the received signals at the destination of link l is given by

$$r_l = \sum_{l'=1}^{L} G(l', l)s_l + z_l, \tag{4.1}$$

where $G(l', l)$ is the cross-path coefficient from the origin of l' to the destination of l, s_l is the signal sent by the transmitter at $o(l)$, and the terms z_l are independent Gaussian noise terms with a joint distribution of $\mathcal{N}(0, N_0 I)$. Similar to the direct gains, the cross-channel gain of interference from $o(l')$ on $d(l)$ is given by $|G(l', l)|^2 := |G_{o(l'),d(l)}|^2 = g_0(D_{o(l'),d(l)})^{-\alpha}$. We assume that links are synchronized so there is no ambiguity around time indices or any problems with synchronization. Channel gains are assumed to take on real values.

[3]For the Euclidean space $\mathbb{R}^m$, the doubling dimension is equal to m, the ordinary notion of dimension of the space.

We make the major simplifying assumption that the links are single-rate, so we impose the uniform SNR threshold of β. We require that

$$\gamma_l = \frac{p(l)G(l)}{N_0 + \sum_{\substack{l' \neq l \\ l' \text{ active with } l}} P(l')G_{l'l}} \geq \beta,$$

for link l to do a successful transfer. In practice, β as a parameter is a function of physical layer technology, RF chain sensitivity and the desired bit-error-rate.

With the elements of the network model described, the next section defines the problem that we are trying to solve.

4.2 Formal definition of scheduling problem

The model of the network was defined in the previous section. We are now ready to give a formal definition of the scheduling problem. The problem input instance is a set of network links $\{1, ..., L\}$. Also, two sets of parameters are used to characterize the exact conditions under which the network operates. The first set of parameters is a function of the ambient space that our link endpoints operate in (which is itself assumed to be a metric space M with doubling dimension m and distance measure D). They are the path loss exponent α, noise spectral density N_0 and the normalized unit distance gain g_0. The second set consists of parameters characterizing the radio hardware at the link endpoints and consists of the normalized transmit power p_0 and receiver SNR threshold β. Each link in the input instance is characterized by the tuple $(o(l), d(l), p(l), G(l))$ where $o(l), d(l)$ are the free parameters characterizing the link's origin and destination. They fully characterize the other two components of the tuple as $p(l) := p_0 D_{o(l),d(l)}^{\alpha/2}$ and $|G(l)|^2 := g_0 D_{o(l),d(l)}^{-\alpha}$. As noted before, the shorthand notation $\ell(l) := D_{O(l),d(l)}$ is used for the length of link l. Using this shorthand, link transmit power and gain can be written as $p(l) = p_0 \ell(l)^{\alpha/2}$ and $|G(l)|^2 = g_0 \ell(l)^{-\alpha}$, respectively. As shown by equation 4.1, the effect of interference is modeled by an additive channel

with cross-gains of $|G(l',l)|^2 = g_0 D^{-\alpha}_{o(l'),d(l)}$ from link l' on link l. As described previously, all gains are real-valued and the noise is white, Gaussian-distributed and additive.

The output corresponding to an input instance, the optimum schedule, consists of a partition of $\{1,...,L\}$ into subsets $\{s_1,...,s_S\}$, such that each l is in exactly one s_t, or

$$t,t' \in \{1,...,S\}, t' \neq t \Rightarrow S_t \cap S_{t'} = \emptyset \tag{4.2}$$

$$\bigcup_{t=1}^{S} s_t = \{1,...,L\}, \tag{4.3}$$

$$\tag{4.4}$$

and for each s_t, all links $l \in s_t$ satisfy

$$\forall s_t, \forall l \in s_t \quad \gamma_l := \frac{p(l)G(l)}{N_0 + \sum_{l' \neq l \in s_t} P(l')G_{l'l}} \geq \beta. \tag{4.5}$$

Also, and crucial to optimality, the cardinality S of the partitioning $\{s_1,...,s_S\}$ is minimum among all such partitions satisfying the above criteria.

It is clear that having such optimum $\{s_1,...,s_S\}$, we can associate each s_t to a different time instant such that all transfer requests are satisfied and this is the shortest length of timeslots in which this can happen. This justifies the name minimum-length schedule.

The sets s_i satisfying the above criteria are called, in analogy to vertex-independent sets of a graph, scheduling-independent sets or **ISets** for short. With this terminology, optimum scheduling is the problem of partitioning the set $\{1,...,L\}$ into the minimum number of ISets.

Formally, the parameters, input and output of the optimum link scheduling problem are:

Parameters: path loss exponent α, metric M, dimension m and associated distance measure D, Noise spectral density N_0, normalized gain g_0 and power p_0 and SNR threshold β.

Input: a set of links $\{1,\cdots,L\}$ with each link l associated to the tuple $(o(l),d(l),p(l),G(l))$ of which $o(l),d(l)$ determine the other two components based on the parameters. Similarly

the cross gains $G(l, l')$ are defined by the path loss formula as described before.

Output: a number S and sets $s_1, s_2, \ldots, s_S$ such that:

1. s_t's are disjoint and cover all of $\{1, ..., L\}$.
2. s_t's are scheduling ISets as defined previously.
3. Their cardinality, S is minimum among all partitions satisfying the previous two conditions.

It is obvious that this problem has a well-defined, computable solution. In particular, the cardinality S of the partitioning is at least 1 and at most L and there is only finitely many partitions of $\{1, ..., L\}$ to examine and find the smallest consisting only of ISets.

If the problem output specification above is relaxed to finding, instead of the exact minimum partition, a partition whose length is approximately close to minimum, we will have the problem of approximate link scheduling. For a positive non-decreasing function h of instance size and parameters, an algorithm for the approximate link scheduling problem is said to $\mathcal{O}(h)$-approximate link scheduling if the ratio of the length of its output schedule to the minimum-length output schedule asymptotically grows like $\mathcal{O}(h)$.

4.2.1 Example of a practical application scenario

A scenario where solving this problem is useful is for a large-scale network of autonomous radio devices arranged in peer-to-peer setup. An example of this is a large scale device-to-device network of autonomous vehicles or drones. The network will operate over consecutive rounds where a static set of demands declared at the start of the round are serviced in a finite number of transmission timeslots before the start of the next round. In order to maximize utilization of the network, we want to conclude each round in as small a number of transmission slots as possible. A scheduling algorithm decides on the number of timeslots and the links transmitting in each timeslot of the round. Therefore, an exact or approximate

solution to finding the minimum number of timeslots required to schedule a set of transmission requests allows achieving the highest possible link utilization and minimizes the time between successive rounds. Having a scalable algorithm is important if the size of network is large or if limited computational resources are available at each node.

4.2.2 Complexity of exact scheduling

As discussed previously, the scheduling problem can in principle be solved by exhaustively examining all partitions of the set $\{1, ..., L\}$ and producing the one that has the smallest cardinality (length) and consists only of β-SINR feasible ISets. This brute force approach does not exploit any structure present in the problem instance and takes exponential time in L. Therefore, solving the SINR-based minimum-length with more clever algorithms and characterizing its worst-case performance is an important problem. Since many simpler versions of link scheduling, for instance graph-based independent set scheduling discussed previously are equivalent to NP-complete problems (such as producing vertex colourings of general graphs), it might seem reasonable to assume that the problem defined above is NP-complete as well. This is not a rigorous argument as there might be a smart way of exploiting the structure inherent in the problem to drastically cut down on the time required to solve it. In contrast, formally showing that this problem is NP-hard, involves finding a way of reducing general instances of an already known NP-complete problem to it. This section reviews the literature on the complexity of link scheduling problem.

Apart from results on graph vertex colouring, Turong et al. in [95] were the first to look at the complexity of wireless link scheduling. They consider an interference model where the two-hop neighbourhood of any two transmitting nodes should be disjoint and show that obtaining the optimal schedule in this setting is NP-complete. Borbash et al. in [96] consider the problem of SINR-based scheduling without geometric constraint on channel gains and show a reduction to it from a generalized graph-matching problem which they hypothesize to be NP-complete. Later, Behzad et al. in [97] gave a linear programming

formulation of the minimum-length scheduling with a SINR metric that is of exponential size in problem parameters. They also proved, by a reduction from instances of graph edge-colouring problem to their formulation, that SINR-based scheduling with general channel coefficients is NP-Complete. Both of these reductions lead to arbitrary matrix of channel gains without any geometric structure. Of more interest to our case, Goussevskaia et al. in [51] showed that even instances of minimum-length SINR-based scheduling where channel gains are constrained to be related to link lengths by a path loss formula are still expressive enough to represent general instances of NP-complete problems. In particular, they show through a reduction from general instances of set-partitioning problem to geometric instances of SINR-based scheduling, that even in this restricted setting, scheduling is NP-complete. This result definitively establishes that the problem we are trying to solve is NP-complete. This makes it imperative to look at approximate solutions for larger instances. The next section shows that scheduling can be formulated as a mixed integer program in a format that is suitable for mathematical programming software and that these software tools can be used to find exact solutions for intermediate-sized instances.

4.3 Exact solution of scheduling using a mixed integer program

Having formally defined the problem of finding exact minimum-length SINR-feasible schedule and looked at its complexity, this section formulates this problem as an instance of Mixed Integer Programming (MIP). This allows us to enlist the well-developed MIP-solving capabilities of general-purpose optimization software for its solution. The standard MIP formulation also serves as a concise alternative formal definition of the problem. Before that, we briefly discuss the subject of mixed integer programming.

A mixed integer program is an optimization problem (mathematical program) in which the objective function and constraints are linear functions of the variables and with the

added constraint that some variables only take on integer values. The addition of integrality constraints distinguishes these problems from linear programs (LP) and gives them extra expressive power. This expressiveness comes with the disadvantage that standard simplex and interior point methods are not adequate for solving MIP instances. In the worst case, solving MIP's requires solving an exponentially large number of LP instances corresponding to fixed choices for the values of discrete variables. Still, there is a group of well-developed techniques for pruning this search tree of discrete choices which, together with heuristics for ordering these choices, makes many mixed integer programs of interest effectively solvable. For general MIP problems, these techniques include the branch and bound [98] and branch and cut [99] methods. Both of these methods require many calls to LP solvers on intermediate instances obtained by relaxations of the original program and use the results to further prune the set of remaining tree nodes they have to examine. This reliance on search tree pruning in general MIP solving strategies has consequences both in formulating combinatorial optimization problems as MIP instances and in the choice of successive relaxations that solvers internally make. One example is that in formulating a problem as a mixed integer program, there is often a trade-off between the work done per node of the search tree and the number of nodes that need to be examined on average. A formulation with more constraints, which leads to an increase in the size of intermediate LP instances, can prune the search tree nodes more effectively on average and therefore be more efficient overall. A formulation with less constraints, on the other hand, solves smaller intermediate LP instances but might need to look at a larger fraction of the search tree nodes. In any case, the framework of general mixed integer programs is so expressive that it can represent every NP-complete problem, so there is good reason to believe that this exponential search tree complexity might be inherent.

The problem we are trying to solve is a general set partitioning problem with extra constraints, where these constraints needs to only hold between elements that are inside the same subset of the partition (namely, the SINR feasibility constraint of link has to hold only

relative to links in the same partition as itself). One way to formulate this problem is to use binary indicator variables indexed by both link and ISet indices to signify when a link belongs to a certain ISet. The SINR feasibility constraint on a link when receiving interference from all links in some ISet, then, will only need to hold if the indicator variable of the link in that ISet is 1. It seems hard at first to express this *logical* constraint activation without using multiplication and when confining ourselves to linear expressions in the indicator variables. The big-M trick [100] is a method in mathematical programming that can be used, among other things, to formulate these logical constraints that are not strictly linear, as linear constraints of a specialized form. As mentioned, the SINR constraint is of this logical type as its incorporation is predicated on the link belonging to a certain ISet. Using the big-M trick, a constraint activation condition such as our SINR feasibility criterion will be incorporated by subtracting from the right hand side of the unconditional constraint written as a $\geq$ inequality, a large number M[4] times the indicator variable of interest. With this extra term, whenever the indicator variable is 1, the inequality is vacuously satisfied and does not further constrain other variables involved.

With this primer, two formulations are presented next. The first formulation uses a smaller number of constraints and is symmetric under renaming ISet indices. It has many feasible solutions corresponding to what is qualitatively the same scheduling solution and as a result can require a larger branch and bound tree. The second formulation, on the other hand tries to order the ISet indices by their cardinality and breaking ties lexicographically. This is done in order to make the correspondence between MIP and original problem solutions one-to-one and the search tree smaller. This comes at the cost of extra constraints to enforce this ISet ordering, which increase the size of relaxed instances that need to be solved at intermediate nodes of the branch and bound tree.

[4]Hence the name big-M.

4.3.1 First approach

Similar to what we did for links in the general formulation, we equate the ISets S_s with their index s, so that both links and ISets are identified with integers. This helps make the mathematical programs that follow cleaner. With this convention, the lowercase l is used to index the set of links and lowercase s the set of ISets. Using indicator variables x_{ls} to denote link l belonging to set s, the problem of minimizing the schedule length can be formulated as the mathematical program below (with the use of shorthand notation $[N]$ for the set $\{1, ...N\}$):

$$
\begin{aligned}
\min_{S,\{x_{ls}\}} \quad & S \\
\text{s.t.} \quad & x_{ls} \in \{0,1\} && S \in \mathbb{N}, \forall l \in [L], \forall s \in [S] \\
& \sum_{s=1}^{S} x_{ls} = 1 && \forall l \in \{1,..,L\} \\
& x_{ls} G(l) p(l) \geq x_{ls} \beta N_0 + \sum_{\substack{l' \neq l \\ l' \in s}} x_{l's} p(l') G(l', l) - M(1 - x_{ls}) && \forall l \in \{1,..,L\}, \forall s \in \{1,...,S\}.
\end{aligned}
\tag{4.6}
$$

As can be seen, The first constraint signifies that x_{ls} is the binary indicator of whether or not link l is put into partition s. The second constraint states that the schedule should be valid, that is every link should be scheduled exactly once. The last constraint uses the big-M trick to enforce the SINR constraint only when a link belongs to an ISet.

This formulation is a straightforward translation of the problem. Unfortunately, since the range of indices for variables x_{ls} depends on another program variable S, this formulation does not strictly conform to the definition of MIP. This can be overcome as will be discussed shortly, but a variation of this formulation, as a mixed integer feasibility problem[5] can be used as a decision procedure to determine whether a certain number of ISets suffices for a given set of links. As such, repeatedly solving of this variation can give successively tighter

[5]In an MIP feasibility problem, the goal is to only determine whether the feasible region of a set of MIP constraints is nonempty.

feasible solutions and infeasible schedule lengths which act as upper and lower bounds for the exact solution. When these bounds meet, we know that we have that exact solution.

Conversion to this feasibility problem works by replacing the range of s in x_{ls} with the constant s_0 for which we want to check if there is a solution with that many ISets. This will lead to the following program:

$$
\begin{array}{lrcll}
\min\limits_{\{x_{ls}\}} & 2 & & & \\
\text{s.t.} & x_{ls} & \in & \{0,1\} & \forall l \in [L], \forall s \leq s_0 \\
& \sum\limits_{s=1}^{s_0} x_{ls} & = & 1 & \forall l \in \{1,..,L\} \\
& x_{ls} G(l) p(l) & \geq & x_{ls}\beta N_0 + \sum\limits_{\substack{l' \neq l \\ l' \in s}} x_{l's} p(l') G(l',l) - M(1-x_{ls}) & \forall l \in \{1,..,L\}, \forall s \in \{1,...,s_0\},
\end{array}
\tag{4.7}
$$

which has the constant 2 as the dummy objective to show that it is only the feasibility of constraints that are important. For any given s_0, the above problem is feasible if and only if s_0 is equal to or greater than the minimum schedule length for that instance. This is a proper MIP feasibility problem. The difference between Programs 4.6 and 4.7 can be seen by way of an example:

Assuming that we have a set of $L = 100$ links represented as $L = \{1, ..., 100\}$ and we like to determine an interval containing the optimum schedule length. It is very unlikely that all links are simultaneously feasible, and it is also unlikely that 100 individual timeslots are required. Therefore a reasonable first guess might put the optimum value in the range of 15

to 35 ISets. If the program

$$
\begin{array}{lrcll}
\min\limits_{\{x_{ls}\}} & 2 & & & \\
\text{s.t.} & x_{ls} & \in & \{0,1\} & \forall \ell \in L, \forall s \in \{1,...,35\} \\
& \sum\limits_{s=1}^{35} x_{ls} & = & 1 & \forall l \in L \\
& x_{ls} G(l) p(l) & \geq & x_{ls}\beta N_0 + \sum\limits_{\substack{l' \neq l \\ l' \in s}} x_{l's} p(l') G(l', l) - M(1 - x_{ls}) & \forall j \in N,
\end{array}
\tag{4.8}
$$

is feasible and replacing 35 by 15 makes it infeasible, our initial guess of $[15, 35]$ is validated and we can tighten the range to, say $[20, 30]$. Otherwise, depending on the solution of the two programs, the range should be enlarged from one end until it contains the exact solution. By successively tightening the upper and lower ends of the range, the exact optimum s can be found. An advantage of this formulation is that the gradual refinement of the interval means that we have an upper and lower bound to the exact answer even computation is stopped before the exact solution is pinpointed.

Program 4.6 can be converted to a proper MIP by noting that S is at most L (the ISets are singleton sets in the worst case). The range of indices for x_{ls} can now be changed accordingly. To accommodate the new possibility that some of the L ISet indices may not be used, a new set of indicator variables y_s can be added that indicates whether the s'th

ISet is empty or not. With these changes, the program below results:

$$
\begin{array}{lrcll}
\min\limits_{S,\{x_{ls}\},y_s} & S & & & \\
\text{s.t.} & x_{ls} & \in & \{0,1\} & \forall l,s \in \{1,...,L\} \\
& \sum\limits_{s=1}^{L} x_{ls} & = & 1 & \forall l,s \in \{1,...,L\} \\
& y_s & \in & \{0,1\} & \forall s \in \{1,...,L\} \\
& y_s & \geq & x_{ls} & \forall l,s \in \{1,...,L\} \\
& y_s & \leq & \sum\limits_{l=1}^{L} x_{ls} & \forall s \in \{1,...,L\} \\
& \sum\limits_{s=1}^{L} y_s & = & S & \\
& x_{ls}G(l)p(l) & \geq & x_{ls}\beta N_0 + \sum\limits_{\substack{l' \neq l \\ l' \in s}} x_{l's}p(l')G(l',l) - M(1-x_{ls}) & \forall j \in N.
\end{array}
\tag{4.9}
$$

The added constraints on the y_s's enforce that y_s represents the non-emptiness of ISet s and that their sum is equal to the variable S which is also the program objective.

As pointed to previously, this mixed integer program has many solutions corresponding to any distinct solution of the original problem, all of which are equivalent up to a relabeling of the ISet indices s. In practice, a general-purpose branch and bound solver without information about this symmetry might need to look at a much larger number of search tree nodes than absolutely required to conclude that it has exhausted all possibly better solutions.

4.3.2 Adding ordering constraints to reduce symmetry

The formulation in the previous section was a straightforward translation of problem objective and constraints. It had the disadvantage that each qualitatively distinct solution of the original problem was equivalent to large number of MIP solutions by relabeling of the indices s. The formulation presented here tries to alleviate this blowup using extra constraints

that favour smaller indices s for the larger ISets and break ties by lexicographic ordering of the elements of $[L]$. Using similar notation and indicator variables names, x_{ls}, y_s as in the previous section, this formulation adds two extra constraints as presented below:

$$
\begin{aligned}
\min_{S,\{x_{ls}\},y_s} \quad & S \\
\text{s.t.} \quad & x_{ls} \in \{0,1\} && \forall l, s \in \{1,...,L\} \\
& \sum_{s=1}^{L} x_{ls} = 1 && \forall l, s \in \{1,...,L\} \\
& y_s \in \{0,1\} && \forall s \in \{1,...,L\} \\
& y_s \geq x_{ls} && \forall l, s \in \{1,...,L\} \\
& y_s \leq \sum_{l=1}^{L} x_{ls} && \forall s \in \{1,...,L\} \\
& y_{s+1} \leq y_s \\
& \sum_{s=1}^{L} y_s = S \\
& x_{ls} G(l) p(l) \geq x_{ls} \beta N_0 + \sum_{\substack{l' \neq l \\ l' \in s}} x_{l's} p(l') G(l', l) - M(1 - x_{ls}) && \forall j \in N \\
& \sum_{l=1}^{L} x_{ls} \geq \sum_{l=1}^{L} x_{l,s+1} && \forall s \in \{1,...,L-1\} \\
& \sum_{l=1}^{L} l x_{ls} \geq \sum_{l=1}^{L} l x_{l,s+1} - M'(\sum_{l=1}^{L} x_{ls} - \sum_{l=1}^{L} x_{l,s+1}) && \forall s \in \{1,...,L-1\}.
\end{aligned}
\tag{4.10}
$$

The last constraint uses a big M' value to enforce ordering by cardinality (the other parts of the constraint enforce lexicographic ordering when cardinalities are equal). The second to last constraint requires smaller s indices to be used for nonempty ISets first before using larger s values. The SINR criterion with the big-M trick is the same as the previous formulation, as are the constraints used to enforce x_{ls} and y_s representing a proper partitioning of $[L]$.

These mixed integer programs can be solved using standard mathematical programming software such as CPLEX [91] and Gurobi [92] to obtain the exact solution. As it will be discussed in Section 4.5, programs 4.9 and 4.10 become prohibitive for instances of larger

than about 50 and 30 links respectively. As such, in that section, we use the bounding approach along with program 4.7 to compare the proposed algorithm with the interval bounds containing the exact solution, for medium-sized instances of up to 250 links. With the MIP formulations of the exact problem provided, the next section turns attention to the proposed approximate scheduling algorithm, where its operation is described and correctness and approximation ratio are established.

4.4 The proposed algorithm for approximating the optimum schedule

This section defines the scheduling algorithm of our schemes and provides an analytical treatment of its performance. We first give a preliminary account of the top-level organization of this algorithm together with required definitions in Section 4.4.1. Detailed discussion of the algorithm and lemmas and theorems characterizing its performance are given in Section 4.4.2.

4.4.1 Notation and preliminaries

We defined the notation used for the links and described what an ISet is while formally defining the minimum-length scheduling problem. This section is both a recap of those and also defines new terms that are used in describing our approximate scheduling algorithm. The set of intended transmit-receive pairs is denoted by set $\{1, 2, \ldots, L\}$ of links as before, where each link l represents a transmitter located at $o(l)$ and a receiver at $d(l)$. The distance metric between two points p_1 and p_2 is being denoted by $D(p_1, p_2) := D_{p_1,p_2}$ and we use the shorthand $D_{l,l'} = D_{o(l),d(l')}$. As previously noted, $\ell(l) := D_{o(l),d(l)}$ denotes the length of link l.

We use the notation $[L]$ for the set $\{1, ..., L\}$, and more generally $[N]$ for a set $\{1, ..., N\}$ of numbers from 1 to N, in subscripts and in other places when it causes less clutter. Also, to harmonize our notation with what is customary for representing instance sizes in discussion

of algorithmic complexity, we denote the size of $[L]$, which is the number L as n. Using this notation, we define Δ, the ratio of longest to shortest link as: $\Delta := \frac{\max_{l\in[L]} \ell(l)}{\min_{l\in[L]} \ell(l)}$.

Length class: a length class or an almost equilength class is a set of links the lengths of which do not vary by more than a factor of two. As we will see, it is sometimes easier to divide a set of links into a number of length-classes and treat each class separately.

Normalized ISR and affectance: we first note that the Interference-to-Signal ratio (ISR) between a two links is

$$I'_{l'}(l) = \frac{p(l')/D(l',l)^\alpha}{p(l)/\ell(l)^\alpha}.$$

This can be generalized to ISR of a set S on a link defined as

$$I'_S(l) = \frac{\sum_{(l'\neq l)\in S} p(l')/D(l',l)^\alpha}{p(l)/\ell(l)^\alpha}.$$

Similarly for total ISR sent from a link to links on a set S, we have

$$I'_l(S) = \sum_{(l'\neq l)\in S} \frac{p(l)/\ell(l)^\alpha}{p(l')/\ell(l')^\alpha}.$$

The definition of $I'_{S_1}(S_2)$ is analogous. An advantage of this notion is additivity, that is $I'_{S_1\cup S_2}(l) = I'_{S_1}(l) + I'_{S_2}(l)$ for disjoint S_1 and S_2.

In order to normalize this metric with respect to β, we use the notion of affectance. This variation of ISR, first introduced in [52] and refined in [54] and [55] is defined as

$$\begin{aligned} I_{l'}(l) &:= \min\left(1, g_l \frac{p(l')/D(l',l)^\alpha}{p(l)/\ell(l)^\alpha}\right) \\ &= \min\left(1, g_l \frac{p(l')}{p(l)} \cdot \left(\frac{\ell(l)}{D(l',l)}\right)^\alpha\right), \end{aligned}$$

where the factor $g_l := \beta/(1-\beta N_0 \ell(l)^\alpha/p(l))$ depends only on properties of the link l and problem parameters. This can be analogously extended to affectance to and from sets as $I_l(S) := \sum_{l'\in S} I_l(l')$, $I_S(l) := \sum_{l'\in S} I_{l'}(l)$, and $I_S(T) := \sum_{l'\in S} I_{l'}(T)$ for sets S and T. It is

easy to see that a set S being an ISet is equivalent to $I_S(l) < 1$ for each $l \in S$.

As a first step in the scheduling algorithm, we pass from the geometric arrangement of nodes in a metric space to an abstract graph. The vertices of this graph correspond to network links and vertex connectivity is based on geometry of the network in a way that helps us build the approximation to the optimum link schedule. In order to go from the notion of an SINR-feasible ISet which defines a hypergraph on vertices representing links to a binary edge-connectivity relation between link pairs, we use a pairwise relaxation of the SINR feasibility of an ISet. What we mean by a relaxation is that if a subset of links forms an ISet, then all of its pairs of links will be disconnected in this relaxed notion. The reverse relationship, of a subset that is vertex-independent according to this binary relation being SINR-feasible and therefore an ISet, does not necessarily hold. This has the advantage of enabling us to use the conceptual simplicity of graph theory while still remaining close to the SINR formulation. We will see that the conflict graph built on this notion provides lower bounds on length of any SINR-feasible schedule. More importantly, the relaxed notion is still strong enough that it can be used advantageously in the design of our scheduling method and yet the conflict graphs built on it are structured enough to be efficiently colourable. This is the notion of q-independence defined next.

q-independence: parameterized by a number q, q-independence is a pairwise relaxation of the notion of SINR-feasibility. Formally, two links l' and l are q-independent if they satisfy $D(l', l) \cdot D(l, l') \geq q^2 \cdot \ell(l)\ell(l')$. This is equivalent to $I_l(l') \cdot I_{l'}(l) \leq \frac{g_l g_{l'}}{q^{2\alpha}}$, independent of the power scheme used. If two links are not q-independent, they are called q-dependent or q-adjacent since they are adjacent in the graphic representation of this relation. Likewise, a set of links is called q-independent if all pairs of its links are q-independent and its called q-dependent otherwise.

The q-dependence graph is defined by having a vertex for each link and having an edge between two vertices whenever they are q-dependent.

From the above definition, it is easy to see that if two links are q-dependent they cannot

be simultaneously active in a $q^\alpha\beta$-SINR feasible ISet (that is an ISet that satisfies the more stringent $q^\alpha\beta$ SINR criterion), but the reverse will not necessarily be true as interference power is accumulative.

The q-dependence graph of $[L]$ is represented by $G_q([L])$. Recall that this graph has vertices corresponding to all links in $[L]$ and an edge between two vertices whenever their respective links are q-dependent.

The top level of the proceeds by first colouring $G_q([L])$. This is done by a greedy subroutine which takes linear time in the number of edges and vertices of the graph. It produces a vertex-colouring that, using a special property of $G_q([L])$, is shown to use at most a constant factor more colours than the its chromatic number, which is by definition, the minimum number of colours required for a vertex colouring of this graph.

Since q-independence is, as noted, a relaxation of the notion of SINR-feasibility of ISets, this colouring can serve as a rough first step in putting links whose simultaneous transmission directly leads to SINR conflicts into different ISets.

The next step considers each monochromatic set, which is by definition a q-independent subset of $[L]$, separately when allocating ISets to avoid obvious pairwise conflicts between q-dependent links. In order to build an ISet partition for each such q-independent monochromatic subset, we will have to define a few more refined properties that allow bounding of sum-interference from other links when putting a link into a tentative ISet. The next few definitions explicitly define these properties between pairs or subsets of links. The roles these conditions play and the way they are satisfied at various stages of the algorithm is clarified in the next section. To motivate the definitions though, they are accompanied by a short description of how they help in different stages of the scheduling.

τ-interfering links: two links l, l' are said to be τ-interfering if $\max\{I_l(l'), I_{l'}(l)\} > \tau$, that is at least one is causing a normalized ISR of more than τ on the other. Later, when we try to assign links to different ISets (timeslots) using a greedy binning approach, we use properties of this definition with an appropriately defined value of τ to show that the chosen

number of bins is enough.

Lengthwise well-separated link set: a set of links is lengthwise well-separated if for every link l, and all longer links l', we either have that $\ell(l') < 2\ell(l)$ or $\ell(l') > 2.(4n)^{\frac{2}{\alpha}}\ell(l)$. Informally, what this says of a set is that the length relationship between any pair of links in it is such that they are either almost equilength or else have widely differing lengths. Noting that the length ratio in the second case is a growing function of instance size n, this notion helps in making an argument based on pigeonhole principle that the total interference from certain types of links on others is bounded.

p-greedily binnable link-set: parameterized by a natural number p, a subset $A \subseteq [L]$ is said to be p-greedily binnable if for every link in $[L]$, there are at most p-links in the set that are both $2.(4n)^{\frac{2}{\alpha}}$-times longer than it and $\frac{1}{2n}$-interfering with it. This is where the τ-interfering property mentioned earlier is used. When a group of subsets have this property, their links can be aggregated in $p+1$ bins corresponding to different ISets in a greedy manner while assuring that the total interference any link receives in its assigned ISet is bounded.

4.4.2 Description of the algorithm, its correctness and performance

In this section, using the definitions just provided, we state our algorithm and characterize its correctness and approximation ratio using a series of lemmas. During its operation, the top-level algorithm invokes a greedy graph colouring algorithm and a special-case link scheduling algorithm (which requires its input instance to satisfy certain conditions) as subroutines. When discussing each of these algorithms, a compact description the steps of the algorithm is given first, then the operation of the algorithm is discussed step-by-step together with lemmas that show correctness and bounds on the approximation ratio. For each lemma, a high-level overview of why it holds and the role it plays in showing the overall correctness and performance of the algorithm is given, but longer derivations and proofs are relegated to Appendix B. As stated previously, the power control scheme assigns power proportional

to $\ell(l)^{\alpha/2}$ due to the advantages this choice offers compared to both uniform power and complete channel inversion. Nevertheless, in one particular stage of the algorithm, when considering links within a length class (where, as we saw, lengths differ within a factor of at most 2), we will state and use a lemma about uniform transmission powers and use the fact that since the lengths do not differ by a factor more than 2, uniform and $\ell(l)^{\alpha/2}$-power are in a certain sense, very similar.

As one last note, the ultimate goal of this algorithm is to partition a set of transmission requests into a set of ISets feasible with SINR β so that members of different ISets can be active over successive timeslots. At the same time, we are trying to make the cardinality of this partition as close to the minimum cardinality as possible. In what follows, therefore, we use the terms timeslots and ISets interchangeably. In principle, different ISets could be assigned to different frequency bands, which would make the objective equivalent to minimizing the number of frequency channels, but this is not relevant to the structure of the problem.

Algorithm 1 is the top-level algorithm that we use for scheduling a set of links:

Algorithm 1 Scheduling arbitrary sets of links.

---

1. Construct the q-dependence graph for $q = 3$, $G_3([L])$ on the set of links $[L]$.
   - We have the graph $G_3([L])$ at this stage.
2. Colour $G_3([L])$ using Hochbaum's greedy colouring algorithm.
   - After this step, we have $[L]$ split into 3-independent subsets $L_1, L_2, \ldots, L_k$ where k is at most $c\chi(G_3([L])$ by Theorem 4.1.
3. For $i = 1, 2, \ldots, k$ apply Algorithm 3 to the set L_i.
   - After this step, we get for each i, a partition of L_i into a collection of ISets $\Sigma_i = \{S_i^1, S_i^2, \ldots, S_i^{k_i}\}$, where k_i is the cardinality of the collection.
4. The output schedule is the set of all ISets in collections calculated for different L_i's: $\cup_i \Sigma_i$

The flowchart of this algorithm is shown in Figure 4.4.2. The algorithm starts by con-

4.4. Proposed approximate algorithm

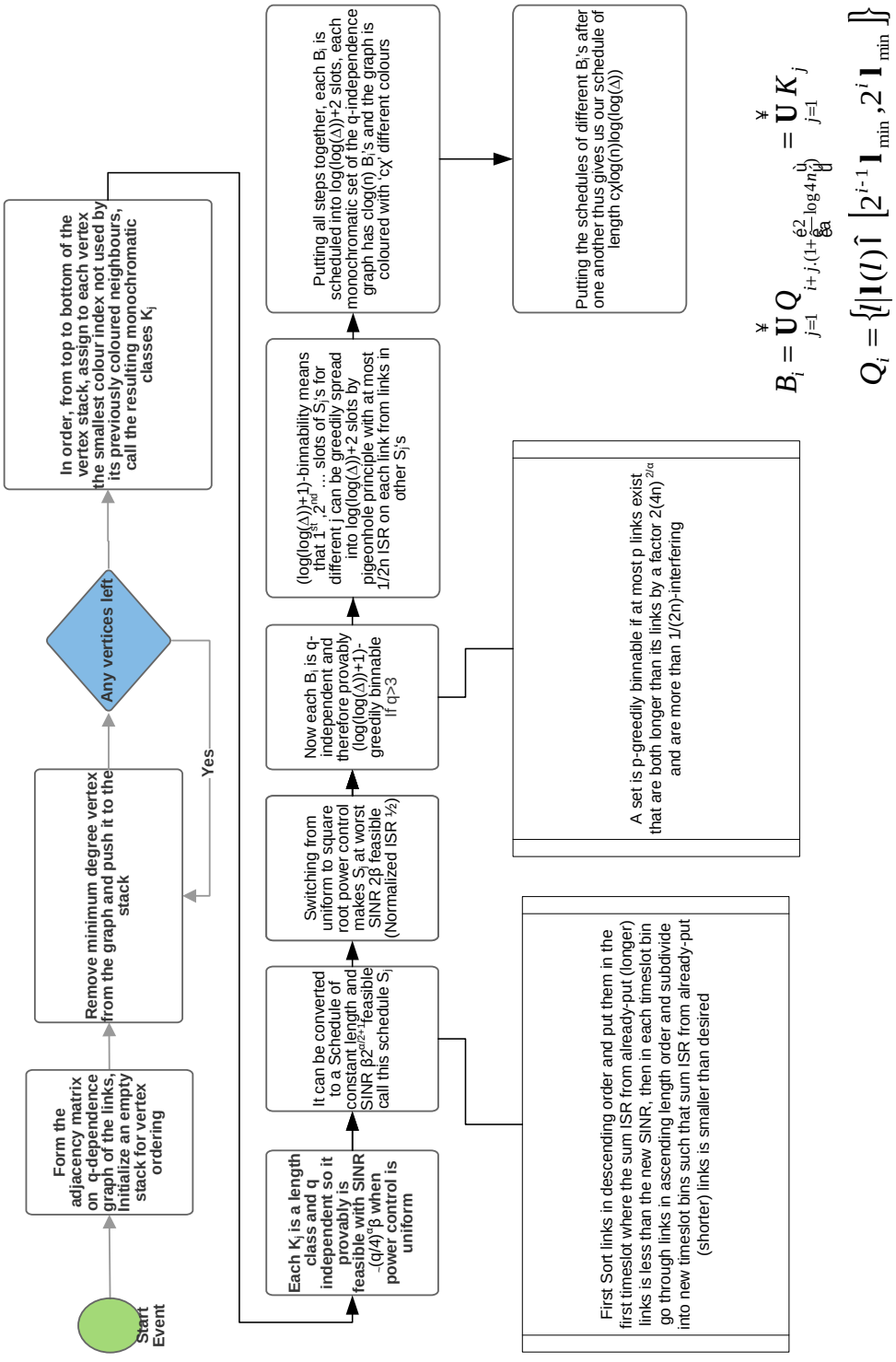

Figure 4.1: Flowchart for Algorithm 1.

structing the q-independence graph. As can be seen, $q = 3$ is chosen in step 1 above. This choice is justified by the fact that it leads to subsets that satisfy the input conditions of Algorithm 3 invocation in step 3 as will be discussed later. This step has worst case time complexity $\mathcal{O}(n^2)$ which is the maximum number of edges of the graph.

Step 2 uses a greedy algorithm to vertex-colour $G_3([L])$. The greedy colouring algorithm used is due to Hochbaum [101]. There, it is shown to have a complexity of $\mathcal{O}(|V|+|E|)$ on a graph with vertex set V and edge set E, which corresponds to a worst-case time complexity of $\mathcal{O}(n^2)$ in our notation. This is Algorithm 2, called greedy Min-Degree-Last and the following lemma characterizes its performance for q-independence graphs.

Algorithm 2 Greedy Min-Degree-Last vertex colouring

---

input: Graph G.
Output: A vertex colouring of graph G, with colours indexed by integers 1 and above.

1. Obtain the total order Min-degree-last on graph G as follows:
   (a) Initialize an empty stack for the vertices.
   (b) Select v a vertex of minimum degree in G.
   (c) Remove v and its incident edges from G, push v onto the stack.
   (d) Go to step (b) and continue removing until all vertices are removed.
   (e) After the above steps are finished, the top-to-bottom order of vertices on the stack is the desired total order, inductively defined such that each vertex is has the minimum degree in the graph induced by itself and vertices coming before it in the order.
2. Go through vertices in Min-degree-last ordering (so the last vertex remaining in the previous step comes first here) and colour each vertex with the smallest colour not used by any of its adjacent predecessors.

Lemma 4.1. *Algorithm 2 is a greedy colouring algorithm that colours any q-independence graph G_q using $c\chi(G_q)$ colours for c a constant.*

Proof: The proof is based on showing that the neighbourhood of each vertex in G_q has a bounded maximum independent set size. To show this, the geometric structure of the connectivity in G_q is used. The details are in Appendix B.

4.4. Proposed approximate algorithm

Step 3 of the Algorithm 1 relies on Algorithm 3. Invocation of this Algorithm is where most of the scheduling work happens. We defined the notion of p-greedily-binnable subsets previously and briefly discussed how this property is useful by allowing greedy aggregation of links from different p-greedily-binnable subsets into a set of $p+1$ separate bins. Algorithm 3 requires the set of links at its input be q-independent and p-greedily binnable. Under these conditions, the output of this algorithm will be a collection of ISets of cardinality $\mathcal{O}(p \log n)$.

Algorithm 3 Scheduling q-independent and p-greedily binnable sets of links.

Input: A q-independent p-greedily binnable set Q, for some $p > 0$ and $q \geq 1$.
Output: A partition of Q into $\mathcal{O}(p \log(n))$ ISets.

1. Let $Q = \cup_i Q_i$, where $Q_i = \{l \in Q | \ell(l) \in [2^{i-1}\ell_{\min}, 2^i \ell_{\min})\}$, for $i = 1, 2, \ldots, i_{\max} \leq \lceil \log(\Delta) \rceil$.
2. Assign $B_i = \cup_i Q_{i+j \cdot (1+\lceil \frac{2}{\alpha} \log 4n \rceil)}$, for $1 \leq i < 1 + \lceil \frac{2}{\alpha} \log 4n) \rceil$.
3. Schedule each $B_i = \cup_j K_j$, where $K_j := Q_{k+j \cdot (1+\lceil \frac{2}{\alpha} \log 4n) \rceil}$, using the steps that follow:
   (a) Using the SINR-strengthening lemma from Theorem 4.3 transform each K_j into an e-SINR feasible partition of ISets: $\Sigma_j = \{S_j^s\}_{s=1}^{k_j}$ under uniform power with $e = 2^{\alpha/2+1}\beta$.
      - After this step we have a collection of subsets of K_j, each of which is an ISet feasible with SINR $2^{\alpha/2+1}\beta$ under uniform power. Since, maximal pairwise length ratio is 2, changing power control from uniform to $\ell(l)^{\alpha/2}$ means that each ISet is at the worst 2β-SINR feasible.
   (b) $s \leftarrow 1$.
   (c) Assign $S \leftarrow \cup_j S_j^s$: if for some j, $k_j < s$, then we take $S_j^s = \emptyset$.
   (d) Sort S in the non-increasing order of link lengths: $\ell_1 \geq \ell_2 \geq \ldots \ell_{|S|}$.
   (e) $T_s^r \leftarrow \emptyset, r = 1, 2, \ldots, p+1$.
   (f) For $k = 1, 2, \ldots, |S|$ do: find a T_s^r not containing links u with $\ell_u > 2.(4n)^{\frac{2}{\alpha}} \ell_k$ which are $1/(2n)$-interfering with k, and assign $T_s^r \leftarrow T_s^r \cup \{k\}$.
   (g) $s \leftarrow s + 1$: if $s \leq \max k_j$, then go to step 3.(c), otherwise the schedule for B_i is $\{T_s^r | T_s^r \neq \emptyset\}$.
4. Output the union of the schedules of all B_i.

The flowchart of Algorithm 3 is shown in Figure 4.4.2. Step 1 of this algorithm decomposes the input set of links into a series length classes, sets where maximum link length ratio

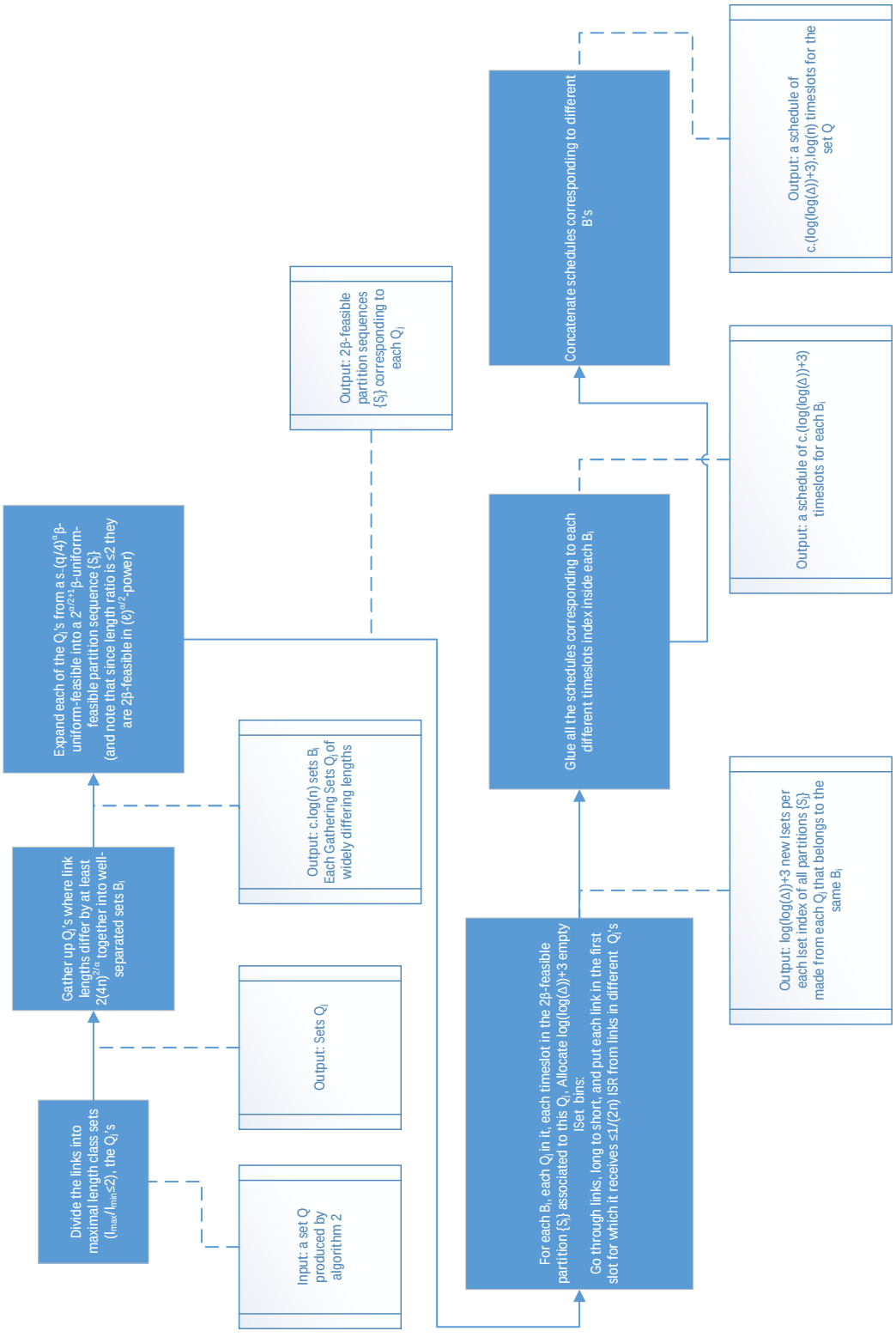

Figure 4.2: Flowchart for Algorithm 3.

is at most 2, namely the Q_i's. Step 2 then takes the union of Q_i's whose indices differ by an integer multiple of $1 + \left\lceil \log\left((4n)^{\frac{2}{\alpha}}\right)\right\rceil$ into the maximal well-separated sets B_i. Recall that a well-separated set is a set of links where for every pair the length ratio is either less than 2 or more than $2.(4n)^{\frac{2}{\alpha}}$, so step 2 forms the maximal well-separated subsets of Q in the most straightforward way. It is also simple to see that exactly $1 + \left\lceil \log\left((4n)^{\frac{2}{\alpha}}\right)\right\rceil$ sets B_i are formed at the end of this step containing all the links in Q. When discussing the operations of step 3, it is useful to denote the j'th subset in the union defining B_i, that is the set $Q_{i+j.(1+\lceil \log((4n)^{\frac{2}{\alpha}})\rceil)}$ by K_j, with its dependence on i being implicit.

Step 3.(a) refines each set $K_j := Q_{i+j.\left(1+\left\lceil \log\left((4n)^{\frac{2}{\alpha}}\right)\right\rceil\right)}$ into a collection of its subsets that are SINR feasible ISets for SINR level $2^{\alpha/2+1}\beta$. The correctness of this step relies on two lemmas.

The first lemma and the only place that we rely on uniform power in this derivation is a result that shows the one-way implication from SINR-feasibility to q-Independence, which results from the latter being a relaxation of the former, has a restricted converse. More precisely, if a set of links is a length class, has maximal pairwise link length ratio at most 2, and all pairs of links are q-independent, It will be SINR-feasible for a certain SINR level $k\beta$ $(k > 1)$ with uniform power control. The sets K_j are by definition q-independent and length class subsets and satisfy this condition. The following lemma states this restricted equivalence between SINR-feasibility and q-independence.

Lemma 4.2. *Any almost equilength class of links in a doubling metric space that is q-independent is also SINR feasible under uniform power assignment.*

Proof: To bound the total interference to a given receiver, this proof accounts for the interference from transmitters of other links located inside different concentric rings around this receiver separately. An argument based on the geometric definition of q-independence, together with all pairs of links being almost equilength, puts an upper bound on the number of transmitters in each ring using the doubling property of the metric. This leads to a bound on total interference from transmitters in k'th concentric ring for each integer k. Summing

over all k gives a desired bound on total interference that implies the link SINR criterion is met. The full details are in Appendix B

The next ingredient in showing the correctness of step 3.(a) is noticing that different SINR levels are interchangeable in the sense that an ISet that is feasible for a certain SINR level can be refined into a collection of a constant number of its subsets, each of which are feasible for a given higher target SINR level. This is shown by the following *SINR-strengthening* result:

Lemma 4.3. *[SINR strengthening] Any collection of S-SINR feasible ISets can be refined into a collection S'-SINR feasible subsets ($S' > S$), increasing the cardinality of the collection by at most $\left\lceil \left(\frac{2S'}{S}\right) \right\rceil^2$.*

Proof: The proof is rather simple and proceeds by judiciously dividing each ISet of the original collection into a number of ISets in a two-step greedy process, such that the new ISets are feasible for the higher SINR level. The complete proof can be found in Appendix B.

Application of the previous lemma in step 3.(a) decomposes each K_j into a constant number of its subsets S_j^s's, each of which is an ISet with SINR level $2^{\alpha/2+1}\beta$ under uniform power. Since the pairwise link length ratios in these ISets is upper-bounded by 2, changing power control from uniform to $\ell(l)^{\alpha/2}$ makes the ISets S_j^s feasible for SINR 2β in the worst case.

Steps 3.(b) to 3.(g) depend on the assumption that the input set of links is p-greedily binnable. The following lemma states that if $q \geq 3$, this property holds for inputs given to Algorithm 3 by the top-level Algorithm 1. This is where the arbitrary-seeming value of 3 comes from in step 1 of that algorithm.

Lemma 4.4. *Any q-independent set of links for $q \geq 3$ is $\mathcal{O}(\log\log\Delta)$-greedily binnable.*

Proof: At a high level, this proof uses the geometric definition of q-independence and the triangle inequality in the metric space of nodes to establish a lower bound on the length ratio of two-links that are both at least $2.(4n)^{\frac{2}{\alpha}}$ times longer than a given link l, mutually

q-independent with each other, and more than $\frac{1}{2n}$-interfering with l. This bound implies a doubly exponential growth rate on the lengths of successively longer links of this type which is then used to upper bound their total number. Full details are in Appendix B.

For a given B_i of fixed index i, these steps (3.(b) to 3.(g)) take all links in the first ISet (or first timeslot), S_j^1, produced in step 3.(a) for K_j's with different j's (note that K_j's and therefore S_j^s's implicitly depend on i which is omitted here to avoid clutter). These links in $\bigcup_j S_j^1$ are put greedily in descending length order into the first one of the $p+1$ initially empty ISet bins where they receive less than $\frac{1}{2n}$ normalized ISR from links already scheduled in that bin. The same process is repeated again for the second ISets (or second timeslots), S_j^2, of all K_j's with $p+1$ new ISet bins and so on until the links corresponding to the highest partition index, $k_{\max} := \max_j k_j$, have been binned. Since by definition links from different K_j's for the same B_i are well-separated, when binning there are at most p links that are longer by more than $2.(4n)^{\frac{2}{\alpha}}$ and at least $\frac{1}{2n}$-interfering by the definition of p-greedily binnable property that the input set of links has to satisfy. This means that the greedy bin selection process always succeeds. Moreover, after all links are assigned to bins, each link is receiving at most $\frac{1}{2\beta}$ total ISR from links in the same S_j (because S_j was a feasible ISet for SINR 2β). But it also receives at most $\frac{1}{2n}$ normalized ISR from each link that is longer or shorter by more than a $2.(4n)^{\frac{2}{\alpha}}$ factor. To see why this is true, note that by construction, the shorter of two links is added to their common bin only if both the ISR it receives from or puts on the other link is bounded above by $\frac{1}{2n}$. This crucially depends on the presence of max in the definition of τ-interfering property. It also deserves noting that in the worst case, finding a suitable bin in this step may add a total of $\mathcal{O}(n^2)$ time steps to the algorithm runtime since each link might need checking against $\mathcal{O}(n)$ other links until a suitable bin is found.

The previous discussion, combined with Lemma 4.2 and Lemma 4.3 as summarized by the following theorem, establish that Algorithm 3 works according to its input-output specification and schedules a p-greedily-binnable and q-independent set in $\mathcal{O}(p\log(n))$ steps.

Theorem 4.1. *Algorithm 3 schedules any p-greedily binnable set of links in $\mathcal{O}(p \log n)$ slots*

Proof: The proof follows from the operation of steps 2 and 3.(a) to 3.(g) of the algorithm together with the aforementioned lemmas establishing correctness conditions at various steps. Appendix B contains the complete proof.

Combining this with Lemma 4.4 means that steps 3 of the top-level Algorithm 1 partitions each monochromatic, and therefore 3-independent, component of $G_3([L])$ into $\mathcal{O}(\log(n)\log(\log(\Delta)))$ ISets. By Theorem 4.1, step 2 of this algorithm outputs at most $c\chi(G_3([L]))$ of these q-independent sets. Therefore, Algorithm 1 produces a $\mathcal{O}(\log(n)\log(\log(\Delta))\chi(G_3([L])))$-ISet partitioning of $[L]$.

To show the approximation ratio of the algorithm, the $\mathcal{O}(\log(n)\log(\log(\Delta))\chi(G_3([L])$ output length should be connected with the minimum β-SINR schedule length. To do this, we first note that since feasibility for $3^\alpha\beta$-SINR is by definition a stronger condition that 3-independence, any partition into ISets with $3^\alpha\beta$-SINR level is at least as large as the minimal vertex colouring of the graph $G_3([L])$. This means $\chi(G_3([L]))$ is a lower bound on the minimum-length schedule with $3^\alpha\beta$ SINR level. Lemma 4.3 means that this schedule is itself within a constant factor of length from the minimal β-SINR feasible schedule (the latter schedule can be converted to an at most constant-factor longer $3^\alpha\beta$-schedule which is by definition longer than the *minimum-length* $3^\alpha\beta$-SINR schedule). These two facts together mean that the best β-SINR schedule is within a constant factor of $\chi(G_3([L]))$. Therefore, the $\mathcal{O}(\log(n)\log(\log(\Delta))\chi(G_3([L])))$-long schedule output by Algorithm 1 is an $\mathcal{O}(\log(n)\log(\log(\Delta)))$-factor approximation for the minimum-length scheduling problem.

Theorem 4.2. *Using Algorithm 1 gives $\mathcal{O}(\log n \log\log \Delta)$-factor approximation to the problem of link-scheduling with $\ell^{\alpha/2}$ power control.*

Proof: The proof follows from the previous discussion using Theorem 4.1 together with Lemma 4.4. A concise writeup is given in Appendix B.

This establishes the correctness of Algorithm 1 and its approximation ratio. As discussed previously in Chapter 2, the reduction of SINR-based scheduling to a graph-theoretical model

used in our algorithm has major differences from approaches previously used in the literature. Having described the operation of the algorithm and its correctness and approximation ratio, we are in a position here to elaborate in greater on the differences of our approach from the existing graph-based methods for SINR-feasible scheduling and the significance of the choices we made in its design. In particular, we describe how our algorithm trades conceptually simple operation and a geometrically more faithful graph model for a more refined network graph with more involved algorithmic and analysis steps but better performance and reference lemmas that highlight what has to be done differently in our algorithm and how our choices confer desirable performance properties. The graph representation that we build, while defined using pairwise geometric distances between network links, is abstracted away from the global geometric structure imposed by the underlying metric space in that it is not necessarily a geometric graph embeddable in this metric space. This contrasts with previous work (for example the work of [51]) where the deployment area is divided into varying-sized square cells and only links that are close in length and whose cells are coloured differently can be simultaneously active. Our graph $G_q([L])$, has network links rather than square cells of the plane as vertices. Its vertex connectivity relation does not admit a simple geometric notion such as a disk radius in a disk graph or cell adjacency in a plane cell-decomposition graph. This means that we cannot use the planar graph colourability results to show that colouring our graph is easy. Instead, we have to painstakingly show that this graph still has enough structure to allow efficient and near-optimal colouring by a simple greedy algorithm (as is shown in Lemma 4.1). In return for this extra complexity, the connectivity relation of our graph relates to SINR-feasibility much more closely than simpler geometric notions (as shown by Lemma 4.2). Concretely, this means that our algorithm does not a priori exclude links from being scheduled together only because they have differing lengths or that they are physically close to one another. The price that we pay for this is the added complexity of analyzing efficient colourability. Also, the sequence of algorithmic steps we use for building up the link schedule from the colouring is slightly more complicated than previous approaches.

The upshot is that this algorithm has more of a global outlook on the set of links and does not suffer from the unnecessary slack that results from a priori separating spatially close or lengthwise far links from each other and that manifests itself in large constant factors that has made previous graph-theoretic SINR-based scheduling algorithms impractical [57]. In the next section, we report on numerical experiments of the performance of our algorithm.

4.5 Simulations and conclusion

In this section, we perform numerical performance experiments to observe how the proposed algorithm compares with the alternatives. First, the methodology and parameters of the simulations are briefly discussed. Next, the feasibility-problem variation of the MIP formulation is used as a benchmark to see how far from optimum the algorithm is for intermediate problem sizes. The section after that will compare the throughput performance of the algorithm in the large-network regime, where exact algorithms will be intractable, with the iterative algorithms FlashLinQ and ITLinQ

4.5.1 Setup and choice of parameters

Before discussing the numerical results, the setup and choice of parameters for simulations is discussed. The parameters chosen are similar to that used in [60, 61] for FlashLinQ and ITLinQ algorithms and represent typical radio equipment and propagation characteristics of an outdoor environment. We assume our transmitters and receivers operate in the 2.4 Ghz ISM[6] band. The receiver front-end is assumed to have a noise figure of 7 dB[7]. This is consistent with the characteristics of 2.4 Ghz band transceivers [102, 103] where the low-

[6]Industrial, scientific, medical bands are a set of radio frequency bands set aside for unlicensed operation by International Telecommunication union (ITU).

[7]Noise figure of a receiver is the ratio of SNR in its output to SNR in its input. It quantifies the amount of extra thermal noise added by receiver's internal circuitry compared to the noise that is received at the antenna input (which is typically attributed to cosmic microwave background and transmitter noise). When expressed in decibels, noise figure values should be subtracted from receiver input SNR before calculating bit error rate and other receiver characteristics.

noise amplifier (LNA) has a typical noise figure of around 4.5 dB and the whole chain of LNA, down-converter and demodulator have a total noise figure of around 6.5 to 7.5 dB. Also, we assume transmitter power to be 20 dBm for a 1 metre link and that antenna gain relative to isotropic transmission be 2.5 dBi[8]. A dipole antenna has a gain of 2.15 dBi and patch antennas may reach up to 3.5 dBi [104], so 2.5 dBi is a good middle of the range value. To model the path loss parameters of the propagation environment, we use the model from ITU-T recommendation 1411 [105], which is compendium of experimentally validated radio channel models for the 900 Mhz-100 Ghz frequency range. We use the line of sight model and use 1.5 metres as both user equipment and base station antenna height since the model is intended for a peer-to-peer setup. What this amounts to in our setup at 2.5 Ghz centre frequency is a path loss exponent of 3 together with a path loss of 68 dB at a distance of 72 metres. We assume a channel bandwidth of 5 Mhz which together with the noise spectral density of -184 dBm/Hz gives a noise variance of around -115 dBm. We use a SINR threshold, β, of 15 dB that is typical for bit error rate (BER) performance of 10^{-4} in a bandwidth-efficient modulation scheme such as 64QAM [106]. We disperse random transmitters in a square area of 1000 metre $\times$ 1000 metre size and the receivers are put in a random direction at a uniformly random distance between 2 and 74 metres from their designated transmitter.

4.5.2 Comparison with exact solution algorithms

In section 4.3 the exact scheduling problem was formulated as a mixed integer program in two slightly different ways. In experimenting numerically, it was discovered that for the MIP formulation in Equation 4.9, with instance sizes of more than about 50 links and for the MIP formulation in Equation 4.10, with instance sizes of more than about 30 links, they will take extremely long to produce a solution[9]. In Section 4.3.1, the feasibility-problem formulation

[8]The gain of an antenna is used to quantify how much better it is in directing electromagnetic energy than an isotropic radiator. It is a measure of antenna directivity and is typically measured in dBi or decibels relative to isotropic.

[9]Not converging to a solution using CPLEX after around 5×10^8 branch and bound iterations which correspond to runtime of 5-6 days using about 20% of the cores of a 48-core Intel Xeon server with 256GB

of Equation 4.7 was given to iteratively refine bounds around the exact minimal schedule length. Further experimentation with this latter formulation showed that for medium-sized networks of up to around 250 links, it can be used to efficiently confine the exact solution to a relatively small interval. This formulation suffers from an increase in runtime for larger instances as well, but this is typically observed when the endpoints of the interval are pushed closer together (and thereby to the exact solution). We use this program, reproduced below, in this section to compare the schedule length produced by our algorithm:

$$
\begin{aligned}
&\min_{\{x_{ls}\}} \quad 2 \\
&\text{s.t.} \qquad x_{ls} \;\in\; \{0,1\} && \forall l \in [L], \forall s \leq s_0 \\
&\qquad \sum_{s=1}^{s_0} x_{ls} \;=\; 1 && \forall l \in \{1,..,L\} \\
&\qquad x_{ls} G(l) p(l) \;\geq\; x_{ls}\beta N_0 + \sum_{\substack{l' \neq l \\ l' \in s}} x_{l's} p(l') G(l',l) - M(1-x_{ls}) && \forall l \in \{1,..,L\}, \forall s \in \{1,...,s_0\}.
\end{aligned}
$$

Simulations were performed for 100, 150, 200 and 250 links. For each data-point, 5 random realizations of the networks were generated as discussed (with transmitters in a 1000 Metre $\times$ 1000 Metre square and receivers distanced randomly from 4 to 72 metres from their designated transmitter) to average out the variation due to randomness. Transmitter, receiver and channel parameters were selected as discussed above. Repeated iterations of the above feasibility problem were performed until tightening the upper and lower range would increase the runtime by a factor of more than 1000. The simulations were performed using approximately 20% of the cores of an Intel Xeon server with 48 cores and 256 GB of RAM. The aggregate time taken by all simulations was about 25 days with individual instances running for around two days at the most. As is expected from our approximate scheduling algorithm being suboptimal, it produces an output that in the case of 100 links is completely outside the interval produced by the MIP. For other values, the output is within the interval containing the exact solution but still not optimal. The graph also shows the beginning of the

of RAM.

trend toward better approximating the optimal schedule as the number of links increases.

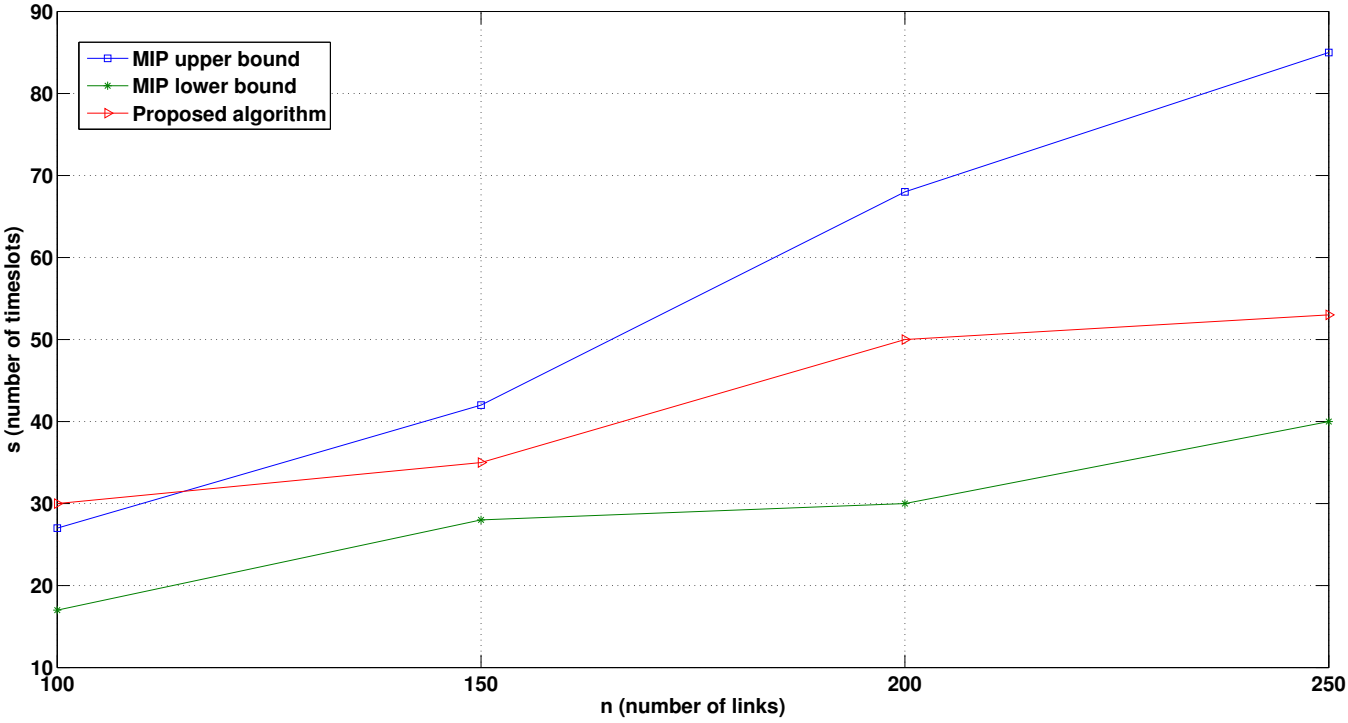

Figure 4.3: Output schedule length of proposed algorithm compared to the bounds obtained by mixed integer programming

4.5.3 Throughput performance in large-network scenario

This section compares the throughput performance of the algorithm with two heuristic link scheduling methods proposed in the literature for large networks. The algorithms are Flash-LinQ and ITLinQ. These were discussed earlier and represent the state of the art for device-to-device network approximate scheduling. Both of these algorithms have worst-case time-complexity of $\mathcal{O}(n^2)$ similar to our algorithm as they also might have to check $\mathcal{O}(n^2)$ link pairs against each other in the worst case. The main difference between the three algorithms is in the pairwise criteria used to allow a link to be added to a tentative Iset of the schedule. The algorithm proposed here is the only one that rigorously relates its performance to the shortest possible schedule in the general case. ITLinQ has the property that treating interference as noise is GDoF optimal for its chosen Isets which under restricted conditions

can be related to it giving an approximation to the optimum schedule. FlashLinQ does not give any theoretical guarantees but it is an end-to-end system implemented on top of 802.11 OFDMA[10] physical layer and therefore much more complicated to analytically study.

For 10 iterations, n transmitters were randomly dispersed in a $1\,\text{Km} \times 1\,\text{Km}$ area with the corresponding receiver placed randomly between 2 and 74 metres away in a uniformly random direction. Transmitter, receiver and channel parameters were selected as discussed above. Figure 4.4 shows the total achievable sum rate of ITLinQ, FlashLinQ, no scheduling and our scheme as a function of n. The algorithms were run side by side on an Apple Macbook pro laptop with an Intel Ivy-bridge core i7 CPU and 16 GB of RAM. The simulations took an aggregate time of about 30 minutes to complete. Individually our algorithm runs at about the same speed as ITLinQ and about 4 times slower than FlashLinQ. The data corresponding to no-scheduling shows the collapse in throughput as interference increases. The results seem to confirm the theoretical advantages of the proposed method compared to other approximation schemes as instance sizes increase. It is easy to see that asymptotically, the algorithm proposed here compares very well in the achieved sum-rate with both FlashLinQ and ITLinQ, specially as the size of the network increases. As discussed in the beginning of the chapter and background part of the thesis, we think that with the advent of very large-scale M2M and D2D deployments, these large network scenarios will become increasingly relevant and it is therefore important to investigate link scheduling methods that offer provable scaling characteristics for large number of users.

[10]Orthogonal Frequency Division Multiple-Access.

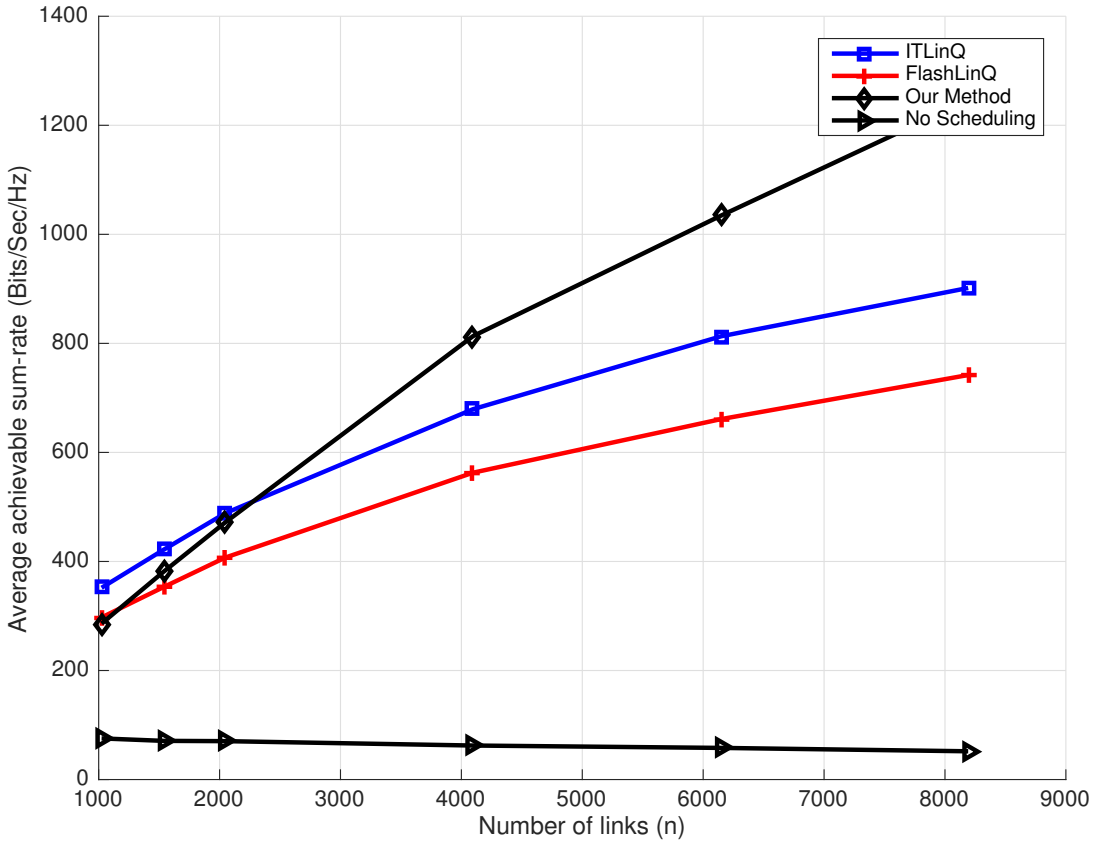

Figure 4.4: Sum-rate comparison of our algorithm with FlashLinQ, ITLinQ and no scheduling

Chapter 5

Summary of contributions and future work

In this chapter, we first provide a summary of the work described in this thesis. We then discusses some of the limitations od our models and approaches to the problems discussed and compare and contrast our work with similar work and state of the art. We then explicitly go over our contributions and finally conclude by discussing future directions.

5.1 Summary of contributions

The contributions of this thesis can be summarized into two parts, corresponding to the two-user channel model and many-user interference networks.

In Chapter 3, a two-user cognitive channel model based on Gaussian interference channel was proposed. The best achievable rate regions via single and multi-layer coding for this channel were characterized through various alternative decoder and encoder designs. The secondary user's rate optimization problem was analyzed and characterized through case by case analysis of the relevant rate expressions. Numerical examples were also provided.

In Chapter 4, an approximate SINR-based wireless scheduling algorithm for large networks was proposed. The essence of the algorithm was based on a graph-based representation

simplifying the SINR-feasibility criterion. Correctness and approximation ratio of this algorithm was analytically established and its performance simulated.

5.2 Limitations

For the model of Chapter 3, to the best of our knowledge, this is the first time this or a similar model has been studied. One limitation with our approach is that the problem setup restricts us to rather simple encoding and decoding schemes due to lack of collaboration from the primary user.

For the model of chapter 4, our model assumes a single SINR threshold to be maintained for all nodes whereas it might be reasonable to consider different SINR requirements for different types of RF front-end equipment.

5.3 Comparison

Our work in Chapter 4, compared to FlashlinQ, which is based on the multi-tone structure of IEEE 802.11 OFDMA physical layer, we study an abstract single-channel model. This simplified setup captures the problem and allows us to prove formal guarantees about the performance of our method, but is limited in scope compared to their work which included hardware implementation and field tests. Compared to ITlinQ, we have a more structured metric assumption on node placement and propagation environment in our model and also a proscribed SINR to be maintained at all links. Their method works by assigning time-slots to subsets of links that are GDoF-optimal under treating-interference-as-noise. Our algorithm does not guarantee GDoF-optimality in each timeslot but we build our schedule in a way that can be shown to be close to the best schedule achieving a given SINR.

Compared to exact formulation of scheduling based on mixed integer programming, our algorithm is suboptimal as it only produces an approximation to the shortest schedule. On the other hand, it can scale to networks of up to thousands of nodes while mixed integer

programming becomes intractable for more than a few hundred links.

5.4 Future Work

One way to extend the model in Chapter 3 is to look at the problem in a cooperative context where one user has much lower data rate compared to the other (for example control data compared to payload data) and relate it to the corner point problem of Gaussian interference channel. This is the maximum rate of the secondary where the required back-off from the optimum point to point rate of the primary user is exactly zero. What we are after therefore is the supremum of rates R_ϵ where the secondary user can send R_ϵ bits per channel use without affecting the primary at all. Namely

$$\inf_{(C_1-\delta(R_\epsilon),R_\epsilon)\in\mathcal{C}_{GIC,P_1,P_2,\sqrt{a_{12}},\sqrt{a_{21}}}} \delta(R_\epsilon)$$

This corner point problem has a long history. It was first studied by [107] by Costa, where he proved rate expressions for the corner points of Gaussian interference channel under some conditions on channel coefficients. Later, Sason in [108] found an error in the proof of [107] but conjectured the expressions to be correct. This led to naming these as the "Missing corner points" ([109, 110, 111]). Sason has recently also studied the achievable rates in the high SNR limit in [112]. Even more recently, Costa and Rioul [113, 114] have shown the equivalence between correctness of the conjectured corner points and a conjectured differential entropy inequality between what they define as "Almost-Gaussian" random variables where certain Markov relationships between them are "Almost lossless".

For the model of Chapter 4, one direction for future work is adding different SINR thresholds for different nodes in the network, perhaps by generalizing our relaxation to take account of different SINRs. This can be useful to model different sensitivity characteristics of radio hardware on different nodes. Another direction is adding fading to model signal strength variations. It has been shown in the literature [115, 116] that a closed form expression for

the cumulative distribution function of SINR can be given for Rayleigh fading. This can be used to generalize our scheduling method to a probabilistic one with precise performance bounds.

Bibliography

[1] J. Mitola, *Cognitive radio: An integrated agent architecture for software defined radio*. PhD thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2000.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[4] "IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput," *IEEE Std. 802.11n-2009*, pp. 1 –565, 2009.

[5] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th annual international conference on Mobile computing and networking*, pp. 53–64, ACM, 2012.

[6] W. Lee, I. Lee, J. S. Kwak, B.-C. Ihm, and S. Han, "Multi-bs mimo cooperation: challenges and practical solutions in 4g systems," *IEEE Wireless Communications*, vol. 19, no. 1, pp. 89–96, 2012.

[7] J. Thompson, X. Ge, H.-C. Wu, R. Irmer, H. Jiang, G. Fettweis, and S. Alamouti, "5g wireless communication systems: prospects and challenges [guest editorial]," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 62–64, 2014.

[8] O. El Ayach, S. W. Peters, and R. Heath, "The practical challenges of interference alignment," *IEEE Wireless Communications Journal*, vol. 20, no. 1, pp. 35–42, 2013.

[9] B. C. Jung and W.-Y. Shin, "Opportunistic interference alignment for interference-limited cellular tdd uplink," *IEEE Communications Letters*, vol. 15, no. 2, pp. 148–150, 2011.

[10] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.

[11] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Transactions on Information Theory,*, vol. 54, no. 8, pp. 3425–3441, 2008.

[12] H. Huang and V. K. Lau, "Partial interference alignment for-user mimo interference channels," *IEEE transactions on signal processing*, vol. 59, no. 10, pp. 4900–4908, 2011.

[13] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in mimo interference networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4771–4782, 2010.

[14] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob*, vol. 1, pp. 611–644, 1961.

[15] R. Ahlswede, *Multi-way communication channels*, pp. 23–51. Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2 - 8, 1971, Akadémiai Kiadó, 1973.

[16] H. H.-J. Liao, *Multiple Access Channels.* PhD thesis, 1972.

[17] H. Liao, "A coding theorem for multiple access communications," in *International Symposium on Information Theory, Asilomar CA*, 1972.

[18] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[19] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, 1973.

[20] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.

[21] P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 279–280, 1974.

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

[23] S. Gel, M. S. Pinsker, *et al.*, "Coding for channel with random parameters," *Problems in Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

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

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

[26] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

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

[28] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.

[29] E. van der Meulen, *Transmission of Information in a T-terminal Discrete Memoryless Channel*. PhD thesis, Stanford University, 1968.

[30] E. Van Der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 1–37, 1977.

[31] T. Cover and A. El-Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[32] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *The Annals of Probability*, pp. 805–814, 1974.

[33] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[34] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, 2006.

[35] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," in *2006 IEEE International Symposium on Information Theory*, pp. 2413–2417, 2006.

[36] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4391–4399, 2007.

[37] I. Maric, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3536–3548, 2007.

[38] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4707–4712, 2008.

[39] X. Kang, R. Zhang, Y.-C. Liang, and H. K. Garg, "Optimal power allocation strategies for fading cognitive radio channels with primary user outage constraint," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 374–383, 2011.

[40] R. Zhang and J. M. Cioffi, "Iterative spectrum shaping with opportunistic multiuser detection," *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1680–1691, 2012.

[41] K. Moshksar and A. Khandani, "Randomized masking in cognitive radio networks," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2635–2647, 2013.

[42] 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, 1992.

[43] P. Chaporkar, K. Kar, X. Luo, and S. Sarkar, "Throughput and fairness guarantees through maximal scheduling in wireless networks," *Information Theory, IEEE Transactions on*, vol. 54, no. 2, pp. 572–594, 2008.

[44] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications societies (INFOCOM), 2005*, vol. 3, pp. 1804–1814, IEEE, 2005.

[45] G. Sharma, N. B. Shroff, and R. R. Mazumdar, "Maximum weighted matching with interference constraints," in *proceedings of Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, (PerCom), 2006.*, IEEE, 2006.

[46] G. Sharma, N. Shroff, and R. Mazumdar, "Maximum weighted matching with interference constraints," in *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, 2006.*, pp. 5 pp.–74, March 2006.

[47] K. Kar, S. Sarkar, A. Ghavami, and X. Luo, "Delay guarantees for throughput-optimal wireless link scheduling," *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2906–2911, 2012.

[48] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[49] S. Kompella, J. E. Wieselthier, and A. Ephremides, "Revisiting the optimal scheduling problem," in *CISS 2008. 42nd Annual Conference on Information Sciences and Systems, 2008.*, pp. 492–497, IEEE, 2008.

[50] T. Moscibroda and R. Wattenhofer, "The Complexity of Connectivity in Wireless Networks," in *25th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Barcelona, Spain*, April 2006.

[51] O. Goussevskaia, Y. A. Oswald, and R. Wattenhofer, "Complexity in geometric sinr," in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, pp. 100–109, ACM, 2007.

[52] O. Goussevskaia, R. Wattenhofer, M. M. Halldórsson, and E. Welzl, "Capacity of arbitrary wireless networks," in *INFOCOM 2009, IEEE*, pp. 1872–1880, IEEE, 2009.

[53] M. M. Halldórsson, "Wireless scheduling with power control," *ACM Transactions on Algorithms (TALG)*, vol. 9, no. 1, p. 7, 2012.

[54] M. M. Halldórsson and R. Wattenhofer, "Wireless communication is in apx," in *Automata, Languages and Programming*, pp. 525–536, Springer, 2009.

[55] T. Kesselheim and B. Vöcking, "Distributed contention resolution in wireless networks," in *Distributed Computing*, pp. 163–178, Springer, 2010.

[56] O. Goussevskaia, Y.-A. Pignolet, and R. Wattenhofer, *Efficiency of wireless networks: Approximation algorithms for the physical interference model*. Now Publishers Inc, 2010.

[57] X. Che, H. Zhang, and X. Ju, "The case for addressing the ordering effect in interference-limited wireless scheduling," *IEEE Transactions on Wireless Communications*, vol. 9, no. 13, pp. 5028–5042, 2014.

[58] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, 2012.

[59] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in lte-advanced networks," *IEEE Wireless Communications*, vol. 19, no. 3, p. 96, 2012.

[60] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic, "Flashlinq: A synchronous distributed scheduler for peer-to-peer ad hoc networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 4, pp. 1215–1228, 2013.

[61] N. Naderializadeh and A. S. Avestimehr, "Itlinq: A new approach for spectrum sharing in device-to-device communication systems," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1139–1151, 2014.

[62] C. Geng, N. Naderializadeh, A. S. Avestimehr, S. Jafar, *et al.*, "On the optimality of treating interference as noise," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1753–1767, 2015.

[63] R. Ash, *Information Theory*. Dover Books on Mathematics Series, Dover Publications, 1990.

[64] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[65] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[66] A. Carleial, "A Case Where Interference Does Not Reduce Capacity (Corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 569–570, 1975.

[67] M. H. Costa and A. El-Gamal, "The capacity region of the discrete memoryless interference channel with strong interference.," *IEEE Transactions on Information Theory*, vol. 33, no. 5, pp. 710–711, 1987.

[68] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 49–60, 1981.

[69] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the han–kobayashi region for theinterference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3188–3195, 2008.

[70] R. H. Etkin, D. N. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[71] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3032–3050, 2009.

[72] A. S. Motahari and A. K. Khandani, "Capacity bounds for the gaussian interference channel," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 620–643, 2009.

[73] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum–rate capacity for gaussian interference channels," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 689–699, 2009.

[74] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric gaussian user interference channel," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3297–3303, 2010.

[75] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over mimo x channels: Interference alignment, decomposition, and performance analysis," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3457–3470, 2008.

[76] B. Nazer, M. Gastpar, S. A. Jafar, and S. Vishwanath, "Ergodic interference alignment," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6355–6371, 2012.

[77] A. S. Motahari, S. Oveis-Gharan, M.-A. Maddah-Ali, and A. K. Khandani, "Real interference alignment: Exploiting the potential of single antenna systems," *Information Theory, IEEE Transactions on*, vol. 60, no. 8, pp. 4799–4810, 2014.

[78] S. A. Jafar, *Interference alignment: A new look at signal dimensions in a communication network*. Now Publishers Inc, 2011.

[79] S.-W. Jeon and M. Gastpar, "A survey on interference networks: Interference alignment and neutralization," *Entropy*, vol. 14, pp. 1842–1863, 2012.

[80] R. H. Etkin and E. Ordentlich, "The degrees-of-freedom of the-user gaussian interference channel is discontinuous at rational channel coefficients," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4932–4946, 2009.

[81] A. G. Davoodi and S. A. Jafar, "Settling conjectures on the collapse of degrees of freedom under finite precision csit," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, pp. 1667–1672, IEEE, 2014.

[82] J. C. Koo, W. Wu, and J. T. Gill III, "Delay-rate tradeoff for ergodic interference alignment in the gaussian case," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pp. 1069–1075, IEEE, 2010.

[83] O. Johnson, M. Aldridge, and R. Piechocki, "Delay-rate tradeoff in ergodic interference alignment," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 2626–2630, IEEE, 2012.

[84] L. Zhou and W. Yu, "On the capacity of the k-user cyclic gaussian interference channel," *Information Theory, IEEE Transactions on*, vol. 59, no. 1, pp. 154–165, 2013.

[85] X. Yi and G. Caire, "On the optimality of treating interference as noise: A combinatorial optimization perspective," *arXiv preprint arXiv:1504.00041*, 2015.

[86] C. Geng and S. A. Jafar, "On the optimality of treating interference as noise for k-user compound interference channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1706–1710, IEEE, 2015.

[87] S. A. Hesammohseni, K. Moshksar, and A. K. Khandani, "A combined underlay and interweave strategy for cognitive radios," in *2014 IEEE International Symposium on Information Theory*, pp. 1396–1400, IEEE, 2014.

[88] C. Nair and A. El Gamal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4479–4493, 2009.

[89] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1996.

[90] R. Descartes, *The Geometry of Ren Descartes: with a Facsimile of the First Edition*. Dover Publications, Inc., 1954.

[91] I. I. CPLEX, "V12. 1: Users manual for cplex," *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009.

[92] G. Inc., "Gurobi optimizer reference manual," *Available at http://www.gurobi.com*, 2014.

[93] S. A. Hesammohseni and M. O. Damen, "Provably near optimal link scheduling and power control for wireless device-to-device networks," in *2016 Australian Communications Theory Workshop (AusCTW)*, pp. 47–52, IEEE, 2016.

[94] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on selected areas in communications*, vol. 17, no. 7, pp. 1205–1211, 1999.

[95] A. Ephremides and T. V. Truong, "Scheduling broadcasts in multihop radio networks," *IEEE Transactions on communications*, vol. 38, no. 4, pp. 456–460, 1990.

[96] S. A. Borbash and A. Ephremides, "Wireless link scheduling with power control and sinr constraints," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5106–5111, 2006.

[97] A. Behzad and I. Rubin, "Optimum integrated link scheduling and power control for multihop wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 194–205, 2007.

[98] J. Clausen, "Branch and bound algorithms-principles and examples," *Department of Computer Science, University of Copenhagen*, pp. 1–30, 1999.

[99] G. Cornuéjols, "Valid inequalities for mixed integer linear programs," *Mathematical Programming*, vol. 112, no. 1, pp. 3–44, 2008.

[100] I. Griva, S. G. Nash, and A. Sofer, *Linear and nonlinear optimization*. Siam, 2009.

[101] D. S. Hochbaum, "Efficient bounds for the stable set, vertex cover and set packing problems," *Discrete Applied Mathematics*, vol. 6, no. 3, pp. 243–254, 1983.

[102] F. Qazi, Q.-T. Duong, and J. J. Dabrowski, "Wideband RF frontend design for flexible radio receivers," pp. 220–223, IEEE, 2011.

[103] Mobile Satellite Services (MSS), *COM-4102 2.4 GHz TRANSCEIVER Datasheet*, 3 2003.

[104] V. R. Gupta and N. Gupta, "Gain and bandwidth enhancement in compact microstrip antenna," *International Union of Radio Science, Proceedings*, 2005.

[105] P. Series, "Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 900 mhz to 100 ghz," *ITU-R Recommendation*, 2012.

[106] L.-L. Yang and L. Hanzo, "A recursive algorithm for the error probability evaluation of m-qam," *IEEE Communications Letters*, vol. 4, no. 10, pp. 304–306, 2000.

[107] M. H. M. Costa, "On the gaussian interference channel," *IEEE Transactions on Information Theory*, vol. 31, no. 5, pp. 607–615, 1985.

[108] I. Sason, "On achievable rate regions for the gaussian interference channel," *Information Theory, IEEE Transactions on*, vol. 50, no. 6, pp. 1345–1356, 2004.

[109] Y. Weng and D. Tuninetti, "On gaussian interference channels with mixed interference," in *Proc. Information Theory and Applications Workshop*, 2008.

[110] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *arXiv preprint arXiv:1504.04419*, 2015.

[111] R. Bustin, H. V. Poor, and S. Shamai, "The effect of maximal rate codes on the interfering message rate," in *2014 IEEE International Symposium on Information Theory (ISIT)*, pp. 91–95, IEEE, 2014.

[112] I. Sason, "On the corner points of the capacity region of a two-user gaussian interference channel," *IEEE Transactions on Information Theory*, vol. 61, no. 7, pp. 3682–3697, 2015.

[113] M. H. Costa and O. Rioul, "From almost gaussian to gaussian," in *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING (MAXENT 2014)*, vol. 1641, pp. 67–73, AIP Publishing, 2015.

[114] O. Rioul and M. H. Costa, "Almost there corner points of gaussian interference channels," in *Information Theory and Applications Workshop (ITA), 2015*, pp. 32–35, IEEE, 2015.

[115] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 46–55, 2002.

[116] X. Liu and M. Haenggi, "Throughput analysis of fading sensor networks with regular and random topologies," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 4, pp. 554–564, 2005.

[117] G. B. Dantzig and B. C. Eaves, "Fourier-motzkin elimination and its dual," *Journal of Combinatorial Theory, Series A*, vol. 14, no. 3, pp. 288–297, 1973.

Appendix A

Proofs from chapter 3

A.1 Proof of Claim 3.1

Claim 3.1. *Rate splitting gives the same region as joint coding, or* $\mathcal{R}^{RS} = \mathcal{R}$.

Proof: Let us separate R_2 as:

$$R_2 = R_{2,p} + R_{2,c}. \tag{A.1}$$

where $R_{2,p}$ is the rate of the codebook used by user 2 on its private band and $R_{2,c}$ is the rate of the codebook used by user 2 on the common band. Then, if $R_1 < \gamma$, user 1's receiver will be able to decode its message. Furthermore, if we define the region $\mathcal{R}_2^{RS}$ as

$$\begin{aligned}\mathcal{R}_2^{RS} \;:=\; \Bigg\{ & R_1 < W_0 \mathsf{C}\left(\frac{a_{1,2}P_1}{W_0}\right), R_2 < W_c \mathsf{C}\left(\frac{Q}{W_c}\right), \\ & R_1 + R_2 < W_c \mathsf{C}\left(\frac{a_{1,2}P_1}{W_0} + \frac{Q}{W_c}\right) + (W_0 - W_c)\mathsf{C}\left(\frac{a_{1,2}P_1}{W_0}\right)\Bigg\},\end{aligned} \tag{A.2}$$

user 2's receiver will be able to decode its desired message if

$$\left((R_1, R_{2,c}) \in \mathcal{R}_2^{RS} \text{ or } R_{2,c} < W_c \mathsf{C}\left(\frac{W_0 Q}{W_c(W_0 + a_{12}P_1)}\right)\right) \text{ and } R_{2,p} < (W - W_0)\mathsf{C}\left(\frac{P_2 - Q}{W - W_0}\right). \tag{A.3}$$

Using Fourier-Motzkin elimination[117] to eliminate $R_{2,p}, R_{2,c}$, we see that:

$$\mathcal{R} = \mathcal{R}_{MUD}^{RS} \cup \mathcal{R}_{TIN}^{RS}, \tag{A.4}$$

where $\mathcal{R}_{MUD}^{RS}$ (MUD stands for Multi-user decoding) is the set of (R_1, R_2) that satisfy the following inequalities:

$$R_1 \leq \min\left\{W_0 C\left(\frac{a_{12}P_1}{W_0}\right), (W_0 - W_c)C\left(\frac{P_1}{W_0}\right) + W_c C\left(\frac{W_c P_1}{W_0(W_c + a_{21}Q)}\right)\right\} \tag{A.5}$$

$$R_2 \leq W_c C\left(\frac{Q}{W_2}\right) + (W - W_0)C\left(\frac{P_2 - Q}{W - W_0}\right) \tag{A.6}$$

$$R_1 + R_2 \leq W_c C\left(\frac{a_{12}P_1}{W_0} + \frac{Q}{W_c}\right) + (W_0 - W_c)C\left(\frac{a_{12}P_1}{W_0}\right) + (W - W_0)C\left(\frac{P_2 - Q}{W - W_0}\right), \tag{A.7}$$

and $\mathcal{R}_{TIN}^{RS}$ (where TIN stands for treating interference as noise) is the set of rates (R_1, R_2) that satisfy

$$R_1 \leq (W_0 - W_c)C\left(\frac{P_1}{W_0}\right) + W_c C\left(\frac{W_c P_1}{W_0(W_c + a_{21}Q_2)}\right) \tag{A.8}$$

$$R_2 \leq W_c C\left(\frac{W_0 Q}{W_c(W_0 + a_{12}P_1)}\right) + (W - W_0)C\left(\frac{P_2 - Q}{W - W_0}\right), \tag{A.9}$$

or in an equivalent way that better shows the chimney structure of this region

$$\mathcal{R}^{RS} = \mathcal{R}_1^{RS} \cap \mathcal{R}_2^{RS}, \tag{A.10}$$

where $\mathcal{R}_1^{RS}$ denotes the region achievable by receiver 1

$$R_1 \leq (W_0 - W_c)C\left(\frac{P_1}{W_0}\right) + W_c C\left(\frac{W_c P_1}{W_0(W_c + a_{21}Q_2)}\right), \tag{A.11}$$

and $\mathcal{R}_2^{RS}$ denotes the chimney region achievable by receiver 2

$$\mathcal{R}_2 = \mathcal{R}_{2,TIN}^{RS} \cup \mathcal{R}_{2,MUD}^{RS}, \tag{A.12}$$

where $\mathcal{R}_{2,TIN}$ is the region given by

$$R_2 \leq W_c C\left(\frac{W_0 Q_2}{W_c(W_0 + a_{12}P_1)}\right) + (W - W_0)C\left(\frac{P_2 - Q_2}{W - W_0}\right), \tag{A.13}$$

and $\mathcal{R}_{2,MUD}$ is the region given by

$$R_1 \leq W_0 C\left(\frac{a_{12}P_1}{W_0}\right) \tag{A.14}$$

$$R_2 \leq W_c C\left(\frac{Q_2}{W_2}\right) + (W - W_0)C\left(\frac{P_2 - Q_2}{W - W_0}\right) \tag{A.15}$$

$$R_1 + R_2 \leq W_c C\left(\frac{a_{12}P_1}{W_0} + \frac{Q_2}{W_c}\right) + (W_0 - W_c)C\left(\frac{a_{12}P_1}{W_0}\right) + (W - W_0)C\left(\frac{P_2 - Q_2}{W - W_0}\right), \tag{A.16}$$

which by inspection of the inequalities, shows that the regions obtainable by rate-splitting and joint coding are the same.

A.2 Proof of Claim 3.2

Claim 3.2. *Replacing the decoder of Section 3.2 with a non-unique joint typicality decoder will result in the same achievable region*

Proof: Assume that we have replaced the rate-maximizing decoder of Section 3.2 with an indirect decoder. If the rates of the transmitted messages are in the portion of $\mathcal{R}$ that corresponds to decoding interference from user 1's transmitter at user 2's receiver, there should have been one unique pair of messages M_1, M_2 such that the tuple $(X_1^N(M_1), X_2^N(M_2), Y_2^N)$ is jointly typical. Therefore, the relaxed decoding condition of indirect decoder (that the undesired messages need not necessarily be unique) does not come into play.

On the other hand, if the rate pair (R_1, R_2) is in the portion of the rate region that treating interference from user 1's transmitter at user 2's receiver as noise is rate-optimal, it is in the vertical strip of the chimney region. In this part of the rate region, the rate of the message transmitted by user 1 is above what can be reliably decoded by receiver 2. This means that there will be more than one tuple $(X_1^N(M_1), X_2^N(M_2), Y_2^N)$ of channel inputs and output that are jointly typical with high probability for any given message index. All such tuples should have the same message index M_2 in the second component though, as otherwise the treat-as-noise decoder would not be able to reliably decode the message either. Therefore, indirect decoding can achieve the same rate pair in this case as well.

This decoder does not enlarge the rate region, since for any rate pair outside this region, unique joint typicality decoding will fail with probability approaching 1, similarly to the case for conventional joint typicality decoders.

A.3 Proof of Lemma 3.1

Lemma 3.1. *For the AWGN capacity function F defined by Equation 3.8, we have*

$$\sum_{i=1}^{n} F(P_i, n + \sum_{j=1}^{i-1} P_i) = F(\sum_{i=1}^{n} P_i, n). \tag{3.11}$$

Proof: Proof is by induction:

Base case: for $n = 2$, the statement of the lemma is equivalent to (3.9).

Inductive step: Assuming the statement of the lemma holds for $n - 1$ we show that it

holds for n as well:

$$F(\sum_{i=1}^{n} P_i, N) = F(\sum_{i=1}^{n-1} P_i, N) + F(P_n, N + \sum_{i=1}^{n-1} P_i) = \tag{A.17}$$

$$= \sum_{i=1}^{n-1} F(P_i, N + \sum_{j=1}^{i-1} P_j) + F(P_n, N + \sum_{i=1}^{n-1} P_i) \tag{A.18}$$

$$= \sum_{i=1}^{n} F(P_i, N + \sum_{j=1}^{i-1} P_j) \tag{A.19}$$

Where the equality in (A.17) holds by the base case and the equality in (A.18) holds because of the assumption that the equality holds for $n-1$.

A.4 Proof of claim 3.3

Claim 3.3. *Multilayer random coding at user 2's transmitter cannot enlarge the rate region calculated in Section 3.2.*

Proof: Take some arbitrary n-layer codebook $\mathcal{C}_{2,1}, \cdots, \mathcal{C}_{2,n+1}$ with codeword powers $P_{2,1}, \cdots P_{2,m}$. Assume that decoding interference from user 1 is done before layer m. So the order of decoding is $\mathcal{C}_{2,1} \to \cdots \mathcal{C}_{2,m-1} \to \mathcal{C}_1 \to \mathcal{C}_{2,m} \to \cdots \mathcal{C}_{2,n}$. Using the Lemma 3.1 we can see that the sum rate achievable by this method is:

$$\sum_{i=1}^{m-1} F(P_{2,i}, N + \sum_{j=1}^{i-1} P_{2,j}) + F(P_1, N + \sum_{i=1}^{m-1} P_{2,i}) + \sum_{i=m}^{n} F\left(P_{2,i}, (N + \sum_{i=1}^{m-1} P_{2,j} + P_1) + \sum_{j=m}^{i-1} P_{2,j}\right) \tag{A.20}$$

$$= F(\sum_{i=1}^{m-1} P_{2,i}) + F(P_1, N + \sum_{i=1}^{m-1} P_{2,i}) + F(\sum_{k=m}^{n} P_{2,k}, N + \sum_{j=1}^{m-1} P_{2,j} + P_1) \tag{A.21}$$

Which is equivalent to the sum rate of a two-layer coding scheme where the total power of user 2 is divided between the two layers as $\sum_{i=1}^{m-1} P_{2,i}$ and $\sum_{i=m}^{n} P_{2,i}$ respectively. Therefore, the achievable maximum sum rate of n-layer coding cannot be higher than 2-layer coding.

A.5 Proof of Claim 3.4

Claim 3.4. *For the case of weak interference* $a_{1,2}, a_{2,1} < 1$, $g(W_c) > f(W_c)$ *for all* $W_c \in [0, W_0]$. *For the case of strong interference* $a_{1,2}, a_{2,1} > 1$, $g(W_c) < f(W_c)$ *for all* $W_c \in [0, W_0]$.

Proof: We only prove the first part of the claim as proof of the second part is similar with the direction of inequalities changed. We only need to show the proposition that$g(W_c) > f(W_c)$ for all $W_c \in [0, W_0]$ for the values of W_c in the interval of interest at which f and g take finite values, that is from the vertical asymptote of either function that is farthest away from the origin to W_0. In this case it is g's vertical asymptote. This is because we have $R^* > s_1$ due to weak interference, hence $\frac{R_{th}}{R^*} < \frac{R_{th}}{s_1}$ and $1 - \frac{R_{th}}{R^*} > 1 - \frac{R_{th}}{s_1}$. To prove $g > f$, we do a change of variable from W_c to k where the two variables are related by $W_c = \frac{W_0}{k}$. So, for the portion of W_c axis where f and g are both finite they are given by the following two expressions:

$$f = \frac{W_0}{k}\left(\frac{2^{\frac{s_1}{W_0}} - 1}{2^{\frac{(1-k)s_1}{W_0}} 2^{\frac{kR_{th}}{W_0}} - 1} - 1\right) \tag{A.22}$$

And:

$$g = \frac{1}{a_{2,1}} \times \frac{W_0}{k}\left(\frac{2^{\frac{R^*}{W_0}} - 1}{2^{\frac{(1-k)R^*}{W_0}} 2^{\frac{kR_{th}}{W_0}} - 1} - 1\right) \tag{A.23}$$

Disregarding the common $\frac{W_0}{k}$ factor and noting that k ranges in the interval $\left[1, \frac{1}{1-\frac{R_{th}}{R^*}}\right]$, it is easy to see that $\frac{1}{a_21} > 1$, $2^{\frac{R^*}{W_0}} - 1 > 2^{\frac{s_1}{W_0}} - 1$ and $2^{\frac{(1-k)R^*}{W_0}} 2^{\frac{kR_{th}}{W_0}} - 1 < 2^{\frac{(1-k)s_1}{W_0}} 2^{\frac{kR_{th}}{W_0}} - 1$ where the last identity holds because $k > 1$, hence $g > f$ for the W_c's of interest.

A.6 Proof of Lemma 3.2

Lemma 3.2. *The number of solutions of the equation* $a_1 e^{b_1 x} + a_2 e^{b_2 x} + \cdots + a_n e^{bnx} = 0$ *in the real variable* x *where* $b_1 < b_2 < \cdots < b_n$ *is at most the number of sign changes in the sequence of coefficients* $(a_1, a_2, \ldots, a_n)$ *and has the same even-odd parity. In particular, any such equation cannot have more than* $n - 1$ *solutions.*

A.6. Proof of Lemma 3.2

Proof:

This lemma generalizes Descartes rule of signs, which relates the number of sign changes in the coefficient sequence of a polynomial to its number of zeros, too transcendental functions expressible as a sum of exponentials. The proof therefore, follows along the same lines as the proof of Descartes rule of signs. First, we claim that the number of sign changes in the sequence $(a_1, \ldots, a_n)$ and the number of zeros have the same even-odd parity. To see why, we note that for $x \to -\infty$ the function is dominated by the term $a_1 e^{b_1 x}$, likewise for $x \to +\infty$ it is dominated by the term $a_n e^{b_n x}$. So, if a_1, a_n have the same sign, the function has an even number of zeros and if they differ in sign, the function has an odd number of zeros. The parity of the number of sign changes in the sequence $(a_1, \ldots, a_n)$ determines whether or not a_1, a_n have the same sign. So the parity of the sign changes in $(a_1, \ldots a_n)$ and the number of zeros is the same. We prove that the number of zeros is strictly less than the number of sign changes by induction on the latter. Denote by $\#SC$ the number of sign changes in the coefficients of the above equation. It is easy to see that the claim holds for $\#SC = 0$. Now assume that the claim hold for any function with $\#SC = k - 1$ sign changes in the ascending sequence of coefficients. To show that it holds for k changes, consider the function $h(x) = a_1 e^{b_1 x} + \cdots a_n e^{b_n x}$ that has k sign changes in the sequence $(a_1, \ldots, a_n)$ and assume one of these changes occurs for index $m \in \{1, \ldots, n-1\}$, that is $sgn(a_m a_{m+1}) = -1$, now choose some $b \in (b_m, b_{m+1})$. It is not hard to see that $e^{bx} \frac{d}{dx} e^{-bx} h(x) = \sum_{l=1}^{n} a_l (b_l - b) e^{b_l x}$ has exactly $m - 1$ sign changes in its coefficients and therefore by the hypothesis of function satisfies $\#Z \leq \#SC$. Since the multiplicative exponential factor is always non-zero, it does not change the number of zeros of the function and since by Rolle's theorem, differentiating a function reduces the number of zeros by at most 1, $h(x)$ has at most one zero more than $e^{bx} \frac{d}{dx} e^{-bx} h(x)$ and since it has exactly one more sign change in its sequence of coefficients, we have that it also satisfies $\#Z \leq \#SC$.

A.7 Proof of proposition 3.1

Proposition 3.1. *Let* $R_{th} < s_1$, $a_{2,1} > \frac{2^{\frac{R^*}{W_0}} - 2^{\frac{R_{th}}{W_0}}}{2^{\frac{s_1}{W_0}} - 2^{\frac{R_{th}}{W_0}}}$ *and* $a_{1,2} < 1$. *Then* f *and* g *intersect at a unique point in the interval* $(0, W_0)$. *Denoting this unique solution of* $f(W_c) = g(W_c)$ *by* $W_c = w$, $\partial_{\mathrm{rm}}\mathcal{D}_1$ *is described by* $Q = f(W_c), W_c \in [0, w]$, $Q = g(W_c), W_c \in [w, W_0]$ *and the vertical line segment* $\{W_0\} \times [0, g(W_0)]$.

Proof: Since $a_{1,2} < 1$, then $s_1 < R^*$. This together with the assumption $a_{2,1} > \frac{2^{\frac{R^*}{W_0}} - 2^{\frac{R_{th}}{W_0}}}{2^{\frac{s_1}{W_0}} - 2^{\frac{R_{th}}{W_0}}}$ implies that $a_{2,1} > 1$. The coefficients of variable T_c in the four exponents appearing in (3.27) are ordered as $2R_{th} - R^* - s_1 < R_{th} - R^* < R_{th} - s_1 < 0$. After sorting the coefficients of the exponential terms in (3.27) according to Lemma 1, we obtain the sequence $(1 - a_{2,1})2^{\frac{R^*+s_1}{W_0}}, 2^{\frac{R^*}{W_0}}(a_{2,1}2^{\frac{s_1}{W_0}} - 1), 2^{\frac{s_1}{W_0}}(a_{2,1} - 2^{\frac{R^*}{W_0}}), 2^{\frac{R^*}{W_0}} - a_{2,1}2^{\frac{s_1}{W_0}}$. Since $a_{2,1} > 1$, the first and second terms are negative and positive, respectively. The third term can be either positive or negative. The forth term $2^{\frac{R^*}{W_0}} - a_{2,1}2^{\frac{s_1}{W_0}}$ can be easily seen to be negative due to the assumption $a_{2,1} > \frac{2^{\frac{R^*}{W_0}} - 2^{\frac{R_{th}}{W_0}}}{2^{\frac{s_1}{W_0}} - 2^{\frac{R_{th}}{W_0}}}$. Hence, we obtain the sequence of signs as $-, +, *, -$ where $*$ stands for the sign of $2^{\frac{s_1}{W_0}}(a_{2,1} - 2^{\frac{R^*}{W_0}})$. Regardless of the status of $*$, the number of sign changes in the sequence $-, +, *, -$ is two. As such, $f = g$ has exactly one solution by Lemma 3.2 and the discussion appearing immediately after this Lemma. According to (3.18) and (3.19), f is smaller that g for W_c sufficiently close to $W_0\left(1 - \frac{R_{th}}{R^*}\right)$. The inequality $a_{2,1} > \frac{2^{\frac{R^*}{W_0}} - 2^{\frac{R_{th}}{W_0}}}{2^{\frac{s_1}{W_0}} - 2^{\frac{R_{th}}{W_0}}}$ is easily seen to be equivalent to $f(W_0) > g(W_0)$. This shows that f an g must intersect in at least one point $W_c = w$ where $W_0\left(1 - \frac{R_{th}}{R^*}\right) < w < W_0$. However, we showed that $f = g$ has exactly one solution. As such, w must be the unique solution to $f = g$.

Appendix B

Proofs from chapter 4

B.1 Proof of Lemma 4.1

Lemma 4.1. *Algorithm 2 is a greedy colouring algorithm that colours any q-independence graph G_q using $c\chi(G_q)$ colours for c a constant.*

The colouring algorithm known as the Min-Degree-Last heuristic, first appeared in [101] and therefore is known as Hochbaum's greedy colouring algorithm. It was shown in [101] that if we define $\delta := \max\{d|\, \exists G" \subseteq G \forall v \in V(G')\, d_{G'}(v) \geq d\}$, this algorithm takes $\delta + 1$ colors to color G. This implies an upper bound on χ, the chromatic number of a graph, of $\delta + 1$.

For graph families on which Min-Degree-Last algorithm works within a constant factor of χ, such as geometric and unit disk graphs, the standard way of showing this is using a corresponding linear lower bound on χ in terms of δ and this is what we set out to prove.

Specifically assume that, for a graph family, we can show that there is a vertex v of degree d such that the subgraph induced by the neighbourhood of v [1] has a bounded maximum independent set size of k, then it is easy to see that $\chi > \frac{d}{k} - 1$ or $d < k\chi - k$. This is a lower bound on χ in terms of d. We say that a graph family has bounded neighbourhood

[1] The subgraph of a graph G containing V and vertices adjacent to v and edges between them.

independence number if has the above property. To relate this bound on d to a bound on δ, it suffices to have a stronger version of this property, which we term having hereditary bounded neighbourhood independence number. What this property says is that there exists a vertex v in every induced subgraph G' of G from the family for which the bounded neighbourhood independence number property holds uniformly. If a graph family has this latter property, by applying the neighbourhood independence number bound to G" in the definition of δ above, we obtain that $\delta < k\chi - k$ and therefore Min-Degree-Last algorithm colours the graphs from the family within a constant of the optimum.

The following shows, by selecting v to correspond to the shortest link, that the q-independence graph has the hereditary bounded neighbourhood independence number property

Theorem B.1. *For $q \geq 2$, the graph $G_q([L])$ has hereditary bounded neighbourhood independence number.*

Proof: Assume j to be the shortest link in the network, we want to show that there is a constant upper bound on the size of the maximum independent set of the neighbourhood $N(j)$ of j in the q-dependence graph. As will be seen, this proof and that of Lemma 4.4 are similar except that here, the shortest link length is only guaranteed to be merely shorter than other links, rather than shorter by a scaling factor depending on network size.

Denote by $I \subseteq N(j)$ one such independent set of links. By the selection of j and I, we have that for each $i \in I$, $\ell_i \geq \ell_j$ and:

$$\min\{D_{i,j}, D_{j,i}\} \leq q\sqrt{\ell_j \ell_i} \tag{B.1}$$

I is by definition a q-independent set, which we show combined with these inequalities, upper bounds its size by a constant. To simplify discussion, we re-index the elements of I as $I = \{1, 2, \ldots, |I|\}$ and j by 0. Since for each different $i, k = 1, 2, \ldots, |I|$, i and k are q-independent, then we have $D_{i,k} > q\sqrt{\ell_i \ell_k}$ and $D_{k,i} > q\sqrt{\ell_k \ell_i}$. Let us assume that $\ell_i \leq \ell_k$.

B.1. Proof of Lemma 4.1

Then using the triangle inequality we have $D(s_k, s_i) \geq D_{k,i} - \ell_i > q\sqrt{\ell_i \ell_k} - \ell_i$, and since $\ell_i \leq \ell_k$, we get

$$D(s_k, s_i) > (q-1)\sqrt{\ell_i \ell_k}. \tag{B.2}$$

With a similar argument we get

$$D(r_k, r_i) > (q-1)\sqrt{\ell_i \ell_k}. \tag{B.3}$$

Now, from equation B.1, we have that for each $k \in I$, at least one of $D_{0,k} \leq q\sqrt{\ell_0 \ell_k}$ or $D_{k,0} \leq q\sqrt{\ell_0 \ell_k}$ holds. Therefore, half or more of the links satisfy one of these inequalities. Without loss of generality we consider the case where the first inequality holds for half or more of links. We denote these links by I'. Now, consider the sender of $j := 0$, s_0 and the set of receivers of links in I', $R = \{r_t | t \in I'\}$ which is re-indexed again without loss of generality to be $I' = \{1, \ldots, t\}$. The argument is symmetric if half or more of the links satisfy the second condition with senders and receivers swapped so we omit the analysis for that case.

We have that $|R| \geq |I|/2$, so any constant bound on $|R|$ implies the same for $|I|$. We have:

$$\sqrt{\ell_0} \leq \sqrt{\ell_m} \text{ for } m = 1, 2, \ldots, t \tag{B.4}$$

$$D(s_0, r_m) \leq q\sqrt{\ell_0}\sqrt{\ell_m} \text{ for } m = 1, 2, \ldots, t \tag{B.5}$$

$$D(r_m, r_n) > (q-1)\sqrt{\ell_m}\sqrt{\ell_n}, \text{ for } m, n = 1, 2, \ldots, t, m \neq n, \tag{B.6}$$

Where B.4 is true by definition of link 0, B.5 holds because all links in I' are adjacent to link 0 and B.6 is a restatement of B.3. From the triangle inequality, for $m, n = 1, 2, \ldots, t, m \neq n$ we have

$$D(r_m, r_n) \leq D(s_0, r_m) + D(s_0, r_n),$$

so using B.5 for the left hand side and B.6 for the right hand side, we get

$$q\sqrt{\ell_0}\sqrt{\ell_m} + q\sqrt{\ell_0}\sqrt{\ell_n} > (q-1)\sqrt{\ell_m}\sqrt{\ell_n} \tag{B.7}$$

Suppose the smallest between ℓ_m and ℓ_n is ℓ_m. Then from (B.7) we get $\sqrt{\ell_0} > \frac{q-1}{2q}\sqrt{\ell_m}$, thus we have that $\sqrt{\ell_0}$ is more than $\frac{q-1}{2q}\sqrt{\ell_m}$ for all but one of $m > 0$. If we suppose without loss of generality these m's are $1, 2, \ldots, t-1$. Then we have that:

$$D(s_0, r_m) < \frac{2q^2}{q-1}\sqrt{\ell_0}^2 \text{ and } d(r_m, r_n) > \frac{q-1}{q^2}\sqrt{\ell_0}^2$$

for $m, n = 1, 2, \ldots, t-1, m \neq n$. The last two inequalities imply that the balls $B(r_m, \frac{q-1}{q^2}\sqrt{\ell_0}^2/2)$ for different m's don't intersect and all of them are contained in the ball $B(s_0, (2q^2/(q-1) + q - 1/2q^2)\sqrt{\ell_0}^2)$. As the metric space has a doubling dimension m, we get $t - 1 \leq C(\frac{4q^4}{(q-1)^2} + 1)^m$. Therefore $t \leq C(\frac{4q^4}{(q-1)^2} + 1)^m + 1$ and $|I| \leq 2C(\frac{4q^4}{(q-1)^2} + 1)^m + 2$ which is a constant independent of network size.

B.2 Proof of Lemma 4.2

Lemma 4.2. *Any almost equilength class of links in a doubling metric space that is q-independent is also SINR feasible under uniform power assignment.*

Proof: Let l, L be the shortest and longest link length in the set X. We first observe that senders of links in X are of situated at least $(q-2)l$ apart. Otherwise, if $D(s_u, s_w) \leq (q-2)l$, for some pair u, w, by triangle inequality we have $D_{u,w} \leq D(s_u, s_w) + \ell_w + \ell_v \leq ql$, and similarly $D_{wu} \leq ql$. Which means u, w are not q-independent, which contradicts our assumption. Let X' be the set of senders of links in X. Let $r = (q-2)l/2$. The previous discussion implies that X' is an r-packing. The definition of a doubling metric implies the

B.2. Proof of Lemma 4.2

following about the packing density in this space:

$$\mathcal{P}(B(x, tZ), Z) \leq Ct^m \ . \tag{B.8}$$

Let n be a natural number and s_i be a sender in X' belonging to link i. Let $X_n = \{s_j \in X' | D(s_i, s_j) < nr\}$ be the set of senders that are less than nr apart from s_i, and define $R_n := X_n \setminus X_{n-1}$. By q-independence, we have that $X_2 = \emptyset$. Each sender s_j in R_n is at least $(n-1)r$ far from s_i, so $D_{j,i} \geq (n-1)r - L \geq (n-2)r$ if $q \geq 6$. Since $\ell_x \leq 2l$, the affectance of s_j on r_i is at most

$$I_j(i) = \frac{1/D_{j,i}^{\alpha}}{1/\ell_i^{\alpha}} \leq \left(\frac{2l}{(n-1)r}\right)^{\alpha} = \left(\frac{4}{(n-2)(q-2)}\right)^{\alpha}, \quad \forall s_y \in T_g.$$

Observe that

$$\frac{1}{(n-2)^{\alpha}} - \frac{1}{(n-1)^{\alpha}} = \frac{(n-1)^{\alpha} - (n-2)^{\alpha}}{(n-1)^{\alpha}(n-2)^{\alpha}} \leq \frac{\alpha(n-1)^{\alpha-1}}{(n-1)^{\alpha}(n-2)^{\alpha}} < \frac{\alpha}{(n-2)^{\alpha+1}} \ .$$

Then,

$$\begin{aligned} I_X(i) &= \sum_{n\geq 3} I_{R_n}(x) \\ &\leq \sum_{n\geq 3} |X_n \setminus X_{n-1}| \cdot \left(\frac{4}{(n-2)(q-2)}\right)^{\alpha} \\ &= \left(\frac{4}{q-2}\right)^{\alpha} \sum_{n\geq 3} |X_n| \left(\frac{1}{(n-2)^{\alpha}} - \frac{1}{(n-1)^{\alpha}}\right) \\ &\leq \left(\frac{4}{q-2}\right)^{\alpha} \sum_{n\geq 3} |X_n| \frac{\alpha}{(n-2)^{\alpha+1}} \ . \end{aligned} \tag{B.9}$$

The balls of radius r centered at points in X_n are all contained within the ball $B(i, (n+1)r)$. For $n \geq 3$, the packing bound (B.8) then implies that $|X_n| \leq \mathcal{P}(B(i, (n+1)r), r) \leq C(n+1)^m$,

and thus we have that

$$\frac{|X_n|}{(n-2)^{\alpha+1}} \leq \frac{C(n+1)^m}{(n-2)^{\alpha+1}} \leq \frac{4^m C}{(n-2)^{\alpha+1-m}} \ .$$

Continuing from (B.9),

$$I_X(i) \leq \left(\frac{4}{q-2}\right)^{\alpha} \alpha \cdot 4^m C \sum_{x \geq 1} \frac{1}{x^{\alpha+1-m}} = \left(\frac{4}{q-2}\right)^{\alpha} \alpha \cdot 4^m C(\zeta(\alpha+1-m)) := \left(\frac{4}{q-2}\right)^{\alpha} C_1 \ .$$

Where ζ is the Riemann zeta function. Thus, S is an SINR feasible with SINR s, where $s = \frac{1}{C_1}(\frac{q-2}{4})^{\alpha}\beta$.

B.3 Proof of Lemma 4.3

Lemma 4.3. *[SINR strengthening] Any collection of S-SINR feasible ISets can be refined into a collection S'-SINR feasible subsets ($S' > S$), increasing the cardinality of the collection by at most $\left\lceil \left(\frac{2S'}{S}\right) \right\rceil^2$.*

Proof: Without loss of generality, assume the S-SINR feasible schedule to consist of one timeslot only. The procedure that follows works the same if applied individually to every timeslot. First, sort all the links in ascending order of length, start with empty slots $S_1, \ldots, S_k$. Going through the links according to order we put each l in first slot for which $I_{S_i}(l) < 1/2S'$. Since the original schedule was S-feasible, the number of sets will be at most $\left\lceil \frac{2S'}{S} \right\rceil$. Now, we sort the links in descending order of length. For each slot S_i, it is partitioned into slots $S_i^1, \ldots, S_i^m$ where we go through the links in the new descending length order and add the link l to the first S_i^j for which $I_{S_i^j}(l) < 1/2S'$. Again, the length of schedule is increased by at most a factor $\left\lceil \frac{2S'}{S} \right\rceil$. Now, since each link is receiving at mos 1/2S' Normalized ISR from shorter and the same amount from longer links in its slot, the total ISR on each link, by the additivity property, is at most $1/S'$ for a total expansion factor of $\left\lceil \frac{2S'}{S} \right\rceil^2$

B.4 Proof of Lemma 4.4

Lemma 4.4. *Any q-independent set of links for $q \geq 3$ is $\mathcal{O}(\log\log\Delta)$-greedily binnable.*

Proof: Such a set S consists of two types of links: those that affect the link $j \notin S$ by at least $\frac{1}{2n}$ under $\ell^{\alpha/2}$ power assignment, and those that are affected by j by this much. At least half of the links should belong to one of these classes. Without loss of generality consider the first type to have more than half; the argument is nearly symmetrical for the second type, and is omitted.

Consider a pair i, i' in S that put normalized ISR on j of at least $1/2n$, and suppose without loss of generality that $\ell_i \geq \ell_{i'}$. The normalized ISR (affectance) condition of i on j means that: $\sqrt{\ell_j \ell_i}^{\alpha} \geq D_{ij}^{\alpha} \cdot 1/2n$, or

$$D_{ij} \leq \sqrt{\ell_j \ell_i} 2n^{1/\alpha} = \sqrt{\ell_j \ell_i (2n)^{\frac{2}{\alpha}}} \ .$$

Similarly, $D_{i'j} \leq \sqrt{\ell_j \ell_{i'} (2n)^{\frac{2}{\alpha}}}$. By triangle inequality it is the case that:

$$D_{i'i} \leq D(s_{i'}, r_j) + D(r_j, s_i) + D(s_i, r_i) \leq \ell_i + \sqrt{4.(2n)^{\frac{2}{\alpha}} \ell_j \ell_i} < 3\ell_i,$$

using the fact that $\ell_i \geq 2.(4n)^{\frac{2}{\alpha}} \ell_j$. Similarly, we have:

$$D_{ii'} \leq D_{ij} + D_{i'j} + \ell_{i'} \leq \ell_{i'} + \sqrt{4.(2n)^{\frac{2}{\alpha}} \ell_j \ell_i} \ .$$

Multiplying these together, we get:

$$D_{i'i} \cdot D_{ii'} \leq 3\ell_{i'}\ell_i + 3\sqrt{4.(2n)^{\frac{2}{\alpha}} \ell_j \ell_i} \cdot \ell_i \ .$$

By the assumption of q-independence for $q \geq 3$ $D_{i'i} \cdot D_{ii'} \geq 9\ell_i \ell_{i'}$. By combining the last

two inequalities and canceling a $6\ell_i$ factor, we have that $\ell_{i'} \leq \sqrt{(2n)^{\frac{2}{\alpha}}\ell_j\ell_i}$, or

$$\ell_i \geq \frac{\ell_{i'}^2}{(2n)^{\frac{2}{\alpha}}\ell_j} \ . \tag{B.10}$$

If we label the links in S as $1, 2, \ldots, t$ in ascending length order, Equation (B.10) implies that

$$\frac{\ell_{u+1}}{\ell_u} \geq \frac{\ell_u}{\ell_j(2n)^{\frac{2}{\alpha}}} \geq \frac{2\ell_u}{\ell_1}, \tag{B.11}$$

for all $u = 2, 3, \ldots, t$. Thus, if we denote $\lambda_u := \ell_u/\ell_1$, we get using (B.11) that $\lambda_{u+1} \geq 2\lambda_u^2$, and by induction it's not hard to see that $\lambda_t \geq 2^{2^{t-1}-1}$. Hence, $|S| = t \leq \log\log\lambda_t + 2 \leq \log\log\Delta + 2$, and the result therefore follows.

B.5 Proof of Theorem 4.1

Theorem 4.1. *Algorithm 3 schedules any p-greedily binnable set of links in $\mathcal{O}(p\log n)$ slots*

Proof: It is easy to see that B_i's are well-separated sets by definition. The number of B_i's is $O(\log n)$. It suffices to show that each B_i is scheduled into $O(p)$ slots using the mean power assignment. According to Theorem 4.2, each Q_i is at least a $k\beta$-SINR feasible set ($k > 1$) using uniform power. This is because each K_j is a nearly equilength set of links that is also q-independent. Using Theorem 4.3, K_j can be transformed into an e-SINR feasible schedule with at most $O((e/s)^2)$ slots, where $e = 2^{\alpha/2+1}\beta$. Let S_j be some slot from the resulting schedule of K_j. Let $S = \cup_j S_j$. For completing the proof it is enough to show that S is binned into $p + 1$ SINR-feasible slots, since SINR strengthening generates at most a constant number of slots.

For scheduling S the algorithm uses $p + 1$ slots T_i for $i = 1, 2, \ldots, p + 1$. It assigns each link v to a slot T_r, which does not contain links w, such that $\ell_w \geq 2.(4n)^{\frac{2}{\alpha}}\ell_v$ and v and w are $1/(2n)$-interfering. Such a T_r exists because the set Q is p-greedily binnable. Consider a link $v \in T_r$ which we took from the slot S_k. The affectance on v by nearly equilength links

(i.e. links from $S_k \cap T_r$) is at most $1/e$ because of the e-feasibility property. Changing the power assignment in the set S_k from uniform to mean power increases the affectance by at most $2^{\alpha/2}$, so overall the affectance by the links with nearly the same length as v is at most $2^{\alpha/2}/e = 1/2$. The links from $T_r \setminus S_k$, each can affects v by less than $1/(2n)$ by definition, and since their number is at most n, the total affectance by those links, according to the additivity of affectance is at most $1/2$. This shows that $I_{T_r}(v) < 1$, i.e. T_r is SINR-feasible, which completes the proof.

B.6 Proof of Theorem 4.2

Theorem 4.2. *Using Algorithm 1 gives $\mathcal{O}(\log n \log\log \Delta)$-factor approximation to the problem of link-scheduling with $\ell^{\alpha/2}$ power control.*

Proof: First, note that feasibility for $3^{2\alpha}\beta$-SINR is by definition a stronger condition that 3-independence. Therefore any partition into ISets with $3^{2\alpha}\beta$-SINR level is at least as large as the minimal vertex colouring of the graph $G_3([L])$. So $\chi(G_3([L]))$ gives a lower bound on the minimum-length schedule with $3^{\alpha}\beta$ SINR level. Lemma 4.3 means that this schedule is itself within a constant factor of length from the minimal β-SINR feasible schedule. This is because the latter schedule can be converted by Lemma 4.3 to an at most constant-factor longer $3^{\alpha}\beta$-schedule. This latter schedule is by definition longer than the *minimum-length* $3^{\alpha}\beta$-SINR schedule. Together, these two observations mean that the best β-SINR schedule is within a constant factor of $\chi(G_3([L]))$. Therefore, the $\mathcal{O}(\log(n)\log(\log(\Delta))\chi(G_3([L])))$-long schedule output by Algorithm 1 is an $\mathcal{O}(\log(n)\log(\log(\Delta)))$-factor approximation for the minimum-length scheduling problem.