COVERAGE AND CAPACITY OF NEXT GENERATION CELLULAR RADIO SYSTEMS: BANDWIDTH SHARING AND MASSIVE MIMO

BY NARAYANAN KRISHNAN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Roy D. Yates and Narayan B. Mandayam

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2013

ABSTRACT OF THE DISSERTATION

Coverage and Capacity of Next Generation Cellular Radio Systems: Bandwidth Sharing and Massive MIMO

by Narayanan Krishnan

Dissertation Director: Roy D. Yates and Narayan B. Mandayam

In the first part of the thesis we investigate bandwidth allocation in next generation cellular systems employing relays similar to LTE advanced systems with type-I relays. We jointly optimize the bandwidth and power usage under constraints on required rate, bandwidth and transmit power. We study scenarios wherein, the relay acts as a forwarder for multiple User Equipments (UEs/users) in both uplink and/or downlink. This includes scenarios when the relay has its own data to send along with forwarding the data of other users. We examine the weighted power minimization problem for relaying with multiple users. We also show specific results with N user scenario and also single user case in order to understand how the bandwidth and power are allocated. Numerical evaluations with N users on a three sector LTE-A cell employing Fractional Frequency Reuse (FFR) indicate that power savings of at least 3 dB can be achieved by optimizing over both bandwidth and power.

Base stations with a large number of transmit antennas have the potential to serve a large number of users simultaneously at higher rates. They also promise a lower power consumption due to coherent combining at the receiver. However, the receiver processing in the uplink relies on the channel estimates which are known to suffer from *pilot contamination*. In the second part of the thesis, we perform an uplink

large system analysis of multi-cell multi-antenna system when the receiver employs a matched filtering and MMSE filtering with a pilot contaminated estimate. We find the asymptotic Signal to Interference plus Noise Ratio (SINR) as the number of antennas and number of users per base station grow large while maintaining a fixed ratio. To do this, we make use of the similarity of the uplink received signal in a multi-antenna system to the representation of the received signal in CDMA systems. The asymptotic SINR expression for both matched filter and the MMSE filter explicitly captures the effect of pilot contamination and that of interference averaging. This also explains the SINR performance of receiver processing schemes at different regimes such as instances when the number of antennas are comparable to number of users as well as when antennas exceed greatly the number of users. Specifically, we explore the scenario where the number of users being served are comparable to the number of antennas at the base station. It is seen that MMSE filter is capable of suppressing the in-cell interference and that the interference power due to pilot contamination is the same as in a matched filter with a pilot contaminated estimate. We find the expression for the amount of interference suppression obtained using an MMSE filter which is an important factor when there are significant number of users in the system as compared to the number of antennas. We validate the asymptotic expression through simulations and compare with an MMSE filter with a perfect estimate. Simulation results for achievable rate is close to theory for even a 10-antenna base station with 3 or more users per cell. In a typical set up, in terms of the five percentile SIR, the MMSE filter is shown to provide significant gains over matched filtering and is within 5 dB of MMSE filter with perfect channel estimate. We also show that the achievable rates are within a 1 bit/symbol of the MMSE with perfect estimate when the number of users is comparable to the number of antennas.

Acknowledgements

Although unaware about the merits of pursuing a doctorate, retrospectively it was one of the best decisions that I had made in my life - it gave me a career and a life.

As for my career, I am grateful to my advisers Roy and Narayan. I treasure the way of critical thinking that I acquired over the four years which has given me the confidence to deal with any scientific literature. I value it more than the specific skills that I acquired in my doctorate studies. Roy nudged my thought process into new ideas that I can immediately connect with. Narayan piqued my thought with questions related to the problem at hand. Most times I learned to proceed in research through their discussions. I am also grateful to have Prof. Predrag Spasojevic, Prof. Waheed Bajwa and Dr. Thomas L. Marzetta in my thesis committee and to have received their feedback.

As for life, it is hard to imagine that I would have met someone as beloved as Remya if I was not in Rutgers. She has been unfailing in her love, understanding and support through out the last three years. I would have struggled to complete my thesis without her by my side.

Finally, I would like to acknowledge all my room-mates, close friends, WINLAB colleagues whom I have spent time personally since I joined in 2009. I refrain from mentioning individual names lest I may forget someone - but life in Rutgers would not be enjoyable and interesting without you guys.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Scope of Thesis

Physical layer transmission in next generation cellular radio system is typically characterized by the Orthogonal Frequency Division Multiplexing (OFDM). In the frequency domain, OFDM is characterized by closely spaced orthogonal sub-carriers each loaded with data symbols. OFDM is motivated by the fact that sinusoids - which are time domain equivalent of sub-carriers, are eigenfunctions of linear time invariant system. In effect this means that when sub-carriers are passed through underspread wireless channels, the received signal is the same sinusoid altered in amplitude and phase. In practice cyclix prefix is used to maintain eigenfunction property as the transmission is time limited to a symbol duration of 71.4μs. In coherent reception, receive filters make use of the amplitude and phase of the channel to reliably decode transmitted data modulated by the sinusoid.

Typically the bandwidth available for a single cell system is divided into number of sub-carriers and the transmission time into different time slots called frames. The smallest unit of bandwidth and time that can be allocated is represented by **Physical Resource Blocks(PRBs)**. For example in LTE, the 12 sub-carriers each of bandwidth $\Delta f = 15$ kHz over a period 0.5 milliseconds(ms) corresponding to 7 symbols constitute one PRB. The PRBs in OFDM provide the flexibility to allocate resources in time axis as well as frequency axis. For example, over a system bandwidth of 20 MHz and frame duration 10ms, a total of 700 PRBs are available for a scheduler to allocate resources to its users. In accordance with some utility maximization each user may be allocated different number of orthogonal PRBs. For example in LTE-A provisions to allocate a bandwidth of 1.25, 2.5, 5, 10 and 20 MHz per user is available while transmitting at the same time slot. These correspond to allocations of 6, 12, 25, 50, 75, 100 PRBs

respectively. When necessary, advanced schedulers may also consider incorporating multiple users sharing at a PRB as along as the per user utility can be met. A great deal of literature is devoted to the study of such multi-user systems in downlink as well as uplink. For details see [46] and the references therein.

Even in orthogonal resource allocation numerous possibilities of assigning PRBs to users arise. For example all users transmit simultaneously during a frame but in different bands. At the other extreme in a time shared system users transmit in all the sub-carriers but in non-overlapping times. Further many intermediate possibilities are also conceivable where a set of users can transmit with no two users sharing the same PRB. While PRBs represent the smallest unit of bandwidth and time that can be allocated to a user, a **Resource Element** in OFDM is represented by the sub-carrier of a symbol. Therefore, in a PRB there are $12 \times 7 = 84$ resource elements.

Motivated by the flexibility in allocating PRBs in OFDM physical layer, this thesis examines two relevant problems related to next generation cellular radio networks. The first part is concerned with introducing relays into a cellular system. We address the problem of a centralized base station flexibly allocating bandwidth in links such that minimum weighted power is consumed by the system. Accordingly, the first part of the thesis is described by the title - "**Bandwidth Sharing for Relaying in Cellular System**". Further we also examine a scenario where the allocation is based on a time sharing and relate the results to bandwidth sharing in terms of weighted power minimization.

The second part of the thesis investigates the performance of a cellular system with large number of antennas at the base station. We focus on an arbitrary PRB consisting of a coherence time and coherence bandwidth. We consider that multiple users within the cell as well as in other cells may be scheduled to transmit in that PRB. Being a multi-cell and a multi-user system it is necessary to deal with the inter-cell and intra-cell interference, respectively. We primarily consider the performance of uplink linear receivers although similar analysis can also be done for downlink linear precoders. From a broad perspective, the multi-user MIMO system within a PRB can be considered as behaving like CDMA systems where the channels gains of each user represent the unique

signature sequence. Accordingly the second part of the thesis is described by the title - "**Multi-cell Multi-user Massive MIMO Systems: Large System Analysis under Pilot Contamination**".

1.1 Bandwidth Sharing for Relaying in Cellular Systems

Fourth generation(4G) cellular standards such as LTE-A and IEEE802.16j are poised to meet the requirements of International Mobile Telecommunications-Advanced (IMT-A) standards [1]. These systems are promising to provide peak data rates of up to 100 Mbps and 1 Gbps in high mobility environments and pedestrian environments respectively, and hence provide high speed mobile broadband access [4,5]. One of the key technology components of next generation cellular systems is relaying, which has been shown to provide better throughput and increased coverage [6,7].

Another aspect of 4G cellular systems is the physical layer OFDMA which enables the centralized base station to flexibly allocate bandwidth to its users orthogonally by partitioning the bandwidth into sets of sub-carriers [3]. This gives the base station the dimension of bandwidth to exploit when optimizing its resources. Coupled with the introduction of coverage extending relays, in this thesis we investigate the power savings that a cellular system is capable of by flexibly allocating bandwidth and power. We denote the term "Bandwidth Sharing" for the flexible allocation of bandwidth depending on the demand on the links in the system.

We formulate a weighted power minimization problem, optimizing over both power and bandwidth under rate, bandwidth and power constraints for serving multiple users. We develop theoretical insights into the nature of optimal solution when the system has unconstrained maximum power. We also examine the case when there are multiple relays which although is mathematically cumbersome is conceptually simple extension to the case of a single relay serving multiple users. A standard method of inter-cell interference coordination in next generation cellular system is through the concept of Fractional Frequency Reuse(FFR). With FFR, the total available bandwidth of a single cell is divided into sub-bands. Users in the cell edge never transmit or receive in

the same sub-band as adjacent cell edge users. However, FFR limits the flexibility in allocation of bandwidth as the cell edge users will now be restricted to transmit from a particular sub-band. Even in such a scenario it is seen that bandwidth sharing with relays provides total power gain of about 3.5 dB as compared to baseline scheme of fixed allocation of bandwidth per link. Further, in terms of power minimization we relate the "Bandwidth Sharing" system with that of a time sharing system as represented in LTE-A relaying standards. We show that when peak power constraints are active then the time sharing system might perform worse than a bandwidth sharing system.

1.2 Massive Multi-cell Multi-user MIMO Systems: Large System Analysis

In view of large future demand for data, the general body of physical layer research in next generation cellular systems focuses on small cells, massive MIMO, network MIMO and combinations of these as potential technologies to increase the system throughput. The goal is to increase the throughput in the physical layer to satisfy the large increase in demand for data. In the second part of the thesis we focus on a Massive MIMO system where it is assumed that the base station is equipped with hundreds of antennas as compared to the conventional MIMO systems where the antennas deployed at the base station is less than 10. Theoretically, within the constraints of bandwidth and power the increase in capacity is brought about by the large increase in number of antennas.

For users in cellular radio system, the available coherence time during which the channel can be considered constant is relatively small due to high mobility. Therefore, it is necessary for the users to perform channel estimation at the start of every coherence time for coherent reception of data in a multi-antenna system. The system considers the estimate to be reliable for all the sub-carriers spanning the coherence bandwidth and the coherence time. In the second part of the thesis, the set of sub-carriers spanning the coherence bandwidth over the coherence time, during which the channel estimate is reliable is considered a PRB. We assume a block fading model in which the channel realizations are independent across the PRBs.

The necessity to do channel estimation and having a large number of antennas implies that the system be Time Division Duplexing(TDD) and use channel reciprocity property. This is because in the downlink of a Frequency Division Duplexing(FDD) system allocating as many orthogonal resources for channel training as the number of antennas is prohibitive. The training and the Channel State Information(CSI) feedback will then take up a significant portion of the PRB at the expense of data transmission. Even in the uplink allocating mutually orthogonal training sequences for all users in system might require as much number of resource elements as the number of users in the system. For example, consider an LTE system with root mean square (r.m.s) delay spread of $T_d = 4.76\mu$s, and coherence time 500μs corresponding to coherence symbol duration of $T_c = 7$ symbols. The number of coherent sub-carriers is then approximately given by $N_c = 1/\Delta f T_d = 14$ where we use the fact that coherence bandwidth is inverse of the r.m.s delay spread. In such a system there are 98 resource elements in a PRB. If we are considering a seven cell system with 14 users per cell, we can already see that in uplink we require 98 resource elements for channel estimation with no time for data transmission. Also, a 100 antenna system using FDD is out of question and we have to rely on TDD and channel reciprocity for communication. This implies that non-orthogonal pilot sequences are needed to employed so as to allow some time for data transmission. As shown in [27], allocating non-orthogonal pilot training sequences for the users in the system results in a fundamental limitation called *pilot contamination*.

Much of the research in large MIMO systems with Rayleigh channel can be borrowed from the considerable literature for CDMA systems. The channel vector with independent and identically distributed(i.i.d) entries for the large MIMO system is analogous to signature sequence in a CDMA system so that antennas contribute to the processing gain. For example, the uplink analysis of an asymptotic regime [35] with both users and signature sequences tending to infinity translates directly to results in a large MIMO system when signatures are replaced by antennas. In has been observed in CDMA that the asymptotic analysis is a good approximation for practical number of users and spreading factor of the signature sequences. While in a CDMA system we assume that the signature sequences are known, there are practical limitations in channel estimation

a mobile radio multi-antenna channel (antenna signatures) in a multi-cellular system. In this work we explore this limitation when users simultaneously estimate the channel and the estimates are subject to pilot contamination.

Recent work in [27] have analysed the effect of pilot contamination problem from a large number of antennas perspective in order to show the pilot interference limitation in the received SINR. When there is a large excess of antennas as compared to the number of users, the channels between two users are asymptotically orthogonal by the law of large numbers. As found in [35] even though the number of users per cell is small as compared to the number of antennas the interference power contributed by each interfering users during linear detection (matched filter and MMSE filter) sum up to non-zero and significant value. We extend the results obtained in [27] and find the asymptotic expression for the SINR when the number of users and the number of antennas per cell go to infinity at a fixed ratio. We also observe that the asymptotic analysis is a good approximation for training based MIMO systems with practical number of antennas and users. Hence our results represent the SINR and achievable rate expression for practical scenarios of uses and antennas per cell for matched filter and MMSE filter receivers employing a pilot contaminated channel estimate. In [35] the channel is assumed to perfectly known, hence our work can also be considered as an extension of asymptotic analysis when the channel estimate is impaired by pilot contamination. In finding the asymptotic SINR we also characterize the effect pilot contamination power and the interference averaging power for both the filters.

1.3 Thesis Organization

The thesis is organized into three chapters. Chapter 2 pertains to "Bandwidth Sharing for Relaying in Cellular Systems". After motivating the problem and presenting the system model in section 2.1 the main results for Bandwidth Sharing are presented in section 2.3. Section 2.4 presents results specific to single user system and section 2.5 relates bandwidth sharing to a time shared system with relays. Chapter 3 similarly corresponds to the "Massive MIMO" systems. After motivating the scope of having a large number of antennas at the base station in sections 3.1 and 3.2, section 3.3 presents

the main results obtained in Massive MIMO systems. Finally section 3.5 concludes the "Massive MIMO" with potential future directions of research.

Chapter 2

Bandwidth Sharing for Relaying in Cellular Systems

2.1 Introduction

As mentioned in the introduction, one of the key technology components of next generation cellular systems is relaying, which has been shown to provide better throughput and increased coverage [6, 7]. Typically, relays help in forwarding data and based on their roles, they have been categorized into two types. The first type is used exclusively to extend the coverage to remote User Equipment (UE), beyond the service range of the base station (aka eNodeB in LTE-A jargon). These are called type-I relays by LTE-A specifications and non-transparent relays in IEEE802.16j specifications. One of the applications of "coverage extending" low powered relays is to provide coverage to indoor or office environments where the signal strength is weak. Apart from providing extended coverage, they also help in deployment of cells in areas where the cost of wired backhaul is prohibitive. On the other hand, the second category of relays, are used to help the UE within the service range of base station to improve its service quality and link capacity. These are called type-II relays by LTE-A or transparent relays in IEEE802.16j.

When relaying terminals aid the UE, they incur costs in the form of power expenditure and usage of relay bandwidth. To compensate the relay for these costs, some well known approaches include reputation based mechanisms, credit based incentives and mechanisms based on forwarding games [10]. A novel approach of bandwidth exchange was introduced in [11] in order to compensate the relay for its incurred costs. Here, the relay node is offered a portion of bandwidth of the destination node as a compensation for forwarding data. The relay can use the compensated bandwidth for purposes that it deems fit, while ensuring forwarding the data of the destination. It was shown

that bandwidth exchange provides significant rate gains and improved coverage areas. In [14], a weighted power minimization problem was formulated for joint power, subcarrier allocation and subframe scheduling for downlink in-band [15] relaying systems with a focus on developing efficient algorithms. Our thesis in contrast provides a theoretical insights into the nature of optimal solution in an out-band relaying system using which we intend to develop efficient algorithms.

In this work, we consider bandwidth sharing in the context of relaying in cellular systems and study optimization problems involving both bandwidth and transmit power under rate constraints. A key aspect of our work is that along with power, we consider the bandwidths allocated to links as optimization variables in the relaying system. We formulate a weighted power minimization problem under rate, bandwidth and power constraints assuming out-band relaying system [15] and develop theoretical insights into the nature of optimal solutions. We formulate the weighted power minimization problem for a system involving a base station and a relay with multiple UEs to be served. We also consider an in-band relaying system [15] in which the eNodeB-relay and relay-UE link use the same set of carrier frequencies but transmits at different time slots. We show that the problem of average power minimization in a time shared relaying system as in [6] can be reformulated to an equivalent bandwidth sharing problem and depending on the nature of power constraints the time sharing system performs worse or same as the bandwidth sharing system. We find that optimizing across power and bandwidth provides scope for better utilization of the available resources, such as minimizing the total power consumption by half while improving the coverage area of the relays. While the theoretical results of the relaying system apply to a wide range of systems which employ multicarrier schemes, the terminology and numerical results used to describe this work loosely conform to LTE-A standards.

2.2 System Model

We start with system with an eNodeB, a relay and N UEs or users who are to be served. The relay can be a dedicated tower or another user that can potentially act as an intermediate node to help users forward its data. We assume that the relays can

forward data simultaneously in uplink and downlink in different carrier frequencies. Our assumption is valid as the LTE-A and IEEE 802.16j standards [16] indeed has dual radio out-of-band relay system in which the relay nodes have two RF chains in order to transmit and receive simultaneously in different carriers. The system is allocated a fixed bandwidth B. The users $i \in \{1, 2, \ldots, N\}$ are subject to transmit at power below $P_{i,\max}$. Also, every UE needs to be served at a rate of $R^d_{i,\min}$ in the downlink and $R^u_{i,\min}$ in the uplink. The link gains are of the form $\rho_{ij} = \kappa d_{ij}^{-\beta}$, where κ is the proportionality constant, d_{ij} is the distance between the i^{th} transmitter and j^{th} receiver and the pathloss exponent is $\beta = 3$. The noise power spectral density is also absorbed into the constant κ that is used to calculate the link gain. In general, any variable with subscript $(\cdot)_{ij}$ corresponds to the variable across the directed link (i, j) from transmitter i to receiver j. We consider that the link gains $\rho_{ij} = \rho_{ji}$ for any links (i, j) and (j, i). Specifically, we denote the users with the subscript i, relay equipment by r and eNodeB by 0, as shown in Fig. 2.1. We look into cases when the relay is employed to help the users achieve its minimum data rate. We consider the transmit power at each node and the bandwidth allocated to each link as a variable to be optimized. Our objective is to minimize the weighted sum of power in the system. Optimizing over bandwidth is relevant as we are employing a multicarrier system. This is because in a multicarrier system like OFDMA the bandwidth allocated to each link is specified by a number of subcarriers. We assume that the subcarrier spacing is small enough so that the bandwidth variables can be approximated as to be continuous. We also assume in our work that the eNodeB does a centralized resource allocation of bandwidth and power.

The relay r can be employed in various possible modes such as, providing only downlink access to users, or providing only uplink access to users, or providing both uplink and downlink access to users and also other intermediate scenarios such as providing uplink support to some users while only downlink to other users. Along with the cases above the relay r may have uplink information to send to the eNodeB, or it may have downlink information to receive for its own. For example, the downlink and uplink information for relay r could be the control information needed for dedicated relaying. In this particular work, we formulate weighted sum power minimization problem when

the relay is serving multiple users as shown in Fig. 2.2 and encompasses all possible cases discussed above. We assume that the eNodeB has the knowledge of channel gain between itself to the relays and between relays to UEs and solves the optimization problem in a centralized manner. Some special cases of the power minimization problem which which are of interest include:

1. The out-band relay r helps the users in forwarding both their downlink and uplink data.

2. The out-band relay r helps the users in forwarding both their downlink and uplink data along with transmitting its own data to the eNodeB.

3. The in-band relay r helps the users in forwarding both their uplink and downlink data.

4. The out-band relay r helps in forwarding both uplink and downlink data of a single user.

In this chapter, these special cases are identified in *Problems (2.1), (2.2), (2.23)* and *(2.18)*. We consider the case of single user as a separate section 2.4 as it provides some intuition on how the bandwidth is allocated in the uplink and downlink. Table 2.1 provides brief descriptions of these problems for quick reference.

A more generic system model arises when the eNodeB is serving directly some UEs and is also be connected to UEs through more than a single relay. The system model in that case might be represented as a tree structure with eNodeB as the root, the UEs as leaf nodes and the relays as intermediate nodes. Although the mathematical formulation and evaluation of the results is cumbersome, this is a straightforward extension of a single hop scenario in the sense that the insights revealed by the weighted power minimization for a multiple relay scenario is qualitatively similar to the case of single relay system. Section 2.3.4, elaborates on this situation with a example.

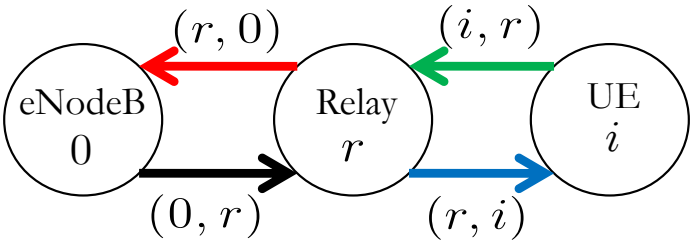

Figure 2.1: Relay acts as forwarder in both directions: This corresponds to *Problem (2.18)* where a single relay serves a user in both uplink and downlink directions.

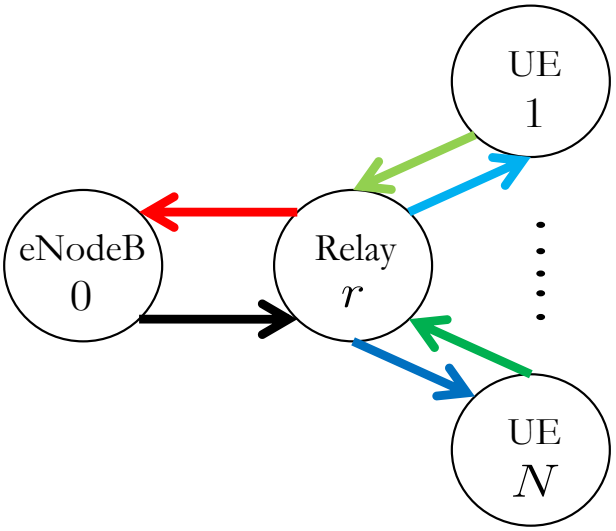

Figure 2.2: Relaying for multiple users: This corresponds to *Problem (2.1)* where a single relay serves multiple users in both uplink and directions.

2.3 Single Relay serving Multiple Users

In this section we formulate a power minimization problem when a base station serving multiple users equipments through a dedicated relay. We present the analytical results and aided by the analytical results we then perform a numerical evaluation to quantify the power savings using relays.

2.3.1 Relaying in Uplink and Downlink to Multiple User Equipments

We consider the case when a relay helps in forwarding data to N users in both uplink and downlink directions. In terms of the power P_{ir}, bandwidth W_{ir}, the signal to noise

Problem	**Description**	**# of relays**	**# of UEs**	**Max. Power Constraint**
(2.1)	Bidirectional Relaying	1	N	Yes
(2.2)	Bidirectional Relaying - with relay uplink data	1	N	Yes
(2.23)	Bidirectional Relaying - Time Sharing	1	N	Yes
(2.18)	Bidirectional Relaying	1	1	Yes
(2.1R)	Bidirectional Relaying - convex formulation	1	N	Yes
(2.4)	Bidirectional Relaying - dual formulation	1	N	No
(2.5)	Link specific optimization problem	1	N	No
(2.13)	Multiple Relays & multiple UEs	M	N	No

Table 2.1: For each numbered optimization problems this table gives a brief description of the problem for convenience. Here bidirectional means that the relay is supporting data in both downlink and uplink for the UEs.

ratio given by

$$\mathsf{SNR}_{ir} \triangleq \rho_{ir} P_{ir} / W_{ir}$$

and the rate function defined as

$$\mathcal{R}(\mathsf{SNR}_{ir}, W_{ir}) \triangleq W_{ir} \log(1 + \mathsf{SNR}_{ir})$$

across the corresponding links (i, r), we formulate the problem of weighted power minimization as

minimize

$$\alpha_0 P_{0r} + \alpha_r \left(\sum_{i=1}^{N} P_{ri} + P_{r0} \right) + \sum_{i=1}^{N} \alpha_i P_{ir} \tag{2.1a}$$

subject to

$$\mathcal{R}(\mathsf{SNR}_{ir}, W_{ir}) \geq R^u_{i,\min} \quad i = 1, \ldots, N, \tag{2.1b}$$

$$\mathcal{R}(\mathsf{SNR}_{r0}, W_{r0}) \geq \sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{ir}, W_{ir}), \tag{2.1c}$$

$$\mathcal{R}(\mathsf{SNR}_{ri}, W_{ri}) \geq R^{d}_{i,\min} \quad i = 1, \ldots, N, \tag{2.1d}$$

$$\mathcal{R}(\mathsf{SNR}_{0r}, W_{0r}) \geq \sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{ri}, W_{ri}), \tag{2.1e}$$

$$\sum_{i=1}^{N} (W_{ir} + W_{ri}) + W_{r0} + W_{0r} \leq B, \tag{2.1f}$$

$$P_{r0} + \sum_{i=1}^{N} P_{ri} \leq P_{r,\max}, \tag{2.1g}$$

$$P_{0r} \leq P_{0,\max}, \tag{2.1h}$$

$$P_{ir} \leq P_{\mathrm{i,max}} \quad i = 1, \ldots, N, \tag{2.1i}$$

variables

$$P_{r0}, P_{0r}, P_{ir}, P_{ri}, W_{r0}, W_{0r}, W_{ir}, W_{ri} \quad i = 1, \ldots, N$$

In *Problem (2.1)*, B is the total available bandwidth and $\alpha_0, \alpha_i, \alpha_r$ are the weights associated with the power spent by the nodes with

$$\alpha_0 + \sum_{i=1}^{N} \alpha_i + \alpha_r = 1.$$

For example, when $\alpha_r = 1$ the problem is just relay power minimization and when

$$\alpha_0 = \alpha_i = \alpha_r = 1/(N+2)$$

implies that the power of all the nodes are equally important and we are interested in system power minimization. Other values gives the engineer flexibility to design the system as required. Here, the constraints (2.1b), (2.1c) and (2.1d), (2.1e) represent the minimum rate requirement in the uplink and downlink simultaneously. By the perspective of a function argument [17, chapter 3], the rate constraints (2.1b), (2.1d) are convex in (P_{ir}, W_{ir}). The limited power constraints are represented by equations (2.1g), (2.1h) and (2.1i). Here, constraint (2.1g) implies that the relay has to use its limited power to provide the required rate in both the uplink and downlink directions for all users. $P_{r,\max}$ is the maximum power constraint on the relay r.

In our formulation we also consider bandwidth to be a variable that is to be optimized across the links. This is represented by the constraint (2.1f). Previous attempts

in joint bandwidth and power optimization have been considered in the context of maximizing the achievable rates in relaying systems in [9]. In this problem we have a relay r which is helping forward data to N users in both uplink and downlink directions. The relaying problem is not convex because of the constraints (2.1c) and (2.1e). Although the problem is not convex we will later show that the optimal solution is same as that of an equivalent convex problem which enables us to exploit the advantages offered by convexity.

An important variant of the *Problem (2.1)* arises when the relay along with forwarding the data of the users has its own data to be sent to the eNodeB at some required rate. For example, this scenario arises when a relay has to allocate some resources (bandwidth and power) to send its control data at a rate of at least $R^u_{r,\min}$. The power minimization problem can then be formulated by replacing the constraint (2.1c) in *Problem (2.1)* by

$$\mathcal{R}(\mathsf{SNR}_{r0}, W_{r0}) \geq \sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{ir}, W_{ir}) + R^u_{r,\min} \tag{2.2}$$

We denote this problem as *Problem (2.2)*.

2.3.2 Results and Discussion

Lemma 2.3.1. *When the constraint set is feasible, an optimal solution to the* Problem (2.1) *has the following characteristics:*

- *The rate constraints given by (2.1b), (2.1c), (2.1d), (2.1e) for* Problem (2.1) *are tight at optimum.*
- *The bandwidth constraint given by (2.1f) for* Problem (2.1) *is tight at the optimum.*

Proof. We will prove that the rate constraints are always tight for *Problem (2.1)* at optimum by a simple contradiction statement. Assume that for *Problem (2.1)*, the constraint (2.1b) is slack at optimum. Therefore,

$$W^\star_{ir} \log\left(1 + \frac{\rho_{ir} P^\star_{ir}}{W^\star_{ir}}\right) > R^u_{i,\min}$$

However, if $P_{ir}^{\star}$ exists then there always exists $P_{ir} < P_{ir}^{\star} \leq P_{i,\max}$ such that

$$W_{ir}^{\star} \log\left(1 + \frac{\rho_{ir} P_{ir}}{W_{ir}^{\star}}\right) = R_{i,\min}^{u},$$

and also has a lesser objective value. Therefore, P_{ir} is the optimum and not $P_{ir}^{\star}$ and at the optimum value the rate constraint is tight. This is true for all rate constraints (2.1b) and (2.1d) in *Problem (2.1)*. For constraint (2.1c), since

$$W_{ir}^{\star} \log\left(1 + \frac{\rho_{ir} P_{ir}^{\star}}{W_{ir}^{\star}}\right) = R_{i,\min}^{u}, \quad \forall i$$

and using the same arguments as before we can say that at optimum,

$$W_{r0}^{\star} \log\left(1 + \frac{\rho_{r0} P_{r0}^{\star}}{W_{r0}^{\star}}\right) = \sum_{i=1}^{N} R_{i,\min}^{u}.$$

Similarly, for *Problem (2.1)* we have at optimum,

$$W_{0r}^{\star} \log\left(1 + \frac{\rho_{0r} P_{0r}^{\star}}{W_{0r}^{\star}}\right) \quad = \quad \sum_{i=1}^{N} R_{i,\min}^{d}.$$

Again, we prove that the bandwidth constraint is tight at optimum by contradiction. Assume that the bandwidth constraint is not tight at optimum. Therefore,

$$\sum_{i=1}^{N} (W_{ir}^{\star} + W_{ri}^{\star}) + W_{r0}^{\star} + W_{0r}^{\star} < B.$$

Consider the constraint (2.1b) and from the earlier conclusion we know that at optimum the rate constraints are tight. Therefore,

$$P_{1r}^{\star} = W_{1r}^{\star} \left(e^{\frac{R_{1,\min}^{u}}{W_{1r}^{\star}}} - 1\right) \tag{2.3}$$

for user 1. Assume the optimal objective of the *Problem (2.1)* is $p^{\star}$. However, clearly we can find another bandwidth allocation $W_{1r}^{\star} + \delta W_{1r}$ with $\delta W_{1r} > 0$ such that the corresponding power P_{1r} consumed to meet the rate $R_{1,\min}^{u}$ is less than $P_{1r}^{\star}$ and

$$\sum_{i=1}^{N} (W_{ir}^{\star} + W_{ri}^{\star}) + W_{r0}^{\star} + W_{0r}^{\star} + \delta W_{1r} = B.$$

Consequently the new objective value $p < p^{\star}$. This implies that our initial assumption that $p^{\star}$ is the optimum is not true and any situation where the bandwidth constraint is slack does not give the optimum solution. Hence, if an optimum solution exist for the *Problem (2.1)* then at that solution the bandwidth constraint has to be tight. $\square$

As a consequence of Lemma 2.3.1 we can formulate an equivalent convex formulation for the *Problem (2.1)*. This is explained in the following theorem.

Theorem 2.3.2. *The optimal solution to* Problem (2.1) *is the same as that of the optimal solution of* Problem (2.1R) *in which the constraints (2.1c) and (2.1e) are replaced by,*

$$\mathcal{R}(\mathsf{SNR}_{r0}, W_{r0}) \geq \sum_{i=1}^{N} R^{u}_{i,\min}, \tag{1R-c}$$

$$\mathcal{R}(\mathsf{SNR}_{0r}, W_{0r}) \geq \sum_{i=1}^{N} R^{d}_{i,\min}, \tag{1R-e}$$

respectively.

Proof. Let $p^\star$ denote the optimal objective to *Problem (2.1)* and $q^\star$ the optimal objective to the reformulated *Problem (2.1R)* with the constraint (2.1c) replaced by (1R-c) and (2.1e) replaced by (1R-e). It follows that the feasible set of *Problem (2.1)* is a subset of the feasible set of *Problem (2.1R)*. Therefore, $p^\star \geq q^\star$. However, from Lemma 2.3.1 since the rate constraints and the bandwidth constraints are tight at optimum, the optimal solution to *Problem (2.1R)* is a feasible solution to *Problem (2.1)*. This implies that there is feasible solution for *Problem (2.1)* with objective value $q^\star$. As $p^\star$ is the optimal solution to *Problem (2.1)*, by definition of an optimal solution $q^\star \geq p^\star$. Therefore, $q^\star = p^\star$ and optimal points for both problems are identical. □

Theorem 2.3.2 shows that the relaying problem for multiple UEs, which is non-convex by nature can be solved optimally by an equivalent convex formulation. The convex formulation allows the optimal solution to be obtained numerically. Although convex, we have only limited analytical insights into the nature of the optimal solution because of the power constraints (2.1g), (2.1h), (2.1i) on the individual nodes. However, under most problem instances specified by relay and UE positions in the cellular system, the nodes will be transmitting strictly below maximum power. In the next section we look into the nature of the optimal solution assuming that the power constraints are slack at optimum. The solution obtained to *Problem (2.1)* when the power constraints are slack are the same as when we assume that each of the nodes have unlimited maximum power.

2.3.3 Power Minimization - Unlimited Maximum Power

In this section we consider a situation when the maximum power constraints in *Problem (2.1)* are eliminated. This is the same as assuming that each of the nodes have unlimited maximum power at their disposal. Henceforth we denote *Problem (2.1R)* as *Problem (2.1)* as both have the same optimal solution. We look into this problem to get some insight into the optimal solution in situations when none of the power constraints are tight at optimum. Writing the partial Lagrangian with respect to the bandwidth constraint (2.1f) in *Problem (2.1R)* results in

$g(\lambda) =$ **minimize**

$$\alpha_0 P_{0r} + \alpha_r \left(\sum_{i=1}^{N} P_{ri} + P_{r0} \right) + \sum_{i=1}^{N} \alpha_i P_{ir} + \lambda \left(\sum_{i=1}^{N} (W_{ir} + W_{ri}) + W_{r0} + W_{0r} - B \right) \tag{2.4a}$$

subject to

$$\mathcal{R}(\mathsf{SNR}_{ir}, W_{ir}) \geq R^u_{i,\min} \quad \forall i = 1, \ldots, N, \tag{2.4b}$$

$$\mathcal{R}(\mathsf{SNR}_{0r}, W_{0r}) \geq \sum_{i=1}^{N} R^d_{i,\min}, \tag{2.4c}$$

$$\mathcal{R}(\mathsf{SNR}_{ri}, W_{ri}) \geq R^d_{i,\min} \quad \forall i = 1, \ldots, N, \tag{2.4d}$$

$$\mathcal{R}(\mathsf{SNR}_{r0}, W_{r0}) \geq \sum_{i=1}^{N} R^u_{i,\min}, \tag{2.4e}$$

where, we have assumed that there are no maximum power constraints. The partial Lagrangian helps us to decouple the problem on a link by link basis where previously it was coupled across the links because of the bandwidth constraint (2.1f). For a given $\lambda \geq 0$ we can split the problem into $2N+2$ different *Link Specific Optimization* problems of the form

$$\textbf{minimize} \quad \alpha_l P_l + \lambda W_l, \tag{2.5a}$$

$$\textbf{subject to} \quad \mathcal{R}(\mathsf{SNR}_l, W_l) \geq R_l, \tag{2.5b}$$

$$\textbf{variables} \quad P_l, W_l, \tag{2.5c}$$

corresponding to each directed link $l \in \{(i,r),(r,i)\}$ in the system. The parameters α_l, ρ_l, R_l are link specific and α_l corresponds to the weights in the weighted power minimization and is associated with the transmitter of link l, ρ_l corresponds to the link gain, and R_l is the minimum rate that needs to be supported on link l. For example, in the link optimization problem corresponding to link $(0,r)$, i.e. downlink from the base station node 0 and relay node r, the parameters correspond to $\alpha_l = \alpha_0$, $\rho_l = \rho_{0r}$ and $R_l = \sum_{i=1}^{N} R^d_{i,\min}$.

Since strong duality holds for *Problem (2.1)* we have at $\lambda = \lambda^\star$, that each *Link Optimization Problem (2.5)* gives the primal optimal points. Lemma 2.3.1 implies that the rate constraints are tight. Thus *Problem (2.5)* is the same as

$$\min_{W_l} \alpha_l \frac{W_l}{\rho_l}\left(\exp\left(\frac{R_l}{W_l}\right) - 1\right) + \lambda^\star W_l$$

Therefore, at optimality for every link l in the system

$$\exp\left(\frac{R_l}{W_l}\right)\left(\frac{R_l}{W_l} - 1\right) + 1 - \rho_l\lambda^\star/\alpha_l \;\; = \;\; 0. \tag{2.6}$$

From equation (2.6) we can gather that the spectral efficiency R_l/W_l depends only the term $\rho_l\lambda^\star/\alpha_l$. This suggests that all the three factors - ρ_l, the link gain, λ, the dual bandwidth cost, and α_l, the node weight decides the spectral efficiency of the link l. This specific property of the *Problem (2.5)* enable us to prove the following claim specified as Lemma 2.3.3. Also, equation (2.6) along with bandwidth constraint provide $N+3$ non-linear equations in $N+3$ variables to solve the optimization problem.

Lemma 2.3.3. ***Symmetry of Spectral Efficiency:*** *When the weights $\alpha_0 = \alpha_r = \alpha_i = \alpha$ are the same then the optimal solution to* Problem (2.1) *satisfies*

$$\frac{\sum_{i=1}^{N} R^d_{i,\min}}{W^\star_{0r}} \;\; = \;\; \frac{\sum_{i=1}^{N} R^u_{i,\min}}{W^\star_{r0}}, \tag{2.7}$$

$$\frac{R^d_{i,\min}}{W^\star_{ri}} \;\; = \;\; \frac{R^u_{i,\min}}{W^\star_{ir}} \quad \forall i. \tag{2.8}$$

Proof. The optimality conditions (2.6) imply

$$\begin{aligned} &\exp\left(\frac{\sum_{i=1}^{N} R^d_{i,\min}}{W^\star_{0r}}\right)\left(\frac{\sum_{i=1}^{N} R^d_{i,\min}}{W^\star_{0r}} - 1\right) \\ &= \exp\left(\frac{\sum_{i=1}^{N} R^u_{i,\min}}{W^\star_{r0}}\right)\left(\frac{\sum_{i=1}^{N} R^u_{i,\min}}{W^\star_{r0}} - 1\right) \end{aligned} \tag{2.9}$$

because, the links $l \in \{(0,r),(r,0)\}$ have equal weights $\alpha_l = \alpha$ and equal link gain $\rho_l = \rho_{0r}{=}\rho_{r0}$. Since $e^x(x-1)$ is monotonic and positive derivative for all $x \geq 0$, (2.9) implies,

$$\frac{\sum_{i=1}^{N} R^d_{i,\min}}{W^\star_{0r}} = \frac{\sum_{i=1}^{N} R^u_{i,\min}}{W^\star_{r0}}.$$

Similarly, using equation (2.6) corresponding to the links (r,i) and (i,r), we obtain

$$\frac{R^d_{i,\min}}{W^\star_{ri}} = \frac{R^u_{i,\min}}{W^\star_{ir}} \tag{2.10}$$

for all $i = 1,\ldots,N$. □

Lemma 2.3.1 also implies that,

$$\frac{\sum_{i=1}^{N} R^d_{i,\min}}{W^\star_{0r}} = \frac{\sum_{i=1}^{N} R^u_{i,\min}}{W^\star_{r0}} = \log(1+\frac{\rho_{0r}P^\star_{0r}}{W^\star_{0r}}) = \log(1+\frac{\rho_{r0}P^\star_{r0}}{W^\star_{r0}}), \tag{2.11}$$

where $\log(1+\rho_{0r}P^\star_{0r}/W^\star_{0r})$ is the spectral efficiency of the link $(0,r)$. Therefore, $\sum_{i=1}^{N} R^d_{i,\min}/W^\star_{0r}$ represents the spectral efficiency of the link $(0,r)$, where $\sum_{i=1}^{N} R^d_{i,\min}$ is the rate that needs to be supported in that link. Similarly, $\sum_{i=1}^{N} R^u_{i,\min}/W^\star_{r0}$ represent the spectral efficiency of the link $(r,0)$ where $\sum_{i=1}^{N} R^u_{i,\min}$ is the rate that needs to supported in that link.

Corollary 2.3.4. *When the weights $\alpha_0 = \alpha_r = \alpha_i = \alpha$ are the same, the optimal solution to* Problem (2.2) *has the property that*

$$\frac{\sum_{i=1}^{N} R^d_{i,\min}}{W^\star_{0r}} = \frac{\sum_{i=1}^{N} R^u_{i,\min} + R^u_{r,\min}}{W^\star_{r0}}. \tag{2.12}$$

Proof. The result follows from Lemma 2.3.3. □

The *Symmetry of Spectral Efficiency* is preserved even when the relay has its own data to deliver to the eNodeB. As a consequence of Lemma 2.3.3 we can reduce the number of equations in the optimality conditions (2.6) by half from $2N+2$ to $N+1$.

2.3.4 Extension to Multihop Relays - An Example

Earlier in section 2.2 we had mentioned that the results that will be obtained for a single relay can be extended to a generic system model represented by a tree structure with

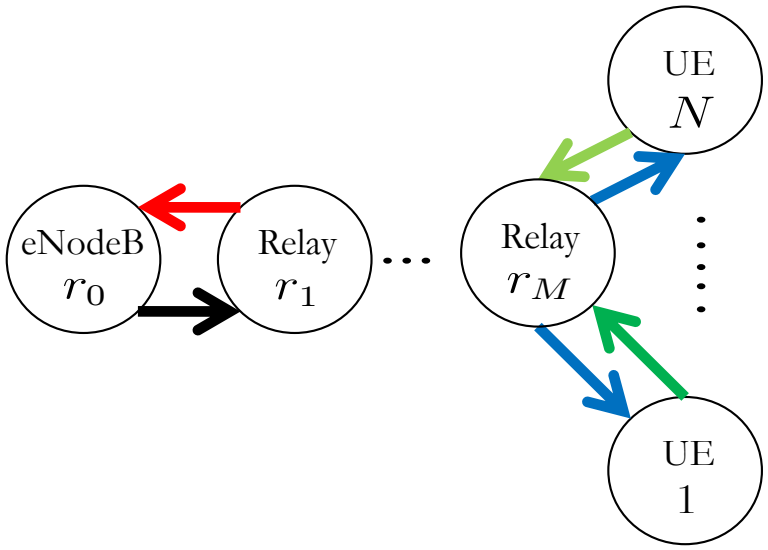

Figure 2.3: Example of a multihop relaying system

eNodeB as the root, the UEs being the leaf nodes and the relays as the intermediate nodes. This follows from the above fact that we can formulate a link specific optimization problem by taking the Lagrangian dual with respect to the bandwidth constraint when we know the minimum rate that each link needs to transmit. This rate depends upon the specific tree structure i.e. the route from eNodeB to different UEs through the relays and the rate demanded in uplink and downlink by each UEs. Fig. 2.4 shows the example of a three sector single cell system with one relay per sector serving the edge UEs as in LTE-A system. The interior users are those who lie within a radius of r_{int} from the eNodeB and others which lie outside are the exterior users. We assume that the interior UEs are served directly by eNodeB and the exterior UEs in a sector are served by their corresponding relays. By defining the dual bandwidth cost $\lambda \geq 0$ for the bandwidth constraint, this system can also be decomposed into link specific optimization problems. In fact the same principle can be extended to a mesh network where given that the routes from different sources to its destination are defined, we can decompose the network into link specific optimization problems which are as many in number as the number of links in the system, for bandwidth and power optimization. However, we focus on the single relay system because as the main problem in larger networks is the identification of the optimal relay node or the optimal route to reach the

destination. Also, the insights for the single relay system which, hinge on Lagrangian dual of the bandwidth constraint, can be easily extended into multihop networks. As a specific and easy example, consider a set of N users that needs to served by a M relays as shown in Fig. 2.3 (with all the users being served by the M^{th} relay). If r_j is the index of relays with $j = 1, 2, \ldots, M$, $P_{r_j r_{j+1}}$ the power transmitted by the j^{th} relay to $j + 1^{th}$ relay and $W_{r_j r_{j+1}}$ the bandwidth allocated to the link (r_j, r_{j+1}), the optimization problem can be formulated as, *Problem (2.13)*

minimize

$$\sum_{j=0}^{M-1} \left(\alpha_{r_j} P_{r_j, r_{j+1}} + \alpha_{r_{j+1}} P_{r_{j+1}, r_j}\right) + \alpha_{r_M} \sum_{i=1}^{N} P_{r_M i} + \sum_{i=1}^{N} \alpha_i P_{i r_M} \tag{2.13a}$$

subject to

$$\mathcal{R}(\mathsf{SNR}_{i r_M}, W_{i r_M}) \geq R^u_{i,\min} \quad i = 1, \ldots, N, \tag{2.13b}$$

$$\mathcal{R}(\mathsf{SNR}_{r_M i}, W_{r_M i}) \geq R^d_{i,\min} \quad i = 1, \ldots, N, \tag{2.13c}$$

$$\mathcal{R}(\mathsf{SNR}_{r_{j+1}, r_j}, W_{r_{j+1}, r_j}) \geq \sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{i r_M}, W_{i r_M}) \qquad j = 0, \ldots, M-1, \tag{2.13d}$$

$$\mathcal{R}(\mathsf{SNR}_{r_j, r_{j+1}}, W_{r_j, r_{j+1}}) \geq \sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{r_M i}, W_{r_M i}) \qquad j = 0, \ldots, M-1, \tag{2.13e}$$

$$\sum_{i=1}^{N} (W_{i r_M} + W_{r_M i}) + \sum_{j=0}^{M-1} (W_{r_j, r_{j+1}} + W_{r_{j+1}, r_j}) \leq B, \tag{2.13f}$$

variables

$$P.\text{'s}, W.\text{'s}$$

The equations (2.13b), (2.13c) represent the rate constraints on the uplink and downlink between the relay r_M and UEs. The rate constraints (2.13d), (2.13e) represent the relay to relay rate constraint requirements in which each relay has to support the end UEs data rate requirement in uplink and downlink. The equation (2.13f) represent the system bandwidth constraint. For the case when $M = 1$ this reduces to the *Problem (2.1)*. The conclusions made in lemma 2.3.1 and theorem 2.3.2 for a single relay system can be extended to *Problem (2.13)* and based on the Lagrangian dual of the bandwidth constraint lemma 2.3.3 can also be derived.

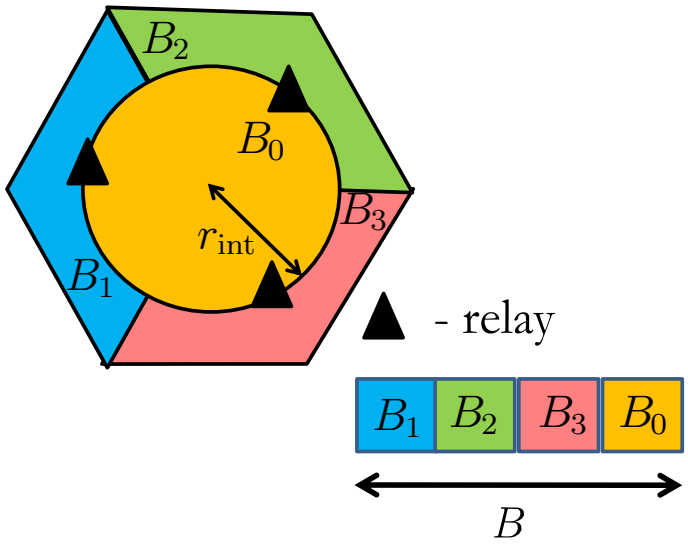

Figure 2.4: Three sector single cell LTE-A system with one relay per sector. The UEs are classified as interior users and exterior users. The interior users are those who lie within a radius of r_{int} from the eNodeB and others which lie outside are the exterior users. The interior UEs are served directly by eNodeB and the exterior users in a sector are served by their corresponding relays. The relays in each sector are located at a point of intersection of edge of the core radius circle and line which bisects the sector into two equal halves.

2.3.5 Numerical Results for Downlink

In this section we provide the numerical simulation results to a N-user downlink scenario and quantify the power saving due to the introduction of relays in a cellular radio system. We consider a LTE-A system with strict Fractional Frequency Reuse (FFR) scheme. Fig. 2.4 shows the diagram of a three sector single cell LTE-A system with one relay per sector. The N users are assumed to be uniformly distributed within the cell and they are classified as interior users and exterior users. The interior users are those who lie within a radius of r_{int} from the eNodeB and others which lie outside are the exterior users. We assume that exterior users in a sector are served by their corresponding relays. The relays in each sector are located at a point of intersection of edge of the core radius circle and line which bisects the sector into two equal halves. The total system bandwidth B is split into four chunks B_0, B_1, B_2 and B_3 as per the LTE-A strict FFR schemes [22], [23]. B_0 is the total bandwidth allocated to the cell interior users and B_1, B_2, B_3 corresponds to that allocated to the cell exterior users in

sector 1, 2, 3 respectively. In [22], the fraction of the total bandwidth B, allocated to the interior users (B_0) is based on the area covered by the interior radius and is given by

$$B_0 = BA_{\text{int}}/A_{\text{tot}}$$

as a result of the uniform distribution of the users, where

$$A_{\text{int}} = \pi r_{\text{int}}^2$$

is the interior area, A_{tot} is the area of the cell and

$$B_r = (B - B_0)/3.$$

The eNodeB allocates bandwidth to the interior users flexibly from B_0 and the relay r allocates bandwidth flexibly to the exterior users in the corresponding sector r from the band B_r. While it is clear that the interior users are allocated bandwidth from B_0 and the exterior users in each sector are served from B_r, $r = 1, 2, 3$, it is not apparent from where the bandwidth for the eNodeB-relay, $(0, r)$ links should be allocated. In one case, the three relays can be allocated bandwidth from B_0 reserved for the interior users as each relay can be viewed as another core user. While in another case the $(0, r)$ link can be allocated bandwidth from B_r as each relays serves only the exterior users. As it is not clear which is better in terms of minimizing the total power in the system, we classify those into two allocation methods,

I The $(0, r)$ links for $r = 1, 2, 3$ are allocated bandwidth from B_0.

II The $(0, r)$ link is allocated bandwidth from the exterior user bandwidth B_r.

In method-I allocation scheme the eNodeB to the interior users link have to operate at higher powers so as to compensate for the bandwidth lost to the $(0, r)$ links. In contrast, for Method-II B_0 is used exclusively reserved to the interior users and B_r is used to support the sum of the data rate of the exterior users served by the relay r in the $(0, r)$ link in addition to the exterior users. This leads to a increase in total power consumption for the exterior users. We compare the above two possible schemes

using the bandwidth partition based on [22] for varying interior radius. The method-I consumes more power than method-II when the interior area is small as compared to the exterior area. This is because when r_{int} is small B_0 will be consequently less and with the less availability of bandwidth the eNodeB has to support both the interior users and the three relays. Coupled with the less availability of B_0, since the number of users classified as exterior is larger in number and that the $(0, r)$ links also has to support the sum of the exterior user rates makes method-I consume more power than method-II. Method-II on the other hand clearly consumes more power when the interior area is large as compared to the exterior area. As the interior area is large, B_r is very low and more power is required to satisfy the both $(0, r)$ link (whose link gain is weaker as the relay is in the edge of interior radius) and the corresponding exterior users in sector-r. This obvious disadvantage of the simple area based bandwidth partitioning can be mitigated by considering some additional bandwidth for $(0, r)$ links given to B_0 or B_r depending on whether we employ method-I or method-II. This is given in the next section.

Area Based Bandwidth Partition

Let δ be the expected number of users per unit area. Therefore the expected number of users in the interior region can be given as

$$n_0 = \delta A_{\text{int}},$$

and the expected number of users served by each relay can be given as

$$n_1 = n_2 = n_3 = \delta(A_{\text{tot}} - A_{\text{int}})/3.$$

Therefore, the expected number of links in the system is given by

$$n_0 + 2\sum_{r=1}^{3} n_r$$

with one link per directly-served users and two links per relay-served users. The bandwidth allocation B_0 is proportional to the number of links that is served by the eNodeB. For method-I we have,

$$B_0 = B\frac{n_0 + \sum_{r=1}^{3} n_r}{n_0 + 2\sum_{r=1}^{3} n_r} = B\frac{1}{2 - A_{\text{int}}/A_{\text{tot}}} \tag{2.14}$$

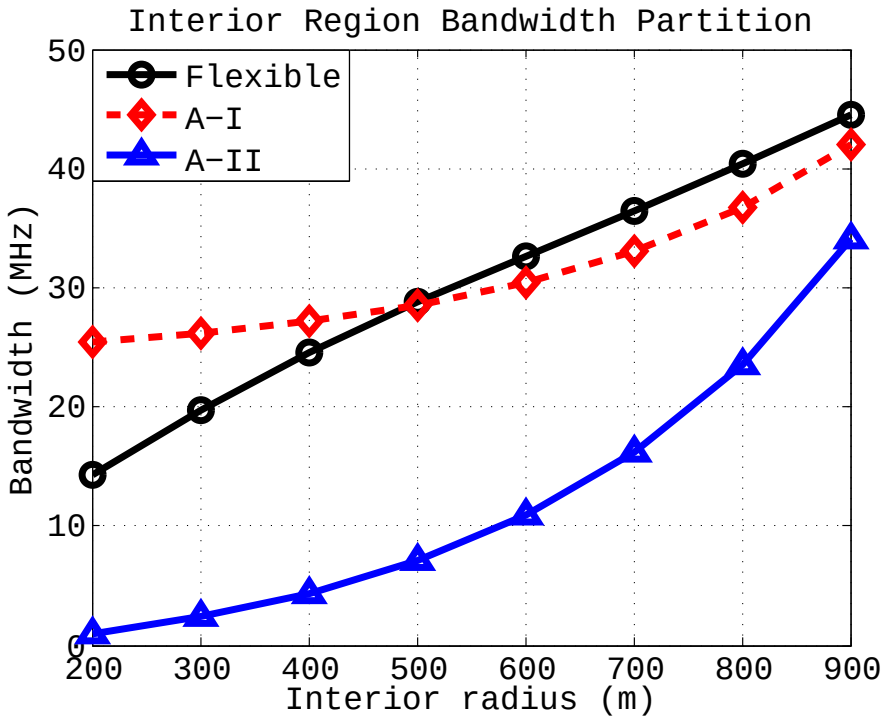

Figure 2.5: The plot shows the bandwidth partitioned for the interior region with link based, area based as well as flexible allocation. For the flexible partition the sum of bandwidth allocated to the interior users and the eNodeb to relay links is plotted.

and,

$$B_r = \frac{B}{3}\frac{1 - A_{\text{int}}/A_{\text{tot}}}{2 - A_{\text{int}}/A_{\text{tot}}} \tag{2.15}$$

We denote the above partition "A-I" which stands for area based bandwidth partition using method-I. Similarly, for method-II we have,

$$B_0 = B\frac{n_0}{n_0 + 2\sum_{r=1}^{3} n_i} = B\frac{A_{\text{int}}/A_{\text{tot}}}{2 - A_{\text{int}}/A_{\text{tot}}} \tag{2.16}$$

$$B_r = \frac{2B}{3}\frac{1 - A_{\text{int}}/A_{\text{tot}}}{2 - A_{\text{int}}/A_{\text{tot}}} \tag{2.17}$$

We denote the above the partition as "A-II". Fig. 2.5 shows the plots of total bandwidth allocated to the interior region for different values of core radius. The bandwidth partition "A-I" allocates sufficient bandwidth to the interior region even for low interior radius by always considering that it is also used to support the $(0, r)$ links. Also, bandwidth for partition "A-II" reduces B_0 when the interior area is large and hence

ensures that there is sufficient bandwidth

$$B_r = (B - B_0)/3$$

to support the $(0, r)$ links. We then compare this with a completely flexible bandwidth allocation where B_0, B_1, B_2 and B_3 are also variables in the optimization process with

$$B_0 + \sum_{r=1}^{3} B_r \leq B.$$

In contrast to the area based bandwidth partition, the flexible bandwidth allocation scheme can be considered as a link gain based optimal allocation of bandwidth per links for minimizing the system power. Fig. 2.5 also shows the sum of bandwidth allocated to the direct links and eNodeB-relay links in a flexible scheme. It is observed that in A-I the link based partition allocates more than optimal bandwidth to the interior region when the interior radius is small but less than optimal bandwidth when the interior radius is large. We also implement a fixed bandwidth allocation where each active links is allotted a fixed bandwidth. In the fixed bandwidth allocation scheme B_0 is given by equation (2.16), B_r, $r = 1, 2, 3$ is partitioned as in equation (2.15) and each $(0, r)$ link is assigned a third of the rest which in turn is equal to B_r itself. Each of the links within the partitioned region is assigned a fixed bandwidth equal to the ratio of bandwidth in the region to the number of users in the region.

The simulations were done for 10 UEs in the system uniformly distributed in the cell. The total bandwidth availability is 50 MHz and the minimum downlink rate requirement is uniformly distributed between $0 - 10$ Mbps. Fig. 2.6 shows the average power consumption for A-I, A-II, flexible and fixed bandwidth allocation schemes. While is it clear that a completely flexible bandwidth allocation should consume the least power, it also achieves at least 3 dB reduction in average power than the fixed bandwidth allocation for all values of interior radius. This is a direct extension and stronger result as compared to the single user case where the gain was up to 3 dB (Fig. 2.9). At interior radius of 200 m the fixed bandwidth allocation scheme seems to consume much more power than for other values of interior radius. The behaviour is due to the fact that the bandwidth allocated to interior users B_0 according to equation (2.16) (same as "A-II'

curve in Fig. 2.5) is very low resulting in high power consumption. Although the bandwidth allocation is exactly same for "A-II" scheme the average power consumed is not as much as the fixed scheme because within the interior region bandwidth is flexibly allocated.

The average power consumption for area based method-I and II schemes lie between that of fixed and flexible schemes. They represent an intermediate and implementable method for LTE-A systems as compared to flexible allocation and fixed allocation schemes. The flexible allocation scheme may not be realistic to implement and fixed allocation results in at least 3 dB additional power consumption which could be avoided. The flexible scheme is unrealistic from the point of view that B_0 and $B_r, r = 1, 2, 3$ are varying for every snapshot instance of the optimization problem while the FFR in LTE-A requires it to fixed. From Fig. 2.6, it is found there is less than 1 dB difference between the area based method-I and flexible allocation scheme for values of interior radius between 500 to 800 m. This can be attributed to the fact that in Fig. 2.5 there is not much difference in bandwidth allotted by the flexible scheme as compared with the method-I. The area based method-II scheme on the other hand has the average power consumption more than that of method-I primarily because of its inherent bandwidth partitioning technique. In method-II the B_0 allotted seems to be more than enough for its direct links while B_r is comparatively less to support the $(0, r)$ links as well as relay to UE links. A supporting viewpoint to this argument is that as the interior radius increases B_0 increases and difference in average power consumption between method-I and method-II increases as seen in Fig. 2.6.

2.4 Single Relay serving Single User

2.4.1 Relaying Problem for a Single User on Uplink and Downlink

For simplicity and understanding of the bandwidth sharing process we also consider the case of the weighted power minimization problem in relaying when there is only one user i.e., $N = 1$. The relay r helps in forwarding data for the UE i in both uplink and downlink direction as shown in the Fig. 2.1. We formulate the problem of minimizing

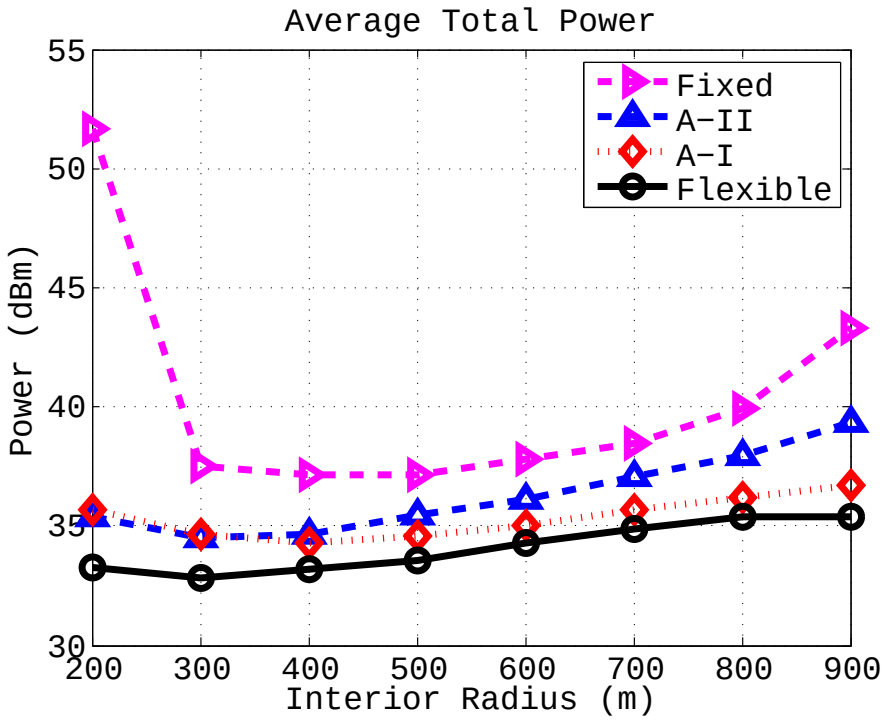

Figure 2.6: The plot the shows the average power consumption for four different schemes - Flexible, Fixed, A-I and A-II. It is seen that the flexible bandwidth allocation consumes at least 3 dB less power as compared to fixed bandwidth allocation.

the weighted sum of powers spent by the system in ensuring the minimum required data rates in both uplink and downlink for the UE i as

$$\begin{aligned}
\text{minimize} \quad & \alpha_0 P_{0r} + \alpha_i P_{ir} + \alpha_r (P_{ri} + P_{r0}) && \text{(2.18a)}\\
\text{subject to} \quad & \mathcal{R}(\mathsf{SNR}_{ir}, W_{ir}) \geq R^u_{i,\min}, && \text{(2.18b)}\\
& \mathcal{R}(\mathsf{SNR}_{r0}, W_{r0}) \geq R^u_{i,\min}, && \text{(2.18c)}\\
& \mathcal{R}(\mathsf{SNR}_{ri}, W_{ri}) \geq R^d_{i,\min}, && \text{(2.18d)}\\
& \mathcal{R}(\mathsf{SNR}_{0r}, W_{0r}) \geq R^d_{i,\min}, && \text{(2.18e)}\\
& W_{ir} + W_{0r} + W_{ri} + W_{r0} \leq B, && \text{(2.18f)}\\
& P_{ri} + P_{r0} \leq P_{r,\max}, && \text{(2.18g)}\\
& P_{0r} \leq P_{0,\max}, && \text{(2.18h)}\\
& P_{ir} \leq P_{i,\max}, && \text{(2.18i)}\\
\text{variables} \quad & P_{0r}, P_{ri}, P_{ir}, P_{r0}, W_{0r}, W_{ri}, W_{ir}, W_{0r}.
\end{aligned}$$

2.4.2 Total Power Minimization

We consider the special case when there is only one user ($N = 1$) in the system and develop an intuitive understanding of the bandwidth sharing mechanism. The weighted power minimization problem for a single user is given by *Problem (2.18)* but without the power constraints. We define the bandwidth allocated in the downlink for *Problem (2.18)* to be

$$B^d = W_{0r} + W_{ri}$$

and similarly that allocated in the uplink to be

$$B^u = W_{ir} + W_{r0}.$$

Theorem 2.4.1. ***Rate Proportional Bandwidth Allocation:*** *When the weights $\alpha_0 = \alpha_r = \alpha_i = \alpha$ are equal, the bandwidths allocated in the downlink and uplink at the optimum for* Problem (2.18) *are independent of the link gains and are given by*

$$B^{d\star} = \frac{R^d_{i,\min}}{R^u_{i,\min} + R^d_{i,\min}} B,$$
$$B^{u\star} = \frac{R^u_{i,\min}}{R^u_{i,\min} + R^d_{i,\min}} B.$$

Proof. By Lemma 2.3.1 for $N = 1$, we have at optimality

$$(W^\star_{0r} + W^\star_{r0}) + (W^\star_{ri} + W^\star_{ir}) = B.$$

Similarly for $N = 1$, using Lemma 2.3.3 implies

$$(W^\star_{0r} + W^\star_{ri}) \left(1 + \frac{R^u_{i,\min}}{R^d_{i,\min}}\right) = B.$$

Since $B^{d\star} = W^\star_{0r} + W^\star_{ri}$ the result follows. Similarly, we get $B^{u\star}$ from $B^{u\star} = B - B^{d\star}$. □

Thus, since the optimal bandwidth split between uplink and downlink is known, the *Problem (2.18)* can be split into two problems one for the downlink and other for the uplink direction. Also, the bandwidth allocated to the downlink problem out of the total available bandwidth, is proportional to that fraction of the downlink rate requirement out of the total rate.

Theorem 2.4.2. *When the weights $\alpha_0 = \alpha_r = \alpha_i = \alpha$ are equal, the optimal bandwidth allocated to each of the links for* Problem (2.18) *depends only on the ratio of link gains ρ_{0r}/ρ_{ri}.*

Proof. Using optimality conditions (2.6) for links $(0, r)$, (r, i) and Theorem 2.4.1 we get,

$$e^{\frac{R^d_{i,\min}}{W_{0r}}}\left(\frac{R^d_{i,\min}}{W_{0r}} - 1\right) + 1 = \beta\left\{e^{\frac{R^d_{i,\min}}{B^{d\star} - W_{0r}}}\left(\frac{R^d_{i,\min}}{B^{d\star} - W_{0r}} - 1\right) + 1\right\}. \quad (2.19)$$

where $\beta = \rho_{0r}/\rho_{ri}$. Since, the LHS is a decreasing function and the RHS is an increasing function in W_{0r}, the above equation has only one point of intersection which is dependent only on β. Therefore, the optimal W_{0r} only depends on the ratio of the link gains. Similarly, we get

$$e^{\frac{R^u_{i,\min}}{W_{r0}}}\left(\frac{R^u_{i,\min}}{W_{r0}} - 1\right) + 1 = \beta\left\{e^{\frac{R^u_{i,\min}}{B^{u\star} - W_{r0}}}\left(\frac{R^u_{i,\min}}{B^{u\star} - W_{r0}} - 1\right) + 1\right\}. \quad (2.20)$$

□

The optimal bandwidth allocations $W^\star_{0r}, W^\star_{r0}, W^\star_{ri}, W^\star_{ir}$ can be found numerically from equations (2.19) and (2.20).

2.4.3 Minimizing the Relay Power

There may also arise situations when the relay power is precious and hence we only intend to minimize the relay power. Again we show the results for a single user case but the results for this section are easily extendible for the case for multiple users.

Minimizing Relay Power for *Problem (2.18)*

We set the value of $\alpha_r = 1$, $\alpha_0 = 0$ and $\alpha_i = 0$, $i \neq r$ when we want to minimize the relay power. When $\alpha_r = 1$, $P^\star_{0r} = P_{0,\max}, P^\star_{ir} = P_{i,\max}$. Consequently, $W^\star_{0r}$ and $W^\star_{ir}$ will be the unique solution to,

$$\begin{aligned} W_{0r}\log\left(1 + \frac{\rho_{0r}P_{0,\max}}{W_{0r}}\right) &= R^d_{i,\min} \quad \text{and,} \\ W_{ir}\log\left(1 + \frac{\rho_{ir}P_{i,\max}}{W_{ir}}\right) &= R^u_{i,\min}, \end{aligned} \quad (2.21)$$

respectively. Intuitively, the eNodeB and the UEs transmit at the maximum power so that the least amount out of total available bandwidth is necessary to maintain the minimum rate from eNodeB to relay and UE to relay links. The rest of the bandwidth is used up by the relay so that it expends minimum power to maintain the minimum rate constraint in uplink and downlink. The relay power minimization problem turns out to be

minimize

$$P_{ri} + P_{r0} \tag{2.22a}$$

subject to

$$\mathcal{R}(\mathsf{SNR}_{r0}, W_{r0}) \geq R^{u}_{i,\min}, \tag{2.22b}$$

$$\mathcal{R}(\mathsf{SNR}_{ri}, W_{ri}) \geq R^{d}_{i,\min}, \tag{2.22c}$$

$$W_{ri} + W_{r0} \leq B - (W^{\star}_{0j} + W^{\star}_{ir}) = B', \tag{2.22d}$$

$$P_{ri} + P_{r0} \leq P_{r,\max}, \tag{2.22e}$$

variables

$$P_{ri}, P_{r0}, W_{r0}, W_{ri}.$$

This can be solved in a centralized manner at the relay.

2.4.4 Numerical Results

We solve the *Problem (2.18)* and plot the resulting bandwidth shared, and power consumed by the links for maintaining a downlink rate of 7 Mb/s with 10 MHz bandwidth. We fix the maximum eNodeB power at 43 dBm, and the maximum power limit for relay and UE at 23 dBm which is the standard for LTE-A systems as given in [20, 21]. Also, the type-I relay is fixed at a distance of 1000 m from the eNodeB. The bandwidth allocation and power consumption is plotted for increasing ρ_{0i}/ρ_{ri} (dB) values until the optimization problem becomes infeasible. Also, increasing sequence of ρ_{0i}/ρ_{ri} (dB) values indicate increasing distance between relay r and UE i when the eNodeB to relay distance is fixed. The optimization problem finds a feasible point at maximum relay

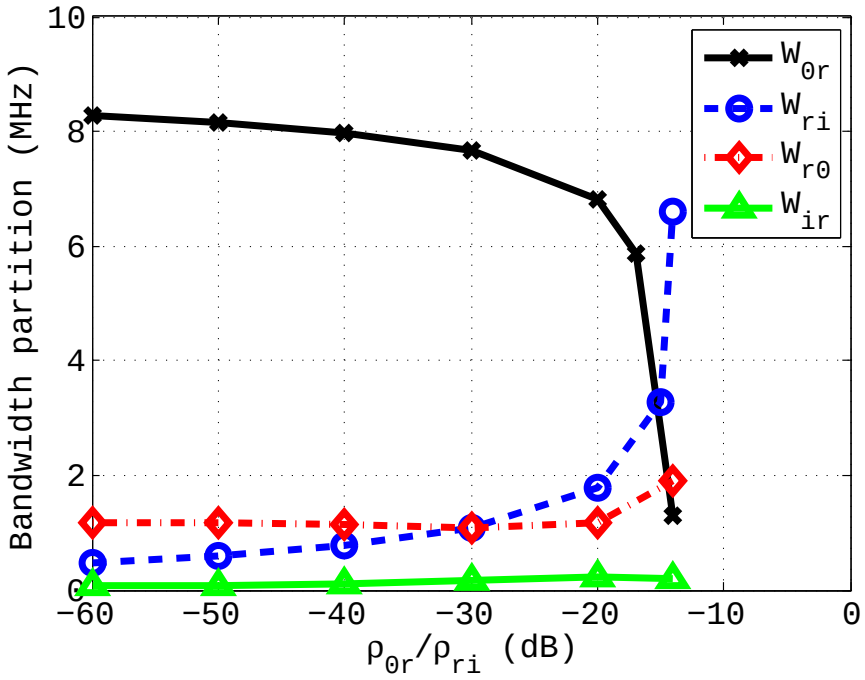

Figure 2.7: Numerical Solution to *Problem (2.18)* showing bandwidth sharing among links. Notice the increased bandwidth allocation W_{r0} and W_{ri} at $\rho_{0r}/\rho_{ri} \geq -18$ dB. This is because the relay power has reached its maximum prescribed limit and minimum required rate is maintained by allocating more bandwidth to the links.

power untill the eNodeB reaches its maximum power and thereafter the system is infeasible. Figs. 2.7 and 2.8 plot the bandwidth partition and power expenditure at each of the links. The plots are for increasing ρ_{0i}/ρ_{ri} (dB) values until the optimization problem becomes infeasible. From Fig. 2.8 we can deduce that the reason for infeasibility is the relay power constraint. While the eNodeB power for downlink P_{0r} and UE i power P_{ir} in the uplink have not reached their maximum but the relay power $(P_{ri} + P_{r0})$ is at its maximum value of 23 dBm. Just before the system becomes infeasible the optimization problem tries to allocate as much bandwidth as possible for the (r, i) and $(r, 0)$ links, so that even with the limited 23 dBm relay power the system can still maintain the required data rate. This can be seen from the increasing W_{ri},W_{r0} and the decreasing W_{0r}, W_{ir} near the infeasibility region in the Fig. 2.7. This can also be verified from Fig. 2.8 where the corresponding to points in Fig 2.7, where the bandwidth allocation is decreasing for W_{0r} and W_{ir} there is a increase in power expenditure for P_{0r} and P_{ir}

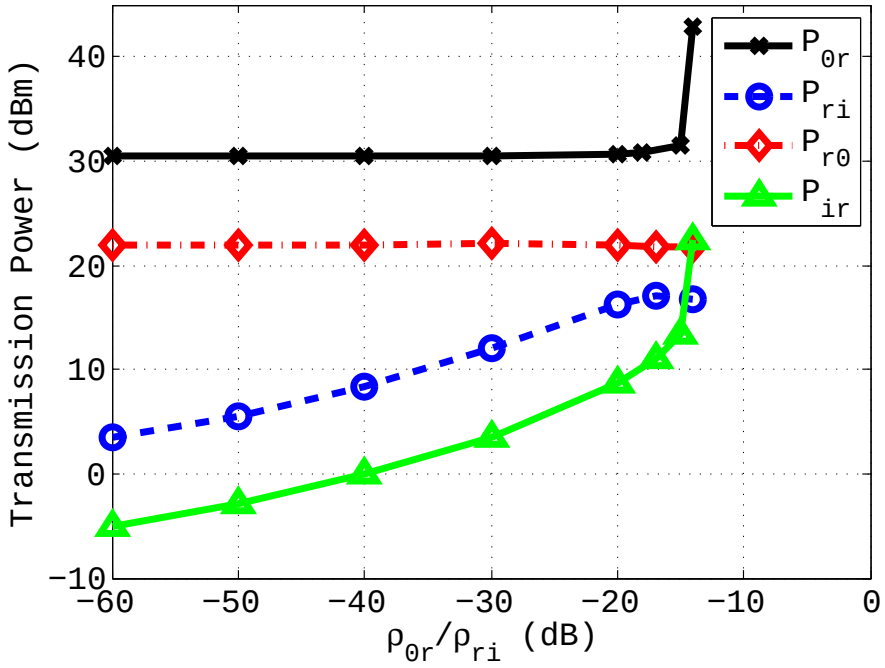

Figure 2.8: Numerical Solution to *Problem (2.18)* showing power consumption on individual nodes. The relay power consumption is the sum $P_{r0} + P_{ri}$. Notice the increased power consumption in eNodeB and UE after $\rho_{0r}/\rho_{ri} \geq -18$ dB i.e. once the relay power reached the maximum prescribed limit. This is because a large portion of the bandwidth is allocated to links $(r, 0)$ and (r, i) and consequently the links $(0, r)$ and $(r, 0)$ are allocated less bandwidth out the total.

to maintain the required data rate. To summarize, the bandwidth sharing utilizes that power that would otherwise be unused in the eNodeB and the UE i while transferring some extra bandwidth to the relatively power constrained relay to maintain required rate. When there is no bandwidth sharing, each link is given a fixed 2.5 MHz bandwidth, and the relay power optimized between uplink and downlink. However, when there is no bandwidth sharing, it is found to be infeasible after when $\rho_{0i}/\rho_{ri} > -16$ dB, while with sharing infeasible region starts when $\rho_{0i}/\rho_{ri} > -14$ dB (Fig. 2.9). This is because there is not enough relay power to maintain the required rate in uplink and downlink. Converting to distances, this corresponds to UE being anywhere in the region of 293 m from relay when there is no bandwidth sharing and 342 m when there is bandwidth sharing. Thus bandwidth sharing results in a 49 m increase in coverage area. This is illustrated in Fig. 2.10. Also, when both the schemes are feasible the

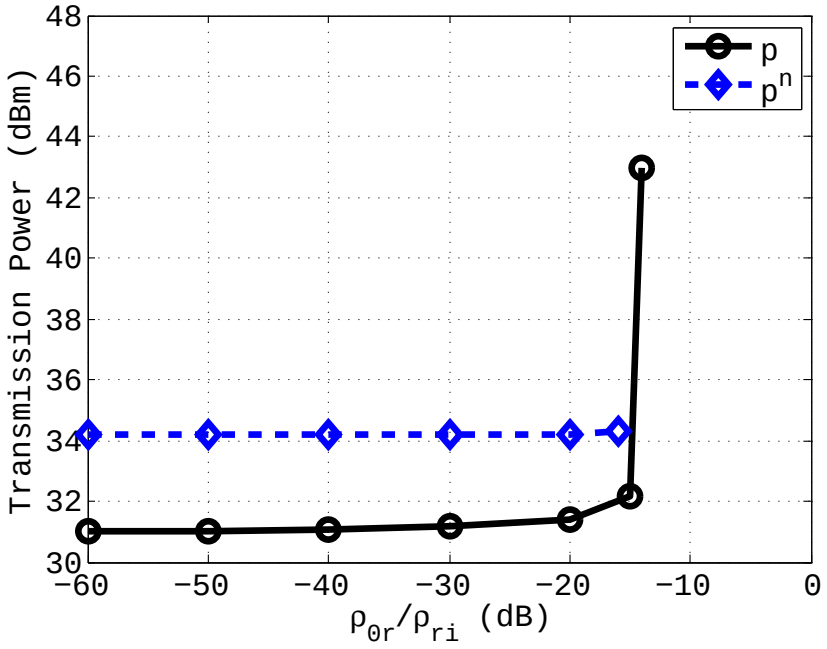

Figure 2.9: Numerical Solution to *Problem (2.18)* showing total Power consumption in the system. P represents the total power consumed when bandwidth is shared and P^n for fixed allocation of bandwidth.

total power consumption with bandwidth sharing is 3 dB less than when there is no bandwidth sharing.

2.5 Relation to Time Shared System

In this section we consider the case of an in-band relaying system with an eNodeB, a single relay and N users. We consider a frame structure similar to that employed in LTE-A and IEEE802.16m schemes where the time is also shared between eNodeB-relay and relay-UE links. Fig. 2.11 shows the typical frame structure of such a system with total frame time T. We denote T_{BS} as the time allocated to eNodeB to relay downlink frame. Similarly T_{RS}^{d} stands for downlink relay to UE frame size, T_{UE} stands for UE to relay uplink frame size, T_{RS}^{u} stands for relay to eNodeB uplink frame size. We also have

$$T_{\text{BS}} + T_{\text{RS}}^{\text{d}} + T_{\text{UE}} + T_{\text{RS}}^{\text{u}} \leq T$$

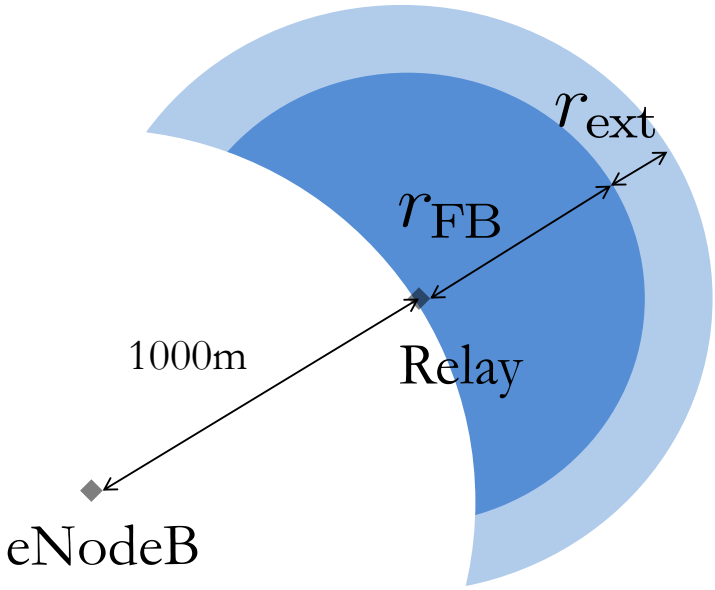

Figure 2.10: The diagram shows coverage extension when bandwidth is shared as compared to fixed allocation of bandwidth. The relay is placed at position of 1000 m from the eNodeB. We assume that the UE does not have direct access to or from the eNodeB and communicates only through the relay. Here r_{FB} denotes the relay coverage till infeasibility for fixed bandwidth allocation scheme (dark blue region) and r_{ext} the coverage extension due to bandwidth sharing (light blue region). For *Problem (2.18)*, $r_{\text{FB}} = 293$ m and $r_{\text{ext}} = 49$ m. Also within the region where both schemes are feasible bandwidth sharing for *Problem (2.18)* consumes up to 3 dB less power.

and that all time slots be positive. During the time slot T^{d}_{RS} relay transmits to all the UEs simultaneously in orthogonal bandwidth allocations and similarly during the time slot T^{u}_{RS} all UEs transmit simultaneously to the relay in uplink in orthogonal bandwidth allocations. All the available bandwidth can be used in the each of the time slots. In terms of the instantaneous power S_{ir} and bandwidth B_{ir} across the link (i, r), we formulate the problem of weighted power minimization as follows,

minimize

$$\alpha_0 \frac{T_{\text{BS}}}{T} S_{0r} + \alpha_r \left(\frac{T^{\text{d}}_{\text{RS}}}{T} \sum_{i=1}^{N} S_{ri} + \frac{T^{\text{u}}_{\text{RS}}}{T} S_{r0} \right) + \frac{T_{\text{UE}}}{T} \sum_{i=1}^{N} \alpha_i S_{ir} \tag{2.23a}$$

subject to

$$T^{\text{d}}_{\text{RS}} \mathcal{R}(\mathsf{SNR}_{ri}, B_{ri}) \geq T R^{d}_{i,\min} \forall i = 1 \ldots N \tag{2.23b}$$

$$T_{\text{BS}} \mathcal{R}(\mathsf{SNR}_{0r}, B_{0r}) \geq T^{\text{d}}_{\text{RS}} \left(\sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{ri}, B_{ri}) \right) \tag{2.23c}$$

$$T_{\mathrm{UE}}\mathcal{R}(\mathsf{SNR}_{ir}, B_{ir}) \geq TR^{u}_{i,\min} \forall i = 1 \ldots N \tag{2.23d}$$

$$T^{\mathrm{u}}_{\mathrm{RS}}\mathcal{R}(\mathsf{SNR}_{r0}, B_{r0}) \geq T_{\mathrm{UE}} \left(\sum_{i=1}^{N} \mathcal{R}(\mathsf{SNR}_{ir}, B_{ir}) \right) \tag{2.23e}$$

$$B_{0r} \leq B \tag{2.23f}$$

$$\sum_{i=1}^{N} B_{ri} \leq B \tag{2.23g}$$

$$\sum_{i=1}^{N} B_{ir} \leq B \tag{2.23h}$$

$$B_{r0} \leq B \tag{2.23i}$$

$$T_{\mathrm{BS}} + T^{\mathrm{d}}_{\mathrm{RS}} + T_{\mathrm{UE}} + T^{\mathrm{u}}_{\mathrm{RS}} \leq T \tag{2.23j}$$

variables

$$T_{\mathrm{BS}}, T^{\mathrm{d}}_{\mathrm{RS}}, T_{\mathrm{UE}}, T^{\mathrm{u}}_{\mathrm{RS}}, S_{0r}, S_{ri}, S_{r0}, S_{ir}, B_{0r}, B_{ri}, B_{r0}, B_{ir}$$

$$\forall i = 1 \ldots N$$

In *Problem (2.23)*, the rate constraints are given by (2.23b), (2.23c), (2.23d), (2.23e) and the bandwidth constraints by (2.23g), (2.23f), (2.23h), (2.23i). The frame time constraint is given by (2.23j). The objective term corresponds to minimizing the weighted average power in the system in a single frame. The power constraint in the system can be either a peak power constraint or average power constraint. The peak power constraints take the form,

$$S_{0r} \leq P_{0,\max}, \tag{3k-P}$$

$$S_{ir} \leq P_{i,\max}, \tag{3l-P}$$

$$\sum_{i=1}^{N} S_{ri} \leq P_{r,\max}, \tag{3m-P}$$

$$S_{r0} \leq P_{r,\max}, \tag{3n-P}$$

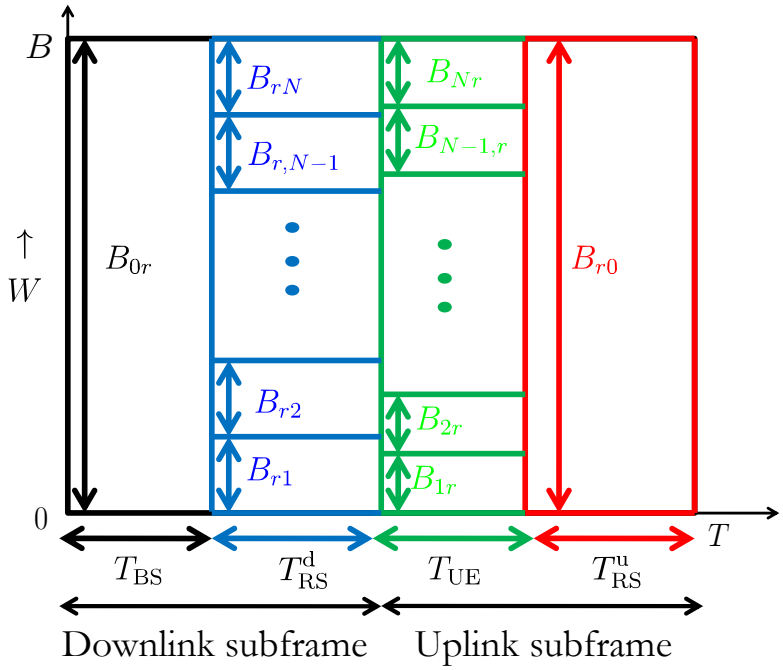

Figure 2.11: The diagram shows the resource block division for a time shared system represented by *Problem (2.23)*. While the total frame time T is divided into uplink and downlink subframe time zones each of them is further divided for enabling single hop relaying. T_{BS} is the time allocated to eNodeB for the downlink frame, $T^{\mathrm{d}}_{\mathrm{RS}}$ stands for downlink relay to UE frame size, T_{UE} stands for UE to relay uplink frame size, $T^{\mathrm{u}}_{\mathrm{RS}}$ stands for relay to eNodeB uplink frame size. In the time slot when the relay transmit to all UEs and also when all UEs transmit to relays the bandwidth is shared orthogonally as shown in the figure.

while the average power constraints are,

$$\frac{T_{\mathrm{BS}}}{T} S_{0r} \leq P_{0,\max} \tag{3k-A}$$

$$\frac{T_{\mathrm{UE}}}{T} S_{ir} \leq P_{i,\max} \tag{3l-A}$$

$$\frac{T^{\mathrm{d}}_{\mathrm{RS}}}{T} \sum_{i=1}^{N} S_{ri} + \frac{T^{\mathrm{u}}_{\mathrm{RS}}}{T} S_{r0} \leq P_{r,\max} \tag{3m-A}$$

While peak power constraints are important from the point of view of avoiding RF power amplifier non-linearities, the average power constraints becomes relevant when there are energy consumption limitations in individual nodes on a per frame basis. The power constraints for *Problem (2.1)* can be interpreted both in terms of peak and average power limits.

2.5.1 Bandwidth Shared System vs. Time Shared System

In this section we show the relation between weighted average power minimization in time shared relaying (*Problem 2.23*) as in 4G systems and weighted power minimization in bandwidth shared system as in *Problem (2.1)*. As mentioned in section 2.5 the time shared system with peak power constraints is represented by *Problem (2.23)* with the constraints (3k-P), (3l-P), (3m-P) and (3n-P). The time shared system with average power constraints is represented by *Problem (2.23)* with the constraints (3k-A), (3l-A) and (3m-A). The behaviour of the two systems as compared to the bandwidth sharing system is explained in the following theorem.

Theorem 2.5.1. *In terms of minimizing the weighted transmit power in the system:*

- *Time sharing with average power constraints is equivalent to bandwidth sharing.*
- *Time sharing with peak power constraints will yield a weighted power objective at least as large as bandwidth sharing.*

Proof. We replace the instantaneous power variables S_{0r}, S_{ri}, S_{r0}, S_{ir} and the bandwidth variables B_{0r}, B_{ri}, B_{r0}, B_{ir} in *Problem (2.23)* by their time averaged power variables P_{0r}, P_{ri}, P_{r0}, P_{ir} and average bandwidth variables W_{0r}, W_{ri}, W_{r0}, W_{ir} respectively as shown in the following the two steps,

1. $TP_{0r} = T_{\mathrm{BS}} S_{0r}$, $TP_{ri} = T^{\mathrm{d}}_{\mathrm{RS}} S_{ri}$, $TP_{r0} = T^{\mathrm{u}}_{\mathrm{RS}} S_{r0}$, $TP_{ir} = T_{\mathrm{UE}} S_{ir}$
2. $TW_{0r} = T_{\mathrm{BS}} B_{0r}$, $TW_{ri} = T^{\mathrm{d}}_{\mathrm{RS}} B_{ri}$, $TW_{r0} = T^{\mathrm{u}}_{\mathrm{RS}} B_{r0}$, $TW_{ir} = T_{\mathrm{UE}} B_{ir}$

The reformulation will result in a convex problem with the resulting objective and the rate constraints the same as that of *Problem (2.1)*. The bandwidth constraints in

Problem (2.23) will be reformulated into,

$$\begin{aligned} W_{0r} &\leq \frac{T_{\mathrm{BS}}}{T} B, \\ \sum_{i=1}^{N} W_{ri} &\leq \frac{T_{\mathrm{RS}}^{\mathrm{d}}}{T} B, \\ \sum_{i=1}^{N} W_{ir} &\leq \frac{T_{\mathrm{UE}}}{T} B, \\ W_{r0} &\leq \frac{T_{\mathrm{RS}}^{\mathrm{u}}}{T} B. \end{aligned} \tag{2.24}$$

However, combining those with the frame time constraint (2.23j), will result in a single constraint given by

$$\sum_{i=1}^{N} (W_{ir} + W_{ri}) + W_{r0} + W_{0r} \leq B$$

which is the same as equation (2.1f).

When the average power constraints given by equations (3k-A), (3l-A) and (3m-A) are employed in the time sharing system, the above reformulation of variables will result in *Problem (2.1)* implying that minimizing the weighted power in the time sharing system with average power constraints will result in the same objective value as in bandwidth sharing system.

When the peak power constraints given by equations (3k-P), (3l-P), (3m-P) and (3n-P) are employed the power constraints will be reformulated into

$$\begin{aligned} P_{0r} &\leq \frac{T_{\mathrm{BS}}}{T} P_{0,\max}, \\ P_{ir} &\leq \frac{T_{\mathrm{UE}}}{T} P_{i,\max}, \\ \sum_{i=1}^{N} P_{ri} &\leq \frac{T_{\mathrm{RS}}^{\mathrm{d}}}{T} P_{r,\max}, \\ P_{r0} &\leq \frac{T_{\mathrm{RS}}^{\mathrm{u}}}{T} P_{r,\max} \end{aligned} \tag{2.25}$$

Clearly these power constraints result in a lesser search space for the reformulated time sharing optimization problem than the power constraints in *Problem (2.1)*. Therefore, the optimal objective of the time sharing system of *Problem (2.23)* with peak power constraints is at least as much as that obtained from the bandwidth sharing system of *Problem (2.1)*. □

When the nodes are transmitting for only a fraction of the frame time in a time sharing system, the effective rate with which they need to communicate is higher than the minimum prescribed rate to transmit the same amount of data in a frame duration. For example, for the rate constraints (2.23b) the effective rate at which the system has to download data in time slot $T_{\mathrm{RS}}^{\mathrm{d}}$ to the users is $TR_{i,\min}^{d}/T_{\mathrm{RS}}^{\mathrm{d}}$ and in order to accomplish that the power has to be increased. In this situation the average power constraints will not affect the system when compared with the bandwidth shared system. However, with peak power constraints, increasing the power levels may result in hitting the power constraints and may perform worse than a bandwidth shared system. Also when we are not considering power constraints time sharing is equivalent to bandwidth sharing and hence the lemma 2.3.1, theorem 2.3.2, and lemma 2.3.3 obtained for bandwidth sharing is also relevant for time sharing system.

2.6 Concluding Remarks

In the first part of the thesis we introduced bandwidth and power optimization in a cellular system employing type-I relaying system. We formulated a weighted power minimization problem, optimizing over both power and bandwidth under rate, bandwidth and power constraints for serving multiple users We developed theoretical insights into the nature of optimal solution when the system has unconstrained maximum power. From the implementation perspective, it is also seen that bandwidth sharing provides total power gain of about 3.5 dB as compared to baseline scheme of fixed allocation of bandwidth per link. This is a significant gain in view of the actual amount of eNodeB power in watts saved and fact that eNodeB power expenditure forms 80% of the total power spent in the cellular system [12]. Future directions of interest include cases with multiple cells and when Inter Cell Interference (ICI) is significant.

Chapter 3

Massive Multi-cell Multi-user MIMO Systems

3.1 Introduction

Cellular systems with large number of base station antennas have been found to be advantageous in mitigating the fading effects of the channel [27] while increasing system capacity. In the downlink, dense low-powered base stations operating with power in the order of milliwatts have the potential to conserve power as compared to current systems. In the uplink, coherent receiver processing with a large number of antennas reduces the transmitted powers of the users. It is shown in [27] that in an infinite antenna regime, and in a bandwidth of 20 MHz, a time division duplexing system has the potential to serve 40 single antenna users with an average throughput of 17 Mbps per user. However, any advantages offered by multiple antennas at the base station can be utilized only by gaining the channel knowledge between the base station and all the users. This requires training data to be sent from the users. In a typical system the time-frequency resources are divided into Physical Resource Blocks (PRBs) of coherence-time coherence-bandwidth product. For each user, it is necessary and sufficient to estimate the channel in every PRB assigned to that user. Thus, some resources (time slots or equivalently frequency channels) are used for channel estimation and the rest are used for transmission in uplink or downlink. However, in [26], it has been shown that the number of orthogonal pilot symbols required for channel estimation is proportional to the total number of users in the system. Because of the orthogonality imposed on training sequences of the users, as the system scales with the number of users, training may take up a significant portion of the PRB. As this is undesirable, only a part of the coherence time is utilized to learn the channel. In this case, the pilot sequences in different cells overlap over time-frequency resources and, as a consequence, the channel

estimates are corrupted. This *pilot interference* is found to be a limiting factor as we increase the number of antennas [29].

It is shown in [27] that in the limit of large number of antennas, the SINR using a matched filter receiver is limited by interference power due to pilot contamination. While the result assumes a regime with finite number of users, we can also envision a regime where the number of users may be comparable to the number of antennas such as a system with 50 antenna base stations serving 50 users simultaneously. In this work, we do a large system analysis of uplink multi-cell, multi-antenna system when the receiver employs an MMSE filter as well as matched filter to decode the received signal. We investigate the SINR in a regime where the number of users per cell is comparable to the number of antennas at the base station. The matched filter and the MMSE filter(which is designed to maximize the SINR) is evaluated in the following cases:

1. when there is a perfect channel estimate,

2. when we have a pilot corrupted channel estimate.

We let the number of antennas and the number of users per base station grow large simultaneously while maintaining a fixed users/antennas ratio α and observe the SINR for the above two cases as a function of α. To do so we make use of the similarity of the uplink received signal in a MIMO system to that of the received signal in a CDMA system [35]. Much of the research in large MIMO systems with Rayleigh channel can be borrowed from the considerable literature for CDMA systems. The channel vector with i.i.d entries for the large MIMO system is analogous to signature sequence in a CDMA system so that antennas contribute to the processing gain. For example, the uplink analysis of an asymptotic regime [35] with both users and signature sequences tending to infinity translates directly to results in a large MIMO system when signatures are replaced by antennas. In both systems it is observed that the asymptotic analysis is a good approximation for practical number of antennas (signatures) and users. While in a CDMA system we assume that the signature sequences are known, there are practical limitations in learning a mobile radio multi-antenna channel (antenna signatures) in a multi-cellular system, as shown in [27]. In this work we explore this limitation when

users simultaneously estimate the channel and the estimates are subject to pilot contamination. We focus our results in the regime $\alpha > 0.1$ as opposed to recent works such as [27, 32, 38, 40] which are found to be approximately in the regime of $0 \leq \alpha \leq 0.1$. Further, we compare the results of the asymptotic SINR expression so obtained with that of the performance of the matched filter.

3.1.1 Related Work

A similar large system analysis in the context of a Network-MIMO architecture was presented in [44]. The authors concluded that high spectral efficiencies can be realized even with 50 antennas in their architecture, paralleling the existing literature results in CDMA systems. In [40], the results obtained suggest to scale the transmission power by the square root of the number of base station antennas, as opposed to scaling by the number of antennas. This assumes that the transmission power during training and data are same. In general we take the approach in [25] where the transmission power can be different for the training and data symbols during a coherence time. In the MU-MIMO literature the asymptotic SINR is also called the "deterministic equivalent" of the SINR. Recent work in [34] finds the deterministic equivalent of SINR with distributed sets of correlated antennas in the uplink. Authors in [38], have done a considerable work in providing the deterministic equivalent for beamforming and regularized zero forcing in the downlink with a generalized channel model taking into account the effect of pilot contaminated channel estimate. They find the number of antennas required to match a fixed percentage of the rate of an infinite antenna regime. Also, the number of extra of antennas required for the matched filter to equal the rate obtained out of the MMSE filter is shown, implicitly showing the interference suppression capability of MMSE filter. We derive in our work the exact amount by which MMSE filter suppresses the interference for a Rayleigh fading channel and provide some fresh engineering insights in the regime with $\alpha > 0.1$. Further, details on our contributions is given in the next section.

There have been significant attempts to mitigate pilot contamination in the recent

works in "Massive MIMO" systems which are, however, specific to the case when antennas far exceed the number of users served i.e. $\alpha < 0.1$. In [32], time shifted pilot schemes were introduced to reduce pilot contamination. There, simultaneous transmission of pilots was avoided by scheduling only a subset of base stations to transmit uplink pilots. Simultaneously, other base stations transmit in the downlink to their users and it is shown that the interference created by these downlink transmission can be cancelled with a large number of antennas at the base station estimating the channel. Pilot contamination is now restricted to base stations in a group that simultaneously transmit uplink pilots. However, this requires that the number of antennas far exceed the number of users. In their recent work, authors in [42,43] show that pilot contamination can be avoided using subspace based channel estimation techniques. They show that the eigenvalues corresponding to the other-cell interference subspace can be separated from the in-cell users in a regime where α is below a threshold. However, their analyses assume an ideal power controlled situation with strict user scheduling and antennas far exceeding the number of users. By contrast, we examine the operating regime in which $0.1 < \alpha < 1$ and determine the effect of pilot contamination on interference and interference suppression capability of MMSE receiver. Also, even with power control, pilot contamination is prevalent when linear MMSE channel estimation is employed.

3.1.2 Contributions of our Work

We develop a large system asymptotic expression for the SINR in a Rayleigh fading environment when using an MMSE filter with a pilot contaminated channel estimate. The SINR expression is dependent on the number of users and the number of antennas only through their ratio. This expression for asymptotic SINR is also strikingly similar to that obtained with matched filter in our earlier work [41]. This SINR explicitly captures the effect of pilot contamination and interference averaging. It is seen that MMSE filter is capable of suppressing the in-cell interference and we derive the exact expression for the interference suppression. As per the authors' knowledge, these contributions have not been reported yet in the literature.

Further, the interference power due to pilot contamination is the same as in a

matched filter with a pilot contaminated estimate. We show that even a system with 50 base station antennas each serving some number of users sufficiently qualifies for the term large system as the users' SINRs are close to the asymptotic limit. Simulation results for achievable rates are close to theory for even a 10-antenna base station with 3 or more users per cell. Operating with a large number of antennas however results in lower transmission power for the symbols and more number of users served simultaneously. The theoretical results are derived assuming that the same set of in-cell orthogonal training signals are repeated across the cells. However, we also show through simulations that in the case of independently generated but non-orthogonal training signals across the cells with orthogonal in-cell training results in similar SINR performance close to the asymptotic limit. In an example seven cell set up, the MMSE filter performs the best in the absence of pilot contamination. We also show an intermediate regime where the MMSE filter with pilot contamination obtains around 7 dB gain over the matched filter with a pilot contaminated estimate. In terms of the five percentile SIR, the MMSE receiver is shown to provide significant gains over matched filtering which is much below the threshold of detection for receivers. Also in most of the operating points α, the performance of the MMSE receiver with pilot estimate is within 5 dB of the MMSE filter with perfect estimate. We also show that the achievable rates are within a 1 bit/symbol of the MMSE filter with perfect estimate when the number of users are comparable to the number of antennas.

3.2 System Model

We consider a system similar to that in [27] with B non-cooperating base stations and K users per base station. We assume that all KB users in the system are allocated the same time-frequency resource. Also, each base station is equipped with M antennas. The channel vector representing the small scale fading between user k in cell j and the antennas in base station l is given by a $M \times 1$ vector $\mathbf{h}_{jk}^{(l)}$. The entries of $\mathbf{h}_{jk}^{(l)}$ are assumed to be independent zero mean i.i.d Gaussian random variables with variance $1/M$ corresponding to the scaling of transmit power by the number of receiver antennas at the base station. This corresponds to an ideal and favourable propagation medium with

rich scattering. A large scale fading coefficient, which represents the power attenuation due to distance and effects of shadowing between base station l and kth user in jth cell is given by $\beta_{jk}^{(l)}$. We assume that $\beta_{jk}^{(l)} < 1$ as we do not expect the received power to be greater than what is transmitted. This is constant across the antennas of the cell l. Accordingly, overall channel vector is given by $\sqrt{\beta_{jk}^{(l)}}\mathbf{h}_{jk}^{(l)}$.

3.2.1 Uplink Transmission

We assume that all users' transmission are perfectly synchronized. Also, while a user's transmission is intended for its own base station, other base stations also hear the transmission. Defining q_{jk} as the symbol transmitted by user k in cell j, $\mathbf{w}^{(l)}$ as the $M \times 1$ noise vector with zero mean Gaussian entries with variance σ^2, the received signal at base station l is given by,

$$\mathbf{y}^{(l)} = \sum_{j=1}^{B}\sum_{k=1}^{K}\sqrt{\beta_{jk}^{(l)}}\mathbf{h}_{jk}^{(l)}q_{jk} + \sigma\mathbf{w}^{(l)}. \tag{3.1}$$

In order to utilize the advantages offered by multiple antennas, the base station has to have an estimate of the channel to all users prior to transmission of uplink information. In a system employing an OFDM physical layer with time-frequency resources, we can divide the resources into physical resource blocks (PRBs) contained in the coherence-time coherence-bandwidth product. Although the channel vector $\mathbf{h}_{jk}^{(l)}$ of each user has to be relearned by the base station at the start of PRB, once learnt for a subcarrier it remains the same for all subcarriers within that PRB. Let the number of coherent symbols be given by T_c and coherent subcarriers be N_c. Therefore, if we fix the number of symbols used for estimation to be T such that $T \leq T_c$, a total of N_cT user's channel can be learnt. This observation was noted in [27]. We would like to point out that it is relevant here as the number of users than can be supported depends on N_c and depending on its value the number of users K that could be supported can be comparable to M. Therefore, it is worthwhile to investigate not only the $M \gg K$ scenario but also the case when M and K large and comparable.

3.2.2 Limitations in gaining Channel Knowledge

During each coherence time, users in a cell spend some pilot symbol times in each PRB for channel estimation at the base station and then data transmission ensues until the end of the block. At base station l, the number of channel vectors $\mathbf{h}_{jk}^{(l)}$ that needs to be learnt is equal to the number of users in the system which is KB where $K = N_c T$. In order to accomplish that, the number of pilots required must at least be KB symbol times in order for the pilot sequences to be orthogonal across the users in the system. However, such a system will not be scalable as there exists some large B for which the product KB will occupy all the coherence time. This is clearly undesirable as pilot training is consuming a significant part of PRB.

In one of the approaches taken in [27], the base station is concerned with only knowing the channel to its own K users and spends only K time-frequency resources for channel estimation instead of KB. Every base station similarly spends its first K time-frequency resources for channel estimation for its K users. The pilot symbols are processed and an MMSE based channel estimate of the channel is formed. MMSE channel estimation is the commonly employed in multiuser MIMO systems [25, 28, 39]. Let $\Psi_{jk} \in \mathbb{C}^{K\times 1}$ denote the training sequence of user k in cell j of duration K symbols. Also assume that the in-cell training sequences are orthogonal i.e $\Psi_{jk}^H \Psi_{jn} = 0$ if $k \neq n$ and 1 otherwise. We assume that the training sequences across the cells are independently generated and hence in general $\Psi_{jk}^H \Psi_{in} \neq 0$ if $i \neq j$ and for all k and n. With $\mathbf{N}$ denoting the additive complex gaussian noise, and ρ_p the pilot transmission power, the received signal across the K training symbols is given by,

$$\mathbf{Y} = \sqrt{\rho_p} \sum_{j=1}^{B} \sum_{k=1}^{K} \sqrt{\beta_{jk}} \mathbf{h}_{jk} \Psi_{jk}^H + \mathbf{N}. \tag{3.2}$$

The MMSE channel estimate for user k in the first cell is then given by,

$$\hat{\mathbf{h}}_{1k} = \mathbf{Y} \left(\frac{\mathbf{I}}{\rho_p} + \sum_{j=1}^{B} \sum_{k=1}^{K} \beta_{jk} \Psi_{jk} \Psi_{jk}^H \right)^{-1} \Psi_{1k} \sqrt{\beta}_{1k}. \tag{3.3}$$

Although it is not common in practice, we assume as in [27–29] that the in-cell pilots are repeated across the cells, in order to get some analytic insight; this implies that

$\Psi_{jk} = \Psi_{ik}$ for all k. Also for convenience we also assume that we have high pilot power. The MMSE channel estimate [47] with pilot contamination when the in-cell orthogonal pilots [28] are repeated across the cells is given by,

$$\hat{\mathbf{h}}_{1k} = \frac{\sqrt{\beta_{1k}}}{\beta^{(k)}} \sum_{j=1}^{B} \sqrt{\beta_{jk}} \mathbf{h}_{jk} \tag{3.4}$$

where, $\beta^{(k)} = \sum_{j=1}^{B} \beta_{jk}$. This estimate is used to design linear detectors to filter the received signal. Later we show that even with actual training given by equation (3.3) the SINRs are very close to when in-cell pilots are reused across the cells. Assuming we have high enough pilot power, we ignore the additive noise affecting the channel estimation in order to focus our results on the pilot contamination problem. In practice, a finite amount of energy is allocated to a coherence time transmission as in [25], then the energy during data transmission is reduced by scaling it with the number of antennas as in section 3.2.1, allowing the rest for pilot transmission power. Thus increasing the number of antennas implies that more power is available for pilot transmissions.

3.2.3 Linear Receivers

Using the channel estimates for all users of the first cell, the MMSE filter for user 1 in the cell 1 is defined as $\arg\min_{\mathbf{c}} \mathbb{E}\left[|q_{11} - \mathbf{c}^H \mathbf{y}|^2 | \hat{\mathbf{h}}_{1k} \forall k\right]$. Defining

$$\mathbf{z} = \sum_{j=2}^{B} \sum_{k=1}^{K} \sqrt{\beta_{jk}} \mathbf{h}_{jk} q_{jk}, \tag{3.5}$$

which represents the other cell interference and $\mathbf{h}_{1k} = \hat{\mathbf{h}}_{1k} + \tilde{\mathbf{h}}_{1k}$ the received signal can be rewritten as,

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\beta_{1k}} \hat{\mathbf{h}}_{1k} q_{1k} + \sum_{k=1}^{K} \sqrt{\beta_{1k}}, \tilde{\mathbf{h}}_{1k} q_{1k} + \mathbf{z} + \sigma \mathbf{w} \tag{3.6}$$

where, $\tilde{\mathbf{h}}_{1k}$ is the result of pilot contamination. The MMSE filter is then given by the expression,

$$\begin{aligned} \hat{\mathbf{c}} &= \left(\mathbf{E}[\mathbf{y}\mathbf{y}^H | \hat{\mathbf{h}}_{1k} \forall k]\right)^{-1} \mathbf{E}[\mathbf{y} q_{11}^* | \hat{\mathbf{h}}_{1k} \forall k] \\ &= \mathbf{S}^{-1} \sqrt{\beta_{11}} \hat{\mathbf{h}}_{11}, \end{aligned} \tag{3.7}$$

where,

$$\mathbf{S} = \sum_{k=2}^{K} \beta_{1k} \hat{\mathbf{h}}_{1k} \hat{\mathbf{h}}_{1k}^H + (\theta_1 + \theta_2 + \sigma^2)\mathbf{I}, \tag{3.8}$$

and,

$$\theta_1 \mathbf{I} = \mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] = \alpha \sum_{j=2}^{B} \left[\frac{1}{K} \sum_{k=1}^{K} \beta_{jk}\right] \mathbf{I}, \tag{3.9}$$

$$\begin{aligned} \theta_2 \mathbf{I} &= \sum_{k=1}^{K} \beta_{1k} \mathbb{E}[\tilde{\mathbf{h}}_{1k} \tilde{\mathbf{h}}_{1k}^H] \\ &= \sum_{j=2}^{B} \left[\frac{1}{M} \sum_{k=1}^{K} \beta_{jk} \left(\frac{\beta_{1k}}{\beta^{(k)}}\right)\right] \mathbf{I}. \end{aligned} \tag{3.10}$$

As seen from the expression for the filter in equation (3.7), the lack of channel knowledge of other-cell users and only a partial channel knowledge of in-cell users shows up as effective noise terms θ_1 and θ_2 respectively. In order to obtain the expression we also use the properties of the MMSE estimate that $\mathbb{E}\left[\hat{\mathbf{h}}\tilde{\mathbf{h}}^H\right] = 0$. In an ideal situation, the channel estimation incurs no error and $\hat{\mathbf{h}}_{1k} = \mathbf{h}_{1k}$ for all k, then

$$\mathbf{c}^* = \left(\sum_{k=1}^{K} \beta_{1k} \mathbf{h}_{1k} \mathbf{h}_{1k}^H + (\theta_1 + \sigma^2)\mathbf{I}\right)^{-1} \sqrt{\beta_{11}} \mathbf{h}_{11} \tag{3.11}$$

This is an optimistic scenario which will serve as a benchmark for the performance of the MMSE filter with pilot contamination. In this section we provide the expressions for the SINR in a large system for MMSE and matched filter. We also provide simulation results to show that performance in realistic scenarios is captured by the asymptotic SINR expression.

3.3 SINR of Linear receivers under Pilot Contamination

We assume that the received signal is projected onto a linear filter $\mathbf{c}_{lk}^{(l)} \in \mathbb{C}^{M \times 1}$ for the k^{th} user in the l^{th} cell. Since the SINR analysis is identical for all users in the system we focus only on user $k = 1$ in base station indexed $l = 1$. We also drop the superscript $(.)^{(1)}$ for notational convenience. After processing the received signal using the linear filter $\mathbf{c}$, let P_{signal}, P_{noise}, P_{pilot}, P_{inter} denote the signal power, noise power,

pilot interference power and interference power respectively. It follows that,

$$P_{\mathsf{signal}} = \beta_{11}\mathbf{c}^H\mathbf{h}_{11}\mathbf{h}_{11}^H\mathbf{c} \tag{3.12}$$

$$P_{\mathsf{noise}} = \sigma^2\mathbf{c}^H\mathbf{c} \tag{3.13}$$

$$P_{\mathsf{pilot}} = \mathbf{c}^H\left(\sum_{j=2}^{B}\beta_{j1}\mathbf{h}_{j1}\mathbf{h}_{j1}^H\right)\mathbf{c} \tag{3.14}$$

$$P_{\mathsf{inter}} = \mathbf{c}^H\left(\sum_{j=1}^{B}\sum_{k=2}^{K}\beta_{jk}\mathbf{h}_{jk}\mathbf{h}_{jk}^H\right)\mathbf{c} \tag{3.15}$$

The received SINR is then given by the expression,

$$\mathsf{SINR} = \frac{P_{\mathsf{signal}}}{P_{\mathsf{noise}} + P_{\mathsf{pilot}} + P_{\mathsf{inter}}} \tag{3.16}$$

Define $\boldsymbol{\beta}_j$ as the random variable representing the large scale fading gain from an arbitrary user in the jth cell. Therefore, β_{jk} can be interpreted as the realization of $\boldsymbol{\beta}_j$ for the user k and let

$$\boldsymbol{\beta} = \sum_{i=1}^{B}\boldsymbol{\beta}_i\,.$$

Next, we state the main theorem of the paper which gives the expression of SINR for a large system when an MMSE filter with a pilot contaminated estimate is used to decode the received signal.

Theorem 3.3.1. *As $M, K \to \infty$, with $K/M = \alpha$, the SINR at the output of filter $\hat{\mathbf{c}}$ given in equation (3.7) converges almost surely to*

$$\widehat{\mathsf{SINR}} = \frac{\frac{\beta_{11}}{1+\left(\sum_{j=2}^{B}\beta_{j1}\right)/\beta_{11}}}{\sigma^2 + \frac{\left(\sum_{j=2}^{B}\beta_{j1}^2\right)/\beta_{11}}{1+\left(\sum_{j=2}^{B}\beta_{j1}\right)/\beta_{11}} + \alpha(\mathbb{E}[\boldsymbol{\beta}] - \mathcal{C}(\alpha))} \tag{3.17}$$

where, $\mathcal{C}(\alpha)$, η_1, η_2 are given by

$$\mathcal{C}(\alpha) = \mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2\eta_1}{1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right] + \frac{\eta_2}{\eta_1}\mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)}{1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right] + \frac{\eta_2}{\eta_1}\mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)}{\left(1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1\right)^2}\right] \tag{3.18}$$

$$\eta_1 = \left(\sigma^2 + \alpha\mathbb{E}[\boldsymbol{\beta}] - \alpha\mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2\eta_1}{1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right]\right)^{-1}, \tag{3.19}$$

$$\eta_2 = \left(\eta_1^{-2} - \alpha\mathbb{E}\left[\left(\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}}{1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right)^2\right]\right)^{-1}. \tag{3.20}$$

Proof. Proof given in Appendix A.2. □

We will see in a large system that Theorem 3.3.1 characterizes the effect of pilot interference power and interference averaging. Specifically, in order to put Theorem 3.3.1 into proper perspective we state two propositions which are the results for SINR of a large system with MMSE filter employing a perfect estimate and a matched filter with pilot contaminated estimate respectively.

Proposition 3.3.2. *As $M, K \to \infty$, with $K/M = \alpha$, the SINR at the output of filter $\mathbf{c}^*$ given in equation (3.11) converges almost surely to*

$$\mathsf{SINR}^* = \beta_{11}\eta_1 = \frac{\boldsymbol{\beta}_{11}}{\sigma^2 + \alpha \sum_{j=2}^{N} \mathbb{E}[\boldsymbol{\beta}_j] + \alpha \mathbb{E}\left[\frac{\boldsymbol{\beta}_1}{1+\boldsymbol{\beta}_1 \eta_1^*}\right]} \tag{3.21}$$

where, $\eta_1^* = \left(\sigma^2 + \alpha \sum_{j=2}^{B} \mathbb{E}[\boldsymbol{\beta}_j] + \alpha \mathbb{E}\left[\frac{\boldsymbol{\beta}_1}{1+\boldsymbol{\beta}_1 \eta_1^*}\right]\right)^{-1}$.

Proof. We state the proposition without proof as it is straightforward to obtain it from the large system analysis techniques used for CDMA systems in [35, 37]. □

SINR^* is the SINR with MMSE filtering with a perfect channel estimate to its own users. This is best case scenario as compared to the MMSE with a channel estimate. We do not expect the SINR of MMSE filter with estimate to exceed this SINR^*.

Proposition 3.3.3. *As $M, K \to \infty$, with $K/M = \alpha$, the SINR at the output of filter $\hat{\mathbf{h}}_{11}$ converges almost surely to*

$$\overline{\mathsf{SINR}} = \frac{\frac{\beta_{11}}{1+\left(\sum_{j=2}^{B} \beta_{j1}\right)/\beta_{11}}}{\sigma^2 + \left[\frac{\left(\sum_{j=2}^{B} \beta_{j1}^2\right)/\beta_{11}}{1+\left(\sum_{j=2}^{B} \beta_{j1}\right)/\beta_{11}} + \alpha \mathbb{E}[\boldsymbol{\beta}]\right]}. \tag{3.22}$$

Proof. Proof given in Appendix A.3. □

It is interesting to see that the expression for $\overline{\mathsf{SINR}}$ converges to a similar expression to the result in matched filtering $\widehat{\mathsf{SINR}}$ where, in both cases, we can separate the effect of pilot interference and interference averaging term. The SINR expression in the limit of infinite antennas but finite number of users per cell are obtained when we put $\alpha = 0$ in equations (3.17) and (3.22) and this corresponds to the expression for

SINR in [27, 28]. It is seen that the SINR expression so obtained is limited by the pilot interference powers at high SNRs. The expression for pilot interference power remains the same irrespective of the system operating load α and $\alpha \neq 0$ implies a non-zero interference averaging power which further affects the SINR. However, as compared $\overline{\mathsf{SINR}}$ in expression (3.22), MMSE filter always obtains interference suppression given by term $\mathcal{C}(\alpha)$ in equation (3.18). The deterministic equivalent for SINR in Rayleigh fading for MMSE filter and the corresponding expression $\mathcal{C}(\alpha)$ representing the interference suppression power are our main contribution of this paper. The terms $\mathbb{E}[\boldsymbol{\beta}]$ and $\mathcal{C}(\alpha)$ can be computed offline for a system with the knowledge of large scale fading distribution. It can also be estimated without the knowledge of the large scale fading distribution from user realizations over time. Depending on the value of $\mathbb{E}[\boldsymbol{\beta}] - \mathcal{C}(\alpha)$ and the operating point α a decision can be made whether to use an MMSE filter or an matched filter. As we will see in the next section, the interference suppression obtained with an MMSE filter is necessary in increasing the outage SINR and achievable rate of the system, when there are considerable number of users as represented by the ratio $\alpha > 0.1$. This is as opposed to the regime in which antennas far outnumber users. In this operating point $\alpha \approx 0$, and then MMSE filter itself may not be necessary as pilot signals are the main contributor to interference. This regime with $\alpha \approx 0$ has been well explored in recent studies in [27,32,38,40]. As mentioned earlier most of our focus for performance analysis is on the regime $\alpha > 0.1$ although the results are perfectly valid for any $\alpha \geq 0$. Across the users in the system the SINR is a random variable by virtue of different received powers of both the signal and the interferers contributing to pilot contamination. Also, the pilot interference power is random by virtue of the choice of the interferers contributing to pilot contamination.

3.4 Performance Analysis

For the numerical evaluation, we consider hexagonal cells with users uniformly distributed in each of the cells, as shown in Fig. 3.1. We consider a scenario where 6 closest cells are interfering with the center cell. We assume $\beta_{1k} = 1$ so that received powers from all the users within a cell are unity. We consider a high SNR of 20 dB

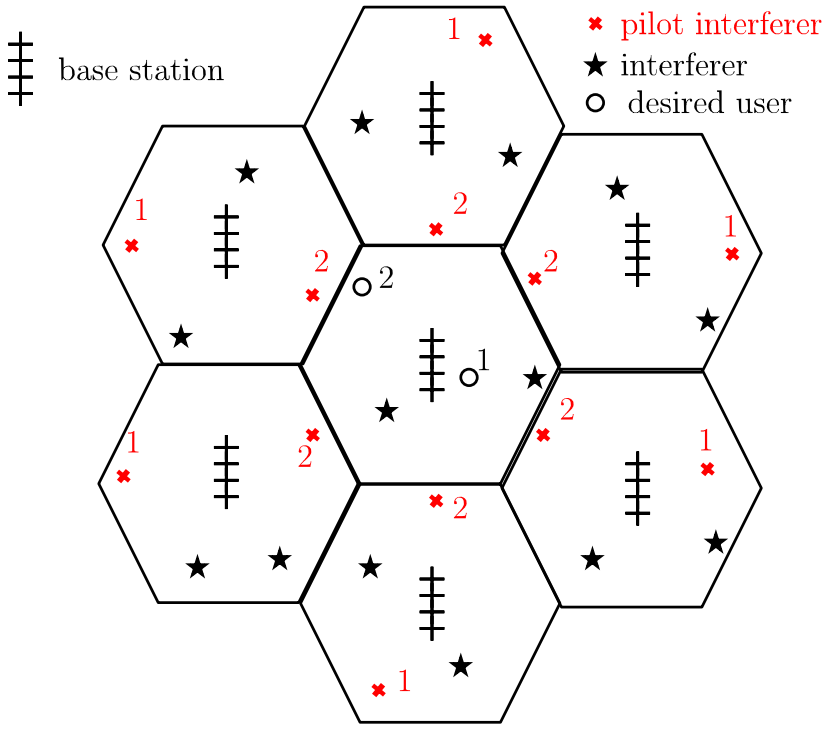

Figure 3.1: In the favourable case, the sum of the received powers of interferers contributing to pilot contamination are very less as compared to that of the desired user. User 1 in the center cell represents such a scenario. The SINR with a pilot corrupted estimate is then comparable to that of perfect estimate. On the other hand for user 2 in the center cell, the pilot interferers received powers are comparable to that of the desired user and represents the worst case scenarios.

and the received powers from all the users in other cells are assumed to take a constant value of $\beta_{jk} = 0.001$, or 0.01, or 0.1 for $j \neq 1$. These represent the contribution of other cell interference for three different idealized scenarios. The interference from other cells is strong as β_{jk} is close to 1. We consider the SINR for the user one in the center cell. Figs. 3.2, 3.3 plot the asymptotic SINR of the MMSE with a pilot corrupted estimate given by $\widehat{\mathsf{SINR}}$ for the case of different received powers. Although in theory the effect for small scale fading vanishes only an with infinite number of antennas it is seen in Fig. 3.3 that even for a 50-antenna base station the actual SINR realizations obtained through simulations are very near to the asymptotic limit. We also plot $\overline{\mathsf{SINR}}$ and SINR^* as baseline for performance comparisons. In Fig. 3.2, it is seen that SINR^* is already affected by other cell interference due $\beta_{jk} = 0.1$ for $j \neq 1$. Hence, the $\widehat{\mathsf{SINR}}$ is

not expected to perform better than that and there is further 4 dB loss due to pilot contamination. However, in the other extreme case when the other cell β_{jk}'s are close to zero, the channel estimate is already better and the $\widehat{\mathsf{SINR}}$ performs close to SINR^*. Useful gains employing an MMSE filter with pilot contaminated channel estimate can be obtained when the other cell β_{jk}'s are neither close to zero or close to unity. In this example when the $\beta_{jk} = 0.01$ for $j \neq 1$, around 7 dB gains are possible in comparison with matched filter with pilot estimate when operating at $\alpha = 0.5$ as seen from fig. 3.3. While there is a loss of 3 dB with respect to the perfect MMSE due nature of channel estimate, the reader is reminded that this is a worse case loss. The curves closes in as we decrease α which represents the $M \gg K$ scenario and also when α increases as in that case the limitation is now the averaged interference term.

In fig. 3.4 we plot the achievable sum rate for users in the first cell. We assume large enough coherence time so that the training time need not be taken into account. This is because our focus is on the sum rate achievable with variation in α. However, if necessary the sum rate can be easily adjusted based on training overhead when coherence time is a significant factor. We fix the number of antennas and calculate the sum rate with varying α as $\alpha M \log_2(1 + \widehat{\mathsf{SINR}})$. The three curves corresponds to the received powers of all users from other cells being either $\beta_{jk} = 0.001$, or 0.01, or 0.1 for j $\neq$ 1 assuming unit received power from the in-cell users. Sum rates of over 20 bits/symbol are achieved for users when the other cell received powers are below 10 dB of the in-cell received powers. Also, the simulation with 50-antenna base station is seen to match the theoretical rates predicted for this set up. The interference limited system has the flexibility to serve a large number of users at low SINR or a few number of users at a high SINR depending on the operating point α. The plot suggest a optimal operating point α for which the sum rate is maximum. For example when $\beta_j = 0.01$ and $\alpha = 0.8$ gives a sum rate of around 88 bits/symbol. Larger α causes the $\widehat{\mathsf{SINR}}$ to be lower so that the α term outside the $\log_2$ is ineffective to increase sum rate while a lower α implies that less users are served and hence lesser sum rate. The peak and the dip thereafter for the sum rate plot is due to the interference suppression action of the MMSE filter. Although not shown, the peak and the dip will not there for a sum

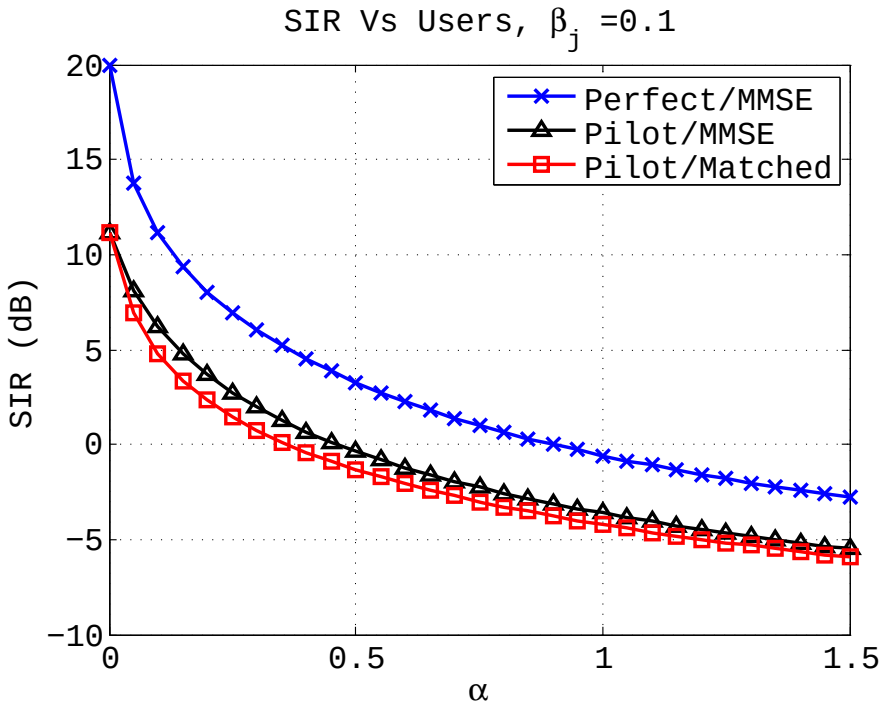

Figure 3.2: The plot shows the asymptotic SINR of the first user in the first cell when MMSE filter with pilot contaminated estimate is used to decode the received signal along with the baseline comparison criterion of MMSE with a perfect estimate and matched filter with a pilot contaminated estimate for an idealized seven cell set up. It is seen that when the other cell received power is just 10 dB of that in cell users the MMSE filter with pilot contaminated estimate performs close to its corresponding matched filter. This is because the limitation is now the other cell interference which MMSE filter is not designed to suppress. Hence, we do not expect the MMSE filter with pilot estimate to be useful when $\beta_j > 0.1$

rate plot of the matched filter receiver. When the other-cell received powers are large, the curve flattens and the sum rate is constant for most of the operating points α. In fig. 3.5, we plot the difference of achievable rate per user between MMSE filter with a perfect estimate and MMSE filter with a pilot estimate for different values of other cell interference power. We do not take into account the training overhead for comparison assuming we obtain the perfect estimate with the same training time. We limit ourselves to $\beta_j < 0.1$ since $\widehat{\mathsf{SINR}}$ is already close to $\overline{\mathsf{SINR}}$ otherwise. Also in $\beta_j > 0.1$ regime, there is significant other cell interference which both the perfect estimate based and the pilot based MMSE filter are not designed to suppress, thereby affecting the achievable

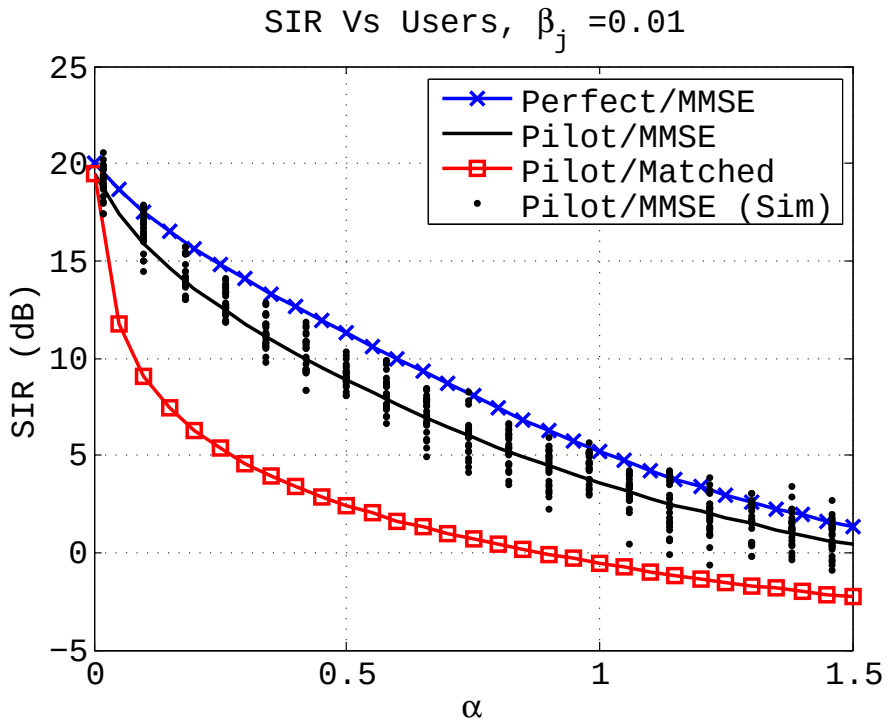

Figure 3.3: The plot shows the asymptotic SIR of the first user in the first cell when MMSE filter with pilot contaminated estimate is used to decode along with the baseline comparison criterion of MMSE with a perfect estimate and matched filter with a pilot contaminated estimate for an idealized seven cell set up. In this case when the other cell received powers is 20 dB lower than that of in-cell received powers significant gains are obtained wit respect using a matched filter with pilot contaminated estimate

rates. We plot five different curves corresponding to system operating points α. As expected when the other cell interference is small, for all values of α the difference is negligible, however for most of β_j's, the rate difference is constant and is higher when α is lower. Larger α leads to the effect of interference being pronounced and hence the rate difference increases at a faster rate for the same change in β_j. Although when $\alpha = 1$ there is only a 0.4 bits/symbol difference the sum rate will be affected differently. For example in a 50-antenna base station with 50 users at $\beta_j = 0.1$ this could mean that sum rate with pilot contaminated MMSE filter is 20 bits/symbol lower than that perfect MMSE filter. On the other hand when $\alpha = 0.2$, the sum rate difference is 12 bits/symbol.

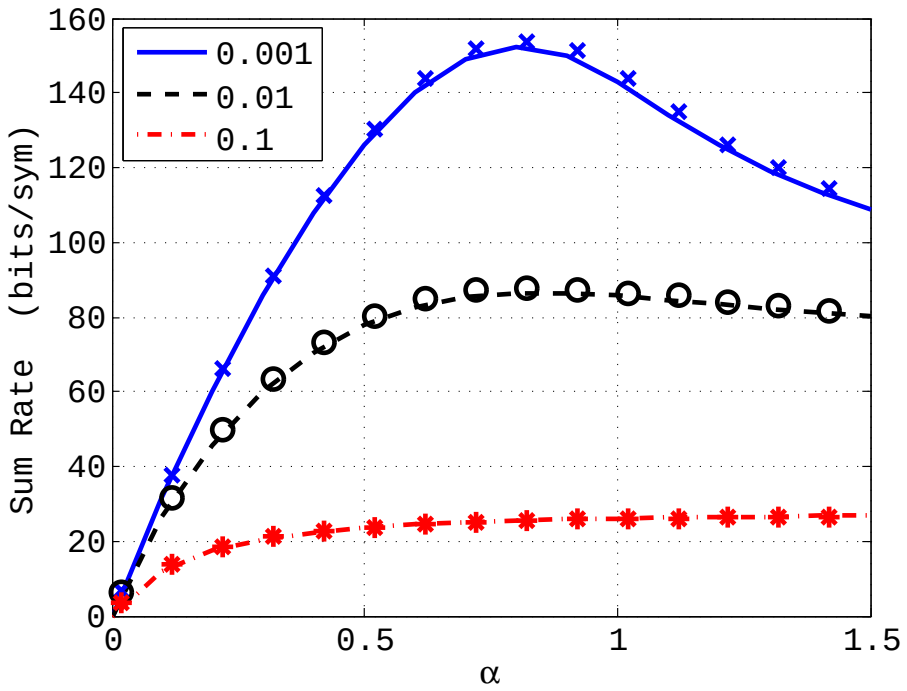

Figure 3.4: The plot shows the sum rate of the users in first cell for different operating points of α. The three curves correspond to three different received powers from other cell users. The markers correspond to simulation with 50-antenna base stations and match the theoretical predictions.

3.4.1 Effect of Pilot Contamination

Through a couple of typical possible realizations of user positions, we explain the effect of pilot interference in $\widehat{\mathsf{SINR}}$, $\overline{\mathsf{SINR}}$ and compare it with that of the SINR with a perfect estimate. For illustration, in Fig. 3.1 consider only distance based pathloss in large scale fading although the result holds when shadowing is also present. This is applicable to both matched filter and MMSE filter. Consider the first scenario when

$$\frac{\sum_{j=2}^{B}\beta_{j1}}{\beta_{11}} \ll 1 \Rightarrow \frac{\sum_{j=2}^{B}\beta_{j1}^2}{\beta_{11}} \ll 1 \tag{3.23}$$

This corresponds to the fact that sum of received powers of the interferers are much less that that of desired user power. Under these conditions the SINR of the received signal in equation (3.22) is,

$$\mathsf{SINR} \approx \frac{\beta_{11}}{\sigma^2 + \alpha(\mathbb{E}[\boldsymbol{\beta}] - \mathcal{I})}$$

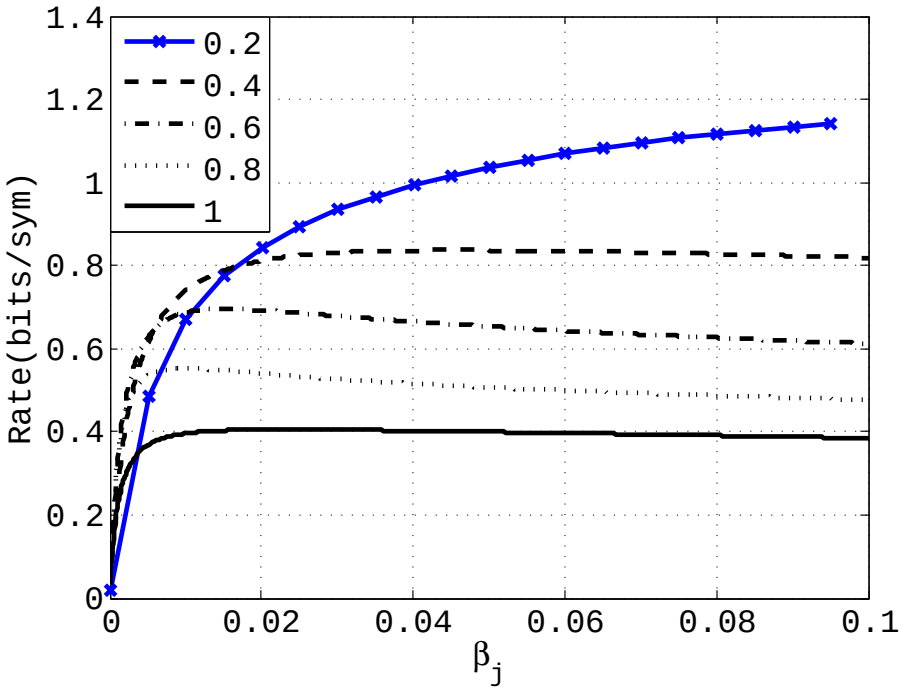

Figure 3.5: The plot shows the loss of rate due to pilot contamination when an MMSE filter with pilot contaminated estimate is used to decode the received signal as compared to perfect estimate MMSE filter. The different curves correspond to different values of α from 0.2 to 1. It is seen that smaller the α the MMSE filter with pilot contaminated estimate performs worse than the ideal MMSE filter. However, the sum rate for users per base station is different.

where, $\mathcal{I} = 0$ if matched filter is employed or $\mathcal{I} = \mathcal{C}(\alpha)$ if MMSE filter $\hat{\mathbf{c}}$ is employed. As we will see in the next section, typically scenarios show that the interference suppression power $\mathcal{C}(\alpha)$ is almost same as what could have been with a filter $\mathbf{c}^*$. Hence the SINR of the filter with the corrupt channel estimate is as good as the SINR with a perfect channel estimate. In Fig. 3.1, the situation of user 1 in the center cell represents the favourable scenario with the interferers contributing to the pilot contaminated channel estimate are far such that the condition (3.23) is satisfied. On the other hand if gains of all the interferers are comparable to that of the desired users, i.e.,

$$\frac{\sum_{j=2}^{B} \beta_{j1}}{\beta_{11}} \approx B - 1 \tag{3.24}$$

then pilot interference contributes negatively to the SINR in addition to interference averaging. This is represented by realization of user 2 of the center cell in Fig. 3.1.

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Perfect Est. R (bits/symbol)	6.0	5.0	4.3	3.8	3.4	3.1	2.8	2.64	2.4	2.2
Pilot Est. R (bits/symbol)	4.7	4.0	3.4	3.0	2.7	2.5	2.3	2.1	1.9	1.9

Table 3.1: The table shows the achievable rate with MMSE filtering for a 50-antenna base station serving different number of users. Both theory and simulations correspond to the similar results.

α	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Perfect Est. R (bits/symbol)	4.4	4.0	3.6	3.3	3.0	2.7	2.4	2.3
Pilot Est. R (bits/symbol)	3.6	3.2	2.9	2.7	2.4	2.2	2.0	2.0

Table 3.2: The table shows the achievable rate with MMSE filtering for a 10-antenna base station serving different number of users. Simulated results closely match with theoretical predictions. This shows that even for a 10 antenna base station the large system analysis gives accurate results for the achievable rate. For $\alpha < 0.3$, there are only 3 or less users which may not be sufficient for the expectation over the rate to converge.

Therefore, we can conclude that, as compared to the linear filter with perfect estimate, the filter with a pilot estimate has higher probability that it is less than a given SINR.

3.4.2 Five Percentile SINR

In the earlier section we showed that pilot contamination has the effect of reducing the outage SINR. In order to get more intuition under practical scenarios of large scale fading gains, we consider the seven cell model with cell radius is $R = 1$ km, and assume a COST231 model for propagation loss between the base station and the users. The noise power is assumed to be -174 dBm and user transmit power of 23 dBm. We plot the five percentile of the SINR in Fig. 3.6 for the perfect MMSE filtering given by SINR^* and MMSE filter with pilot contamination given by $\widehat{\mathsf{SINR}}$, for varying values of α. To that extend, we compute the interference terms $\mathbb{E}[\boldsymbol{\beta}]$ and $\mathcal{C}(\alpha)$ offline by averaging over a sufficient number of user positions. Also, η_1 and η_2 can be computed offline for different values of α as they are constant for the system at a given α and dependent on the distribution of large scale fading characteristics. Notice that $\widehat{\mathsf{SINR}}$ is devoid of the small scale fading parameters. Also, for SINR^* we compute the terms, η_1^*

and the average interference $\mathbb{E}[\boldsymbol{\beta}] - \mathbb{E}\left[\frac{\boldsymbol{\beta}_1^2 \eta_1^*}{1+\boldsymbol{\beta}_1 \eta_1^*}\right]$. It is found that $\mathbb{E}[\boldsymbol{\beta}] - \mathcal{C} = 38$ dB and $\mathbb{E}[\boldsymbol{\beta}] - \mathbb{E}\left[\frac{\boldsymbol{\beta}_1^2 \eta_1^*}{1+\boldsymbol{\beta}_1 \eta_1^*}\right] = 36$ dB. This shows that in terms of the interference suppression the performance of both filters $\mathbf{c}^*$ and $\hat{\mathbf{c}}$ are almost same. In order to compare the theoretical expression we also plot the five percentile SINR which is generated using simulations. These involves computing the SINRs for various small scale fading channel realizations along with large scale fading. For the simulation we use 50 antenna base stations each serving a different number of users corresponding to different α. The channel estimate is based on same in-cell orthogonal training sequences being repeated across the cells. Further, we also compare the five percentile SINR obtained out of actual channel estimation. In order for that we assume different independently generated in-cell orthogonal training sequences which are non-orthogonal across the cells. We perform an MMSE estimation of the channel and generate MMSE filter using the actual channel estimate given by equation (3.3). It is seen through Fig. 3.6, that $\widehat{\mathsf{SINR}}$ in equation (3.17) matches the five percentile SINR obtained through simulation and is typically less by about 0.3 dB of the theoretical expression. This is true for both in-cell pilot sequences being repeated across the cells as well as different and independent pilot training sequences across the cells. This also implies that the even a 50 antenna base station is large enough for the theoretical predictions to be effective in addition to being independent of the effect of small scale fading in the resulting SINR. As we increase the number of antennas the theoretical expression exactly matches the SINR obtained through simulation. Also, the MMSE filter with pilot contamination performs just 5 dB below the MMSE filter with perfect channel estimate and this gap is unambiguously a result of the pilot contaminated channel estimate. Also it is seen that even at $\alpha = 1$, which implies a heavily loaded system the five percentile $\widehat{\mathsf{SINR}}$ is -9 dB which is well within the sensitivity of base station receivers.

Table 3.1 shows the achievable rates per symbol for a user in the central cell using MMSE based detection with perfect estimate and pilot contaminated estimate. The achievable rate is given by $R = \mathbb{E}[\log(1+\widehat{\mathsf{SINR}})]$ which here is calculated by averaging the instantaneous rate over 2×10^3 realizations of user positions for user 1. This is the same for all users in the central cell. Both theory and simulations based on 50 base stations

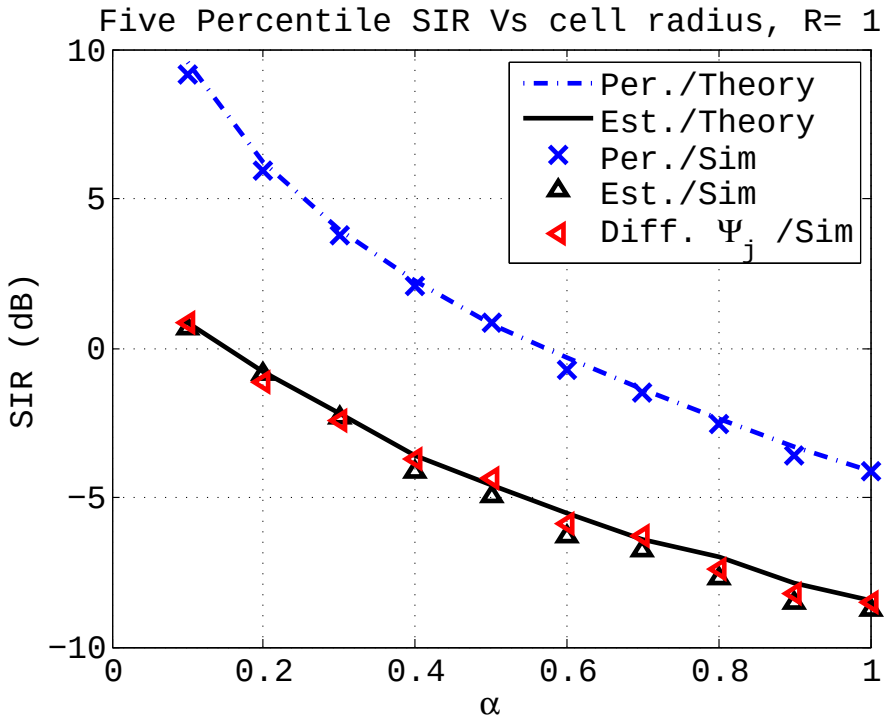

Figure 3.6: The plot shows the five percentile SIR of the first user in the first cell for a seven cell set up in a non-idealized scenario. The received powers from users can be different depending on their positions and shadowing and hence the received SIR is random. The theoretical curves are matched with simulation. The details are described in section 3.4.2.

antennas serving different number of users agree to the numbers shown in the table. It is seen that the difference between them is approximately 1 bit/symbol for small $\alpha > 0.1$ and closes in when it increases. However, for $\alpha \ll 1$, the difference between the rates increases as effect of pilot interference will never let the SINR approach SNR.

Further, Table 3.2, shows the simulated results for the achievable rate for a 10-antenna base station with 3 to 10 users. It is seen that even for a 10 antenna base station, the simulated results agree closely to the earlier results obtained from theory in Table 3.1. This highlights the usefulness of the large system analysis in providing accurate predictions for achievable rates for not necessarily large number of antennas but also contemporary MIMO systems. However, we would like to point out that with more number of antennas we can serve more users at the same rate given below.

3.5 Conclusion

In this work we found the expression for SINR for a large system when a MMSE filter with a pilot contaminated estimate is employed to decode the received signal. We validated the expression through simulations and showed that a 50-antenna base station serving different number of users is sufficient enough to employ our large system results. We characterized the effect of pilot contamination in that it has the effect of reducing the five percentile SINRs as compared to the MMSE with perfect estimate for all values of α. We also found an explicit expression for the interference suppression power due to MMSE filter and compared it with that of matched filter. We showed that five percentile SIR of the MMSE with a pilot contaminated estimate is within 5 dB of MMSE filter with a perfect estimate. We also found that the results with actual channel estimation match the theoretical results. It would also be interesting to see if the considerable work done by the authors in [38], in getting a generalized expression for the deterministic equivalent of the SINR can be further simplified into intuitive expressions for other channel models. This will be of help in realizing engineering conclusions tailored for different channels models like distributed antennas, correlated antennas, distributed sets of correlated antennas [34] to name a few. A brief description of future possible research directions in multi-cell multi-user massive MIMO systems are given in the next section.

3.6 Future Work

3.6.1 Training Vs Antennas: Subspace based Channel Estimation

In this work we have assumed that the training time allocated to users is K symbols. In future work, we wish to study the training overhead for different values of training time and users depending on the coherence time. We believe this can be done through a large system analysis of training time versus the number of antennas per cell. Further, it would be interesting to investigate the relationship between the training time T versus increasing the number of antennas M when both linear MMSE based channel estimation as well as subspace based non-linear channel estimation techniques are employed. In

their recent work, authors in [42, 43] show that *pilot contamination* can be avoided using subspace based channel estimation techniques. If $\mathbf{u}$ is an arbitrary vector then $\mathbb{E}\left[||\mathbf{u}^H\mathbf{y}||^2\right]$ represents the mean energy of the received signal in the direction of $\mathbf{u}$. The correlation matrix

$$\mathbf{R} = \mathbb{E}[\mathbf{y}\mathbf{y}^H]$$

of the received signals is encountered in the mean energy criterion given by $\mathbb{E}\left[||\mathbf{u}^H\mathbf{y}||^2\right]$. In practice the received signal obtained over T symbol periods is used to get an estimate of the correlation matrix given by

$$\hat{\mathbf{R}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}[t]\mathbf{y}^H[t].$$

Using the estimates, they show that the eigenvalues corresponding to the other-cell interference subspace can be separated from the in-cell users when α is below a threshold. *Taking it a step further, I would like to investigate the influence of T/M in systems with subspace based channel estimation.* Techniques from Random Matrix Theory(RMT) in [48] are the necessary tools to approach the problem.

3.6.2 Algorithms from MUD to combat Pilot Interference

Also, recent work has proposed that pilot contamination as an artefact of linear channel estimation techniques [42]. While they have provided a theoretical understanding in an ideal situation, practical solutions applicable to a regime with large number of users are still to be found. Algorithms from multi-user detection for CDMA systems are a useful tool when we have enough coherence time [45]. We are currently looking into such adaptive algorithms that could be implemented for short coherence time scenarios. To elaborate on the usefulness of such algorithms, consider adaptive MMSE filtering [45] being employed for each user to decode its uplink data. Adaptive MMSE does not require the explicit knowledge of the channel. Instead, pilot training sequences known at the base station are sent by every user simultaneously so that each user converges to its desired MMSE filter directly. The explanation for the implementation of adaptive MMSE is as follows. Given the received signal in (3.1), we wish to design an MMSE filter

$\mathbf{c}_{11}$ for the first user in the first cell so that $\mathbb{E}[\left|q_{11} - \mathbf{c}_{11}^H \mathbf{y}\right|^2]$ is minimized. From [45], the filter which minimizes the mean square objective is given by

$$\mathbf{c}_{11} = (\mathbb{E}[\mathbf{y}\mathbf{y}^H])^{-1}\mathbb{E}[q_{11}^* \mathbf{y}].$$

In order to form the filter $\mathbf{c}_{11}$ directly, the channel knowledge $\mathbf{h}_{jk}$ from the first base station to all the users is required.

A way to overcome this difficulty is by using the stochastic gradient descent algorithm, so that the filter adaptively converges to MMSE as proposed for CDMA systems [45]. With $0 < t \leq T$ being the index of time-frequency resources spent for training data, $\{\psi_{11}[t]\}$ the T length training sequence for user indexed one and $\mu > 0$, the progressively decreasing step size, a step in the algorithm is given by the equation,

$$\mathbf{c}_{11}[t] = \mathbf{c}_{11}[t-1] - \mu(\psi_{11}[t] - \mathbf{c}_{11}^H[t-1]\mathbf{y}[t])\mathbf{y}[t]. \tag{3.25}$$

It has been shown that there exist positive values of μ so that the iteration (3.25) converges to the MMSE filter. Also, as long as the training data satisfies $\mathbb{E}[\psi_{jk}\psi_{li}^*] = 0$ for $\{j,k\} \neq \{l,i\}$ the filter converges to the MMSE of user 1 in cell 1. *However, the real challenge appears when implementing the algorithms in short coherence time scenarios of* 100 *to* 200 *symbols.* More research effort is required in this direction.

Appendix A

Proofs of asymptotic SINR in Multi-cell Multi-user MIMO

A.1 Results from Literature

In this section we briefly describe the necessary results from literature to derive the asymptotic SINR expressions. These results were used previously used in the context of CDMA systems in [35–37].

Lemma A.1.1. *[37, Lemma 1] If* $\mathbf{S}$ *is a deterministic* $M \times M$ *matrix with uniformly bounded spectral radius for all* M*. Let* $\mathbf{q} = \frac{1}{\sqrt{M}} \begin{bmatrix} q_1 & q_2 & \dots & q_M \end{bmatrix}^T$ *where* q_i*'s are i.i.d complex random variables with zero mean, unit variance and finite eight moment. Let* $\mathbf{r}$ *be a similar vector independent of* $\mathbf{q}$*. Then,*

$$\mathbf{q}^H \mathbf{S} \mathbf{q} \rightarrow \frac{1}{M} \text{trace}\{\mathbf{S}\}, \tag{A.1}$$

$$\mathbf{q}^H \mathbf{S} \mathbf{r} \rightarrow 0 \tag{A.2}$$

almost surely as $M \rightarrow \infty$*.*

Results in linear MMSE filters for large dimensions have been obtained using Stieltjes transform result on symmetric matrices in [35–37]. For completeness and clarity of understanding of MMSE for multi-user MIMO with pilot contamination we first define the Stieltjes transform of a random variable and state the result without proof here.

Definition A.1.2. *Let a real valued random variable be given by the distribution* G*. Then, the Steiljes transform* $m(z)$ *with complex argument* z *and positive imaginary part is defined as*

$$m(z) = \int \frac{1}{\lambda - z} dG(\lambda). \tag{A.3}$$

Theorem A.1.3. *Let* $\mathbf{X} \in \mathbb{C}^{M\times K}$ *be a matrix with independent and identically distributed complex entries each with variance* $1/M$*. Also, let* $\mathbf{T} \in \mathbb{C}^{K\times K}$ *be a random hermitian non-negative definite matrix independent of* $\mathbf{X}$ *such that the empirical distribution of its eigenvalues converges to a fixed distribution* F *as* $M \to \infty$*. Then almost surely the empirical distribution of eigenvalues of* $\mathbf{XTX}^H$ *converges to a non-random distribution function* G *whose Stieltjes transform* $m(z)$ *satisfies,*

$$m(z) = \frac{1}{-z + \alpha \int \frac{p}{1+pm(z)} dF(p)}, \tag{A.4}$$

for $z \in \mathbb{C}^+$

Next we state a corollary from [36] which is also a consequence of Steiltjes transform result in Theorem A.1.3.

Corollary A.1.4.

$$\alpha \mathbb{E}\left[\frac{p\frac{d}{dz}m(z)}{(1+pm(z))^2}\right] = \int \frac{\lambda}{(\lambda - z)^2} dG(\lambda) \tag{A.5}$$

Proof.

$$\begin{aligned}
\alpha \mathbb{E}\left[\frac{p\frac{d}{dz}m(z)}{(1+pm(z))^2}\right] &\overset{(a)}{=} \alpha \frac{d}{dz}\mathbb{E}\left[1 - \frac{pm(z)}{1+pm(z)}\right], \\
&= -\frac{d}{dz}\alpha\mathbb{E}\left[\frac{pm(z)}{1+pm(z)}\right], \\
&\overset{(b)}{=} -\frac{d}{dz}(1 + zm(z)), \\
&= -\frac{d}{dz}\int \frac{\lambda}{\lambda - z} dG(\lambda),
\end{aligned} \tag{A.6}$$

where (a) is due to dominated convergence theorem and (b) is due to equation (A.4). □

A.2 Proof of Theorem 3.3.1

Let the overall channel matrix representing the system be defined as,

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_K \end{bmatrix},$$

where,

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{h}_{1i} & \mathbf{h}_{2i} & \dots & \mathbf{h}_{Bi} \end{bmatrix}.$$

Also define the large scale fading coefficient vector to be

$$\mathbf{a}_i = \left[\sqrt{\beta_{1i}} \quad \sqrt{\beta_{2i}} \quad \dots \quad \sqrt{\beta_{Bi}}\right]^T,$$

and $\mathbf{e}_i \in \mathbb{C}^{B\times 1}$ as a unit vector with 1 in the i^{th} position for $i \in \{1, 2, \dots, B\}$. When the filter in equation (3.7) is used to decode the received vector at the base station, the signal power, noise power, and pilot interference power, are respectively given by,

$$P_{\mathsf{signal}} = \beta_{11}(\mathbf{a}_1^H \mathbf{H}_1^H \mathbf{S}^{-1} \mathbf{H}_1 \mathbf{e}_1)^2 \tag{A.7}$$

$$P_{\mathsf{noise}} = \sigma^2 \mathbf{a}_1^H \mathbf{H}_1^H (\mathbf{S}^{-1})^2 \mathbf{H}_1 \mathbf{a}_1, \tag{A.8}$$

$$P_{\mathsf{pilot}} = \sum_{j=2}^{B} \beta_{j1} \left|\mathbf{a}_1^H \mathbf{H}_1^H \mathbf{S}^{-1} \mathbf{H}_1 \mathbf{e}_j\right|^2. \tag{A.9}$$

For further analysis let us define,

$$\mathbf{Z} = \mathbf{S}^{-1} \left(\sum_{j=2}^{B} \sum_{k=1}^{K} \beta_{jk} \mathbf{h}_{jk} \mathbf{h}_{jk}^H + \sum_{k=2}^{K} \beta_{1k} \mathbf{h}_{1k} \mathbf{h}_{1k}^H \right) \mathbf{S}^{-1} \tag{A.10}$$

then the interference power is given by

$$P_{\mathsf{inter}} = \mathbf{a}_1^H \mathbf{H}_1^H \mathbf{Z} \mathbf{H}_1 \mathbf{a}_1 \tag{A.11}$$

Also, if

$$\nu_k = \left(\frac{\beta_{1k}}{\beta^{(k)}}\right)^2,$$

then define $\mathbf{D}_1 \in \mathbb{R}^{(K-1)B\times(K-1)B}$ and $\mathbf{S}_1 \in \mathbb{C}^{M\times(K-1)B}$ as

$$\mathbf{D}_1 = \operatorname{diag}\left\{\left[\nu_2 \mathbf{a}_2 \mathbf{a}_2^H, \quad \nu_3 \mathbf{a}_3 \mathbf{a}_3^H, \quad \dots \quad \nu_K \mathbf{a}_K \mathbf{a}_K^H\right]\right\}, \text{ and} \tag{A.12}$$

$$\mathbf{S}_1 = \left[\mathbf{H}_2 \quad \dots \quad \mathbf{H}_K\right]. \tag{A.13}$$

Then the matrix $\mathbf{S}$ can be rewritten as,

$$\mathbf{S} = \sum_{k=2}^{K} \nu_k \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H + (\theta_1 + \theta_2 + \sigma^2)\mathbf{I} = \mathbf{S}_1 \mathbf{D}_1 \mathbf{S}_1^H + (\theta_1 + \theta_2 + \sigma^2)\mathbf{I}. \tag{A.14}$$

For the matrix $\mathbf{D}_1$ there are $(K-1)(B-1)$ eigenvalues which are equal to zero and $K-1$ non-zero values given by $\frac{\beta_{1k}^2}{\beta^{(k)}}$, for all $k \in \{2, \dots, K\}$. Also, define $\boldsymbol{\beta}_j$ as the random variable representing the large scale fading gain from an arbitrary user in the

j^{th} cell. Therefore, β_{jk} can be interpreted as the realization of $\boldsymbol{\beta}_j$ for the k^{th} user. Therefore, Theorem A.1.3 takes the form

$$m(z) = \left(-z + \alpha \mathbb{E}\left[\frac{\boldsymbol{\beta}_1^2 / \boldsymbol{\beta}}{1 + \boldsymbol{\beta}_1^2 \, m(z) / \boldsymbol{\beta}} \right] \right)^{-1} \tag{A.15}$$

where, the expectation is now over the joint distribution of the $\boldsymbol{\beta}_j$s and $\boldsymbol{\beta}$.

Also, notice that the spectral radius of $\mathbf{S}$ is bounded by $(\theta_1 + \theta_2 + \sigma^2)^{-1}$. Therefore, with

$$\begin{aligned} \boldsymbol{\beta} &= \sum_{j=1}^{B} \boldsymbol{\beta}_j, \\ \bar{\theta_1} &= \alpha \sum_{j=2}^{B} \mathbb{E}\left[\boldsymbol{\beta}_j \right], \\ \bar{\theta_2} &= \alpha \sum_{j=2}^{B} \mathbb{E}\left[\boldsymbol{\beta}_j \left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}} \right) \right] \end{aligned}$$

and using Lemma A.1.1 we can conclude that,

$$\mathbf{H}_1^H \mathbf{S}^{-1} \mathbf{H}_1 \rightarrow \frac{1}{M} \operatorname{trace}\{\mathbf{S}^{-1}\} \mathbf{I} = \eta_1 \mathbf{I}, \tag{A.16}$$

$$\mathbf{H}_1^H (\mathbf{S}^{-1})^2 \mathbf{H}_1 \rightarrow \frac{1}{M} \operatorname{trace}\{\mathbf{S}^{-2}\} \mathbf{I} = \eta_2 \mathbf{I} \tag{A.17}$$

almost surely as $M \rightarrow \infty$ where, if G is the non-random limiting distribution of the eigenvalues λ of the matrix $\mathbf{S}_1 \mathbf{D}_1 \mathbf{S}_1^H$, then

$$\eta_1 = \int \frac{1}{\lambda + \bar{\theta_1} + \bar{\theta_2} + \sigma^2} dG(\lambda) \text{ and} \tag{A.18}$$

$$\eta_2 = \int \frac{1}{(\lambda + \bar{\theta_1} + \bar{\theta_2} + \sigma^2)^2} dG(\lambda). \tag{A.19}$$

From equations (A.18), (A.19) and using the definition of Stieltjes transform [48] we can find that

$$\eta_1 = \lim_{z \rightarrow -\bar{\theta_2} - \bar{\theta_2} - \sigma^2} m(z),$$

and since $m(z)$ is complex analytic

$$\eta_2 = \lim_{z \rightarrow -\bar{\theta_2} - \bar{\theta_2} - \sigma^2} \frac{d}{dz} m(z).$$

The values of η_1 and η_2 can then also be from obtained from solving equation (A.15) and its derivative. The equations are given by,

$$\eta_1 = \left(\sigma^2 + \alpha\mathbb{E}[\boldsymbol{\beta}] - \alpha\mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2 \eta_1}{1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1} \right] \right)^{-1}, \tag{A.20}$$

$$\eta_2 = \left(\eta_1^{-2} - \alpha\mathbb{E}\left[\left(\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}}{1+\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1} \right)^2 \right] \right)^{-1}. \tag{A.21}$$

Similarly, in the expression for interference P_{inter},

$$\begin{aligned}
\text{trace}\{\mathbf{Z}\} &= \text{trace}\left\{ \sum_{j=1}^{B}\sum_{k=2}^{K} \beta_{jk}\mathbf{S}^{-1}\mathbf{H}_k\mathbf{e}_j\mathbf{e}_j^H\mathbf{H}_k^H\mathbf{S}^{-1} + \sum_{j=2}^{B}\beta_{j1}\mathbf{S}^{-1}\mathbf{H}_1\mathbf{e}_j\mathbf{e}_j^H\mathbf{H}_1^H\mathbf{S}^{-1} \right\} && \text{(A.22)}\\
&\overset{(a)}{=} \sum_{j=1}^{B}\sum_{k=2}^{K} \beta_{jk}\mathbf{e}_j^H\mathbf{H}_k^H\left(\mathbf{S}^{-1}\right)^2\mathbf{H}_k\mathbf{e}_j + \sum_{j=2}^{B}\beta_{j1}\mathbf{e}_j^H\mathbf{H}_1^H\left(\mathbf{S}^{-1}\right)^2\mathbf{H}_1\mathbf{e}_j && \text{(A.23)}\\
&\overset{(b)}{=} \sum_{j=1}^{B}\sum_{k=2}^{K} \beta_{jk}\mathbf{e}_j^H\mathbf{H}_k^H\mathbf{S}_k^{-2}\left(\mathbf{I} - 2\frac{\nu_k\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}}{1+\nu_k^2\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k} + \right.\\
&\quad \left. \frac{\nu_k^2\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}}{(1+\nu_k^2\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k)^2} \right)\mathbf{H}_k\mathbf{e}_j + \sum_{j=2}^{B}\beta_{j1}\mathbf{e}_j^H\mathbf{H}_1^H\mathbf{S}^{-2}\mathbf{H}_1\mathbf{e}_j && \text{(A.24)}\\
&= \sum_{j=1}^{B}\left(\sum_{k=2}^{K}\beta_{jk}\mathbf{e}_j^H\mathbf{H}_k^H\mathbf{S}_k^{-2}\mathbf{H}_k\mathbf{e}_j - 2\sum_{k=2}^{K}\beta_{jk}\frac{\nu_k\mathbf{e}_j^H\mathbf{H}_k^H\mathbf{S}_k^{-2}\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{e}_j}{1+\nu_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k} + \right.\\
&\quad \left. \sum_{k=2}^{K}\beta_{jk}\frac{\nu_k^2\mathbf{e}_j^H\mathbf{H}_k^H\mathbf{S}_k^{-2}\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{e}_j}{(1+\nu_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k)^2} \right) + \\
&\quad \sum_{j=2}^{B}\beta_{j1}\mathbf{e}_j^H\mathbf{H}_1^H\mathbf{S}^{-2}\mathbf{H}_1\mathbf{e}_j && \text{(A.25)}
\end{aligned}$$

where, in step (a) we use $\text{trace}\{\mathbf{q}\mathbf{q}^{\mathbf{H}}\} = \mathbf{q}^{\mathbf{H}}\mathbf{q}$ and if

$$\mathbf{S}_k = \sum_{k\neq 1,k}\nu_k\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H + (\theta_1+\theta_2+\sigma^2)\mathbf{I} \tag{A.26}$$

then in step (b) use matrix inversion lemma as,

$$\mathbf{S}^{-1} = \mathbf{S}_k^{-1}\left(\mathbf{I} - \frac{\nu_k\mathbf{H}_k\mathbf{a}_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}}{1+\nu_k\mathbf{a}_k^H\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{a}_k} \right). \tag{A.27}$$

Notice that using Lemma (A.1.1), the terms $\mathbf{H}_k^H\mathbf{S}_k^{-1}\mathbf{H}_k \xrightarrow{a.s.} \eta_1\mathbf{I}$ and $\mathbf{H}_k^H\mathbf{S}_k^{-2}\mathbf{H}_k \xrightarrow{a.s.} \eta_2\mathbf{I}$ and it appears repeatedly in equation (A.25). Therefore, in the limit of infinite number

of antennas, again using Lemma A.1.1 we have

$$
\begin{aligned}
\mathbf{H}_1^H \mathbf{Z} \mathbf{H}_1 \quad &\xrightarrow{a.s.} \quad \frac{1}{M} \operatorname{trace}\{\mathbf{Z}\}\mathbf{I} \\
&\xrightarrow{a.s.} \quad \left(\alpha \sum_{j=1}^{B} \mathbb{E} \left[\boldsymbol{\beta}_j \left(\eta_2 - 2 \frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^2 \boldsymbol{\beta}_j \, \eta_1 \eta_2}{1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}} + \frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^4 \boldsymbol{\beta}_j \, \boldsymbol{\beta} \, \eta_2 \eta_1^2}{\left(1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2} \right) \right] \right) \mathbf{I} \\
&= \quad \left(\alpha \sum_{j=2}^{B} \mathbb{E}\left[\boldsymbol{\beta}_j\right] \eta_2 + \alpha \sum_{j=2}^{B} \mathbb{E}\left[\boldsymbol{\beta}_j \frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right] \eta_2 + \alpha \mathbb{E}\left[\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right] \eta_2 \right. \\
&\qquad \left. - \alpha \mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^2 \sum_{j=1}^{B} \boldsymbol{\beta}_j \, \eta_1 \eta_2}{1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}} \right] - \alpha \mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^2 \sum_{j=1}^{B} \boldsymbol{\beta}_j \, \eta_1 \eta_2}{\left(1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2} \right] \right) \mathbf{I} \\
&= \quad \left((\bar{\theta_1} + \bar{\theta_2}) \eta_2 + \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_1\right)^2} \right] \right. \\
&\qquad \left. - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \left(\sum_{j=2}^{B} \frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}} \right) \eta_1 \eta_2}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_1} \right] - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \left(\sum_{j=2}^{B} \frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}} \right) \eta_1 \eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_1\right)^2} \right] \right) \mathbf{I} \qquad \text{(A.28)} \\
&= \quad \left(\int_0^{\infty} \frac{\lambda + \bar{\theta}_1 + \bar{\theta}_2}{(\lambda + \bar{\theta}_1 + \bar{\theta}_2 + \sigma^2)^2} dG(\lambda) \right. \\
&\qquad \left. - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \left(\sum_{j=2}^{B} \frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}} \right) \eta_1 \eta_2}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_1} \right] - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \left(\sum_{j=2}^{B} \frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}} \right) \eta_1 \eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_1\right)^2} \right] \right) \mathbf{I}. \qquad \text{(A.29)}
\end{aligned}
$$

In equation (A.28) the third term $\mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}} \eta_1\right)^2} \right]$ is equal to $\int_0^{\infty} \frac{\lambda}{(\lambda + \bar{\theta}_1 + \bar{\theta}_2 + \sigma^2)^2} dG(\lambda)$. This follows from corollary A.1.4.

Using equations (A.16), (A.17), (A.29) in expressions (A.7), (A.8), (A.9) and (A.11) the SINR given in equation (3.16) converges almost surely to $\widehat{\mathsf{SINR}}$ as in expression (3.17).

A.3 Proof of Proposition 3.3.3

Using the matched filter given by $\mathbf{c} = \mathbf{H}_1 \mathbf{a}_1$, the signal power, noise power, pilot interference power is given by $P_{\mathsf{signal}} = \left| \mathbf{a}_1 \mathbf{H}_1^H \mathbf{H}_1 \mathbf{e}_1 \right|^2$, $P_{\mathsf{noise}} = \sigma^2 \mathbf{a}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{a}_1$, $P_{\mathsf{pilot}} =$

$\sum_{j=1}^{B} \beta_{j1} \left|\mathbf{a}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{e}_j\right|^2$. Since, $\mathbf{H}^H\mathbf{H} \xrightarrow{a.s.} \mathbf{I}$ as $M \to \infty$, we have, $P_{\mathsf{signal}} \xrightarrow{a.s.} \beta_{11}$, $P_{\mathsf{noise}} \xrightarrow{a.s.} \sigma^2 \sum_{j=1}^{B} \beta_{j1}$, $P_{\mathsf{pilot}} \xrightarrow{a.s.} \sum_{j=2}^{B} \beta_{j1}^2$. Using Lemma A.1.1 in the interference term we have,

$$
\begin{aligned}
P_{\mathsf{inter}} &= \mathbf{a}_1 \mathbf{H}_1^H \left(\sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_k \mathbf{h}_k^H \right) \mathbf{H}_1 \mathbf{a}_1 && \text{(A.30)} \\
&\xrightarrow{a.s.} \mathbf{a}_1 \left(\frac{1}{M} \operatorname{trace} \left\{ \sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_k \mathbf{h}_k^H \right\} \mathbf{I} \right) \mathbf{a}_1 && \text{(A.31)} \\
&= \alpha \sum_{j=1}^{B} \beta_{j1} \frac{1}{K} \left(\sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_k^H \mathbf{h}_k \right) && \text{(A.32)} \\
&= \alpha \sum_{j=1}^{B} \beta_{j1} \left(\sum_{j=1}^{B} \mathbb{E}[\boldsymbol{\beta}_j] \right) && \text{(A.33)}
\end{aligned}
$$

Rearranging the terms in equation (3.16) we get the expression for $\overline{\mathsf{SINR}}$.

References

[1] S. Kumar, N. Marchetti, "IMT-Advanced: Technological requirements and solution components", *2nd Int. Workshop Cognitive Radio Advanced Spectrum Manage.*, pp. 1-5, May 2009.

[2] M. K. Karakayali, G. J. Foschini, R. A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems", *IEEE Wireless Commun.*, vol. 13, pp. 56-61, Aug. 2006.

[3] D. Astely, E. Dahlman, P. Frenger, R. Ludwig, M. Meyer, S. Parkvall, P. Skillermark, N. Wiberg "A Future Radio-Access Framework", *J. Sel. Areas Commun.*, vol. 24, March 2006.

[4] S. Parkvall, E. Dahlman, A. Furuskar, Y. Jading, M. Olsson, S. Wanstedt, K. Zangi, "LTE-Advanced - Evolving LTE towards IMT-Advanced", *IEEE 68th Vehicular Tech. Conf.* pp. 1-5, Sept. 2008.

[5] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, S. Parkvall, "LTE: The Evolution of Mobile Broadband", *IEEE Commun. Mag.*,vol. 47, pp. 44-51, April 2009.

[6] S.W. Peters, R.W. Heath, "The future of WiMAX: Multihop relaying with IEEE 802.16j", *IEEE Commun. Mag*, vol. 47, pp. 104-111, Jan. 2009.

[7] Y. Yang, Honglin Hu and Jing Xu, "Relay Technologies for WiMAX and LTE-Advanced Mobile Systems", *IEEE Commun. Mag., Oct. 2009.*

[8] L. Long; E. Hossain; "Multihop Cellular Networks: Potential Gains, Research Challenges, and a Resource Allocation Framework", *IEEE Commun. Mag*, vol. 45, pp. 66-73, Sep. 2007.

[9] Chris T. K. Ng, G. J. Foschini; "Transmit Signal and Bandwidth Optimization in Multiple-Antenna Relay Channels", [Online] Available: `http://arxiv.org/PS_cache/arxiv/pdf/1001/1001.2938v1.pdf`

[10] D. Zhang, O. Ileri, N. B. Mandayam. "Bandwidth Exchange as an Incentive for Relaying", *42nd Annual Conference on Information Sciences and Systems (CISS)* 2008, pp. 749 - 754, Princeton, NJ, Mar. 2008.

[11] D. Zhang, R. Shinkuma, and N. B. Mandayam, "Bandwidth Exchange: An Energy Conserving Incentive Mechanism for Cooperation". *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, Jun 2010.

[12] F. Richter, A.J. Fehske, G.P. Fettweis; "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks", *IEEE Veh. Tech. Conf.* pp. 1-5, Sept. 2009.

[13] J. Acharya, *"Utility Maximization in Dynamic Spectrum Allocation"*, Phd Thesis, Rutgers University Dept. of Electrical and Computer Engineering, May 2009.

[14] T. Girici, "Joint power, subcarrier and subframe allocation in Multihop relay networks", *Int. J. Commun. Syst.*, vol. 22, pp: 835 - 855, Jan. 2009.

[15] 3GPP TR 36.814: "Further Advancements for E-UTRA Physical Layer Aspects (Release 9)". [Online], Available: `http://www.3gpp.org/ftp/Specs/html-info/36-series.htm`

[16] M. Hart et al., "Out-of-Band Relay Clarification", IEEE C802.16j-08/079r1, 2008.

[17] S. Boyd, L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[18] M. Grant, S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 1.21", `http://cvxr.com/cvx`, May 2010.

[19] R. Madan, S. P. Boyd, ; S. Lall, "Fast Algorithms for Resource Allocation in Wireless Cellular Networks", *IEEE/ACM Trans. Networking*, vol. 18, pp. 973-984, Jun 2010.

[20] 3GPP TS 36.101: "User Equipment (UE) radio transmission and reception". [Online], Available: `http://www.3gpp.org/ftp/Specs/html-info/36-series.htm`.

[21] 3GPP TS 36.104: "Base Station (BS) radio transmission and reception". [Online], Available: `http://www.3gpp.org/ftp/Specs/html-info/36-series.htm`.

[22] T. Novlan, J.G. Andrews, Illsoo Sohn, R.K. Ganti, A. Ghosh, "Comparison of Fractional Frequency Reuse Approaches in the OFDMA Cellular Downlink", *IEEE Globecom Conf.*, pp. 1-5, Dec. 2010.

[23] N. Chun, L. Pei, S. Panwar, "Interference management using frequency planning in an OFDMA based wireless network", *IEEE Wireless Commn. Netw. Conf.*, pp. 998-1003, Mar. 2011.

[24] N. Krishnan, J. S. Panchal, N. B. Mandayam, R. D. Yates, "Bandwidth Sharing in LTE-A systems", *48^{th} Annual Allerton Conf. on Commun. Control and Computing*, Oct. 2010.

[25] B. Hassibi, B. M. Hochwald, "How Much Training is Needed for Multiple-Antenna Wireless Links?", *IEEE Trans. Info. Th.*, vol. 49, pp. 951-963, Apr. 2003.

[26] T. L. Marzetta, "How Much Training is Required for Multiuser MIMO?", *Asilomar Conf. Signal Syst. Comput.*, pp. 359-363, Nov. 2006.

[27] T. L. Marzetta, "Non-cooperative Multiantenna Base Stations with Unlimited Number of Antennas", *IEEE Trans. on Wireless Commun.*, vol. 9, pp. 3590-3600, Nov. 2010

[28] J. Jose, A. Ashikhmin, T. L. Marzetta, S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems, in *Proc. of IEEE ISIT*, June 2009.

[29] J. Jose, A. Ashikhmin, T. L. Marzetta, S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems", *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2640-2651, Aug. 2011

[30] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, F. Tufvesson, "Scaling up MIMO: Opportunities and Challenges with very large arrays", *IEEE Signal Processing Mag.*, vol. 30, pp. 40-60, Jan. 2013

[31] H. Q. Ngo, E. G. Larsson, T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems", arXiv:1112.3810v1.

[32] F. Fernandes, A. Ashikhmin, T. L. Marzetta, "Inter-Cell Interference in Noncooperative TDD Large Scale Antenna Systems", *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 192-201, Feb. 2013.

[33] H. Q. Ngo, M. Matthaiou, T. Q. Duong, E. G. Larsson, "Uplink Performance Analysis of Multicell MU-MIMO Systems with ZF Receivers", *IEEE Swedish Commun. Tech. Workshop*, pp. 40 - 45, Oct. 2011

[34] J. Zhang, C. K. Wen, S. Jin, X. Gao, K. Wong, "On Capacity of Large-Scale MIMO Multiple Access Channels with Distributed Sets of Correlated Antennas" *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 133-148, Feb. 2013.

[35] D. Tse, S. Hanly, "Linear multiuser receivers: Effective interference, Effective bandwidth and User capacity", *IEEE Trans. Inform. Th.*, vol. 45, pp. 641-657, Mar. 1999

[36] J. Zhang, E.K. P. Chong, D. N. C. Tse, "Ouput MAI Distributions of Linear MMSE Multiuser Receivers in DS-CDMA Systems", *IEEE Trans. Info. Th.*, vol. 47, pp. 1128 - 1144, March 2001.

[37] J. Evans, D. N. C. Tse, "Large System Performance of Linear Multiuser Receivers in Multipath Fading Channels", *IEEE Trans. Info. Th.*, vol. 46, pp. 2059 - 2077, Sept. 2000.

[38] J. Hoydis, S. ten Brink, M. Debbah, "Massive MIMO in UL/DL of Cellular Networks: How Many Antennas Do We Need?", *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 160-171, Feb. 2013.

[39] D. Samardzija, N. B. Mandayam, "Impact of Pilot Design on Achievable Data Rates in Multiple Antenna Multiuser TDD System", *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1370-1379, Sept. 2007.

[40] H. Q. Ngo, E. G. Larsson, T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays", *49th Allerton Conf. Commun., Control, Comput.*, pp. 1272-1279 , Sept. 2011

[41] N. Krishnan, R. D. Yates, N. B. Mandayam, "Cellular System with Many Antennas: Large System Analysis under Pilot Contamination", *50th Allerton Conf. Commun., Control, Comput.*, pp. 1220-1224, Oct. 2012

[42] R. Mueller, M. Vekhapera, L. Cottatellucci, "Blind pilot decontamination", *17th Intl ITG Workshop on Smart Antennas (WSA)*, March 2013.

[43] R. Mueller, M. Vehkapera, L. Cottatellucci, "Analysis of Pilot Decontamination Based on Power Control", *IEEE Vehicular Tech. Conf.* , June 2013

[44] H. Huh, G. Caire, H.C. Papadopoulos, S. A. Ramprashad, "Achieving "Massive MIMO" Spectral Efficiency with a Not-so-Large Number of Antennas", *IEEE Transactions on Wireless Communication*, vol. 11, pp. 3226-3239, Sept. 2012

[45] S. Verdu, "*Multiuser Detection*", Cambridge University Press.

[46] D. Tse, P. Viswanath, "*Fundamentals of Wireless Communications*", Cambridge University Press.

[47] H. V. Poor, "*An Introduction to Signal Detection and Estimation*", Springer Texts in Electrical Engineering.

[48] A. M. Tulino, S. Verdu, "Random Matrix Theory for Wireless Communications", *Foundations and Trends in Communications and Information Theory*, vol. 1, 2004.

[49] Y. Lin, L. Shao,Z. Zhu, Q. Wang, R.K. Sabhikhi, "Wireless network cloud: Architecture and system requirements", *IBM Journal of Research and Development*, vol. 54,pp. 4:1-4:12

[50] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu,G. Kumar, A. Muralidhar1 P. Polakos, V. Srinivasan, T. Woo, "CloudIQ: A Framework for Processing Base Stations in a Data Center", *Int. Conf. Mob. Comput. Net.* , Aug. 2012.

[51] P. Gupta, A. Vishwanath, S. Kalyanaraman and Y. H. Lin, "Unlocking Wireless Performance with Co-operation in Co-located Base Station Pools", *Int. Conf. on Commun. Sys Net.*, pp. 1-8, Jan. 2010.

[52] J. Zhang, J. G. Andrews,"Distributed Antenna Systems with Randomness", *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3636-3646,Sept. 2008

[53] J. Zhang, C. K. Wen, S. Jin, X. Gao, K. K. Wong, "On Capacity of Large Scale MIMO Multiple Access Channels with Distributed Sets of Correlated Antennas", *IEEE J. Sel. Areas. Commun.*, vol. 31, pp. 133-148, Feb. 2013.

Vita

Narayanan Krishnan

Education

2009-13 Ph. D. in Electrical and Computer Engineering, Rutgers University.

2007-09 M. Sc. in Electrical and Computer Engineering, Kansas State University.

2000-04 B. Tech in Electronics and Communication Engineering,
College of Engineering, Trivandrum.

Employment

2009-13 Graduate Assistant, WINLAB, Rutgers University, NJ, USA.

2012 Interim Engineering Intern, Qualcomm Incorporated, NJ, USA.

2007-09 Research Assistant, Dept. of ECE, Kansas State University.

2004-07 Technical Specialist, Infosys Technologies Ltd., Bangalore, India.

Relevant Publications

N. Krishnan, Narayan B. Mandayam, Roy D. Yates, Jignesh S. Panchal, "Bandwidth Sharing for LTE-A Relaying Systems", *48^{th} Annual Allerton Conf. on Commun. Control Computing*, Oct. 2010.

N. Krishnan, Roy D. Yates, Narayan B. Mandayam, Jignesh S. Panchal, "Bandwidth Sharing for Relaying in Cellular Systems", *IEEE Transactions on Wireless Communication*, vol. 11, pp. 117-129, Jan. 2012.

N. Krishnan, Roy D. Yates, Narayan B. Mandayam, "Cellular Systems with Many Antennas: Large System Analysis under Pilot Contamination", *50^{th} Annual Allerton Conf. on Commun. Control Computing*, Oct. 2012.

N. Krishnan, Roy D. Yates, Narayan B. Mandayam, "Uplink Linear Receivers for Multi-cell Multiuser MIMO with Pilot Contamination: Large System Analysis", *submitted to IEEE Trans. Wireless Commun.*.

N. Krishnan, Roy D. Yates, Narayan B. Mandayam, "Massive MIMO TDD Systems: Uplink Training Dependence on Antennas", *to be submitted*