University of Alberta

EFFICIENT TRANSMISSION AND RESOURCE ALLOCATION METHODS FOR MULTI-USER MIMO DOWNLINK

by

Boon Chin Lim

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in Communications

Department of Electrical and Computer Engineering
Edmonton, Alberta
Fall 2008

Your file Votre référence
ISBN: 978-0-494-46361-1
Our file Notre référence
ISBN: 978-0-494-46361-1

To Linda, Annette, Quianna, my parents and the Almighty

ABSTRACT

The consideration of multiple-input multiple-output (MIMO) techniques for the cellular downlink of future wireless systems is motivated by the demand for high transmission rates to multiple users over limited frequency spectrum. Although the optimal approach for MIMO broadcast channels is dirty paper coding (DPC), it incurs very high complexity that limits its practicality. This thesis focuses on enhancing the feasibility of deploying multi-user MIMO techniques in practical downlink systems. In line with this, emphasis is placed on improving the performance of transmit zero-forcing beamforming (TZFBF), which has lower complexity but is sub-optimal.

To narrow the sum rate performance gap with DPC systems, it is shown that receive antenna selection (RAS) is necessary for maximizing the achievable sum rate for TZFBF systems. This is true for TZFBF systems with multi-antenna terminals even when all receive antennas are equipped with RF chains and RAS reduces the upper bound on the broadcast sum capacity, and when the orthogonalized channels use optimal processing. Similarly, spatial mode selection (SMS) is necessary when receive-weight matrices are used for spatial mode allocation. Significantly, RAS/SMS helps to reduce the performance gap even for small user pool sizes. Optimal user selection for sum rate maximization is subsumed within an optimal RAS/SMS process for multi-antenna terminals and both selection processes become identical for single-antenna terminals. For a system with M transmit antennas, RAS/SMS increases the probability of scheduling M spatial modes compared to the case with sole reliance on user selection, especially when the potential user pool is small.

However, optimal RAS/SMS incurs very large system overhead because the channel matrices of all potential users must be fed back to the base station. Another challenge is posed by the flexibility for spatial mode allocation at multi-antenna terminals to meet individual transmission rate requirements. A streamlined process that encompasses efficient selection with feedback reduction and systematic resource allocation with rate loss minimization, is developed for the sum rate maximization of TZFBF systems. In addition, bounds are developed for estimation of the ergodic TZFBF sum rates where RAS/SMS, user selection, signal-to-noise ratio and number of transmit elements are taken into account.

ACKNOWLEDGEMENT

It is my great privilege to have worked with my supervisors, Dr. Witold A. Krzymień and Dr. Christian Schlegel and I thank them for their very valuable guidance that has charted the development of this work. I thank Dr. Sofiène Affes, Dr. Ioanis Nikolaidis, Dr. Vincent Gaudet, and Dr. Sergiy Vorobyov for their very thorough and constructive comments and recommendations. I am grateful to DSO National Laboratories for giving me the opportunity and providing the financial support that has made this pursuit possible.

I deeply appreciate Dr. David Mazzarese for his constructive comments during the early stages of this work. My heartfelt thanks go to Mohsen Eslami and Kevin Jacobson for the many insightful discussions and their friendship. I would also like to thank my colleagues Dr. Shreeram Sigdel, Robert Elliot, Dr. Dmitri Trukhachev, Sumeeth Nagaraj, Lukasz Krzymień and others in TRLabs and HCDC Lab who have assisted me.

I am very grateful to Chow Seng and Grace Liu for their great hospitality especially during the last year of this program. Many thanks go out also to Dr. Yan Xin, Dr. Fengqin Zhai, Yvonne and William Lim, Kim and William Lim, I.T. and Leng, Mike and Katie Gao, the Phoa family and many other friends here in Edmonton for their friendship and support. I treasure the memories of my friends at the McLaurin Memorial Baptist Church and thank them for their fellowship and prayers.

My heartfelt gratitude goes to my wife Linda for her constant encouragement and love, and to my daughters Annette and Quianna for bringing me cheer and comic relief.

This academic journey has shown me aspects of the glory of the one who is called Wonderful Counselor.

Table of Contents

List of Figures

List of Acronyms

BAS	Block antenna selection
BC	Broadcast channel
BD	Block diagonalization
BD-SDM	Block diagonalized space-division multiplexing
BMS	Block mode selection
BF	Beamforming
BS	Base station
CBM	Correlation based method
CCI	Co-channel interference
CDF	Cumulative density function
CSI	Channel state information
CSIFR	Channel state information feedback reduction
CTR	Coordinated transmit receive
DPC	Dirty paper coding
DRAS	Decremental receive antenna selection
DRAS-SAS	Decremental receive antenna selection done on a single antenna selection basis
DSEL	Decoupled transmit antenna selection and user selection (decoupled TAS-USEL
GDS	Gorokhov decremental selection
GIS	Gorokhov incremental selection
FDM	Frequency-division multiplexing
i.i.d.	Independent and identically-distributed

IRAS	–	Incremental receive antenna selection
IRAS-SAS	–	Incremental receive antenna selection done on a single antenna selection basis
JREUS	–	Joint rate evaluation and user selection
JREUS-MAX	–	JREUS version that is guided by the $\max(1/b_i)$ value.
JREUS-ALT1	–	JREUS version that is also guided by the second highest $(1/b_i)$ value.
JREUS-ALT2	–	JREUS version that is also guided by the third highest $(1/b_i)$ value.
MDR	–	Maximum determinant ranking
MDSEL	–	Modified DSEL = modified decoupled transmit antenna selection and user selection
MEMS	–	Micro-electromechanical systems
MIBM	–	Mutual information based method
MIMO	–	Multiple-input multiple-output
MIMO-BC	–	MIMO broadcast channel
MMS	–	Multimedia messaging service
MMSE	–	Minimum mean square error
MUD	–	Multi-user diversity
NB-CSIFR	–	Norm-based channel state information feedback reduction
NBS	–	Norm-based selection
Nu-SVD	–	Null space directed singular value decomposition
OFDM	–	Orthogonal frequency-division multiplexing
PDF	–	Probability density function
PF-TCIBF	–	Proportional fair transmit channel inversion beamforming
PMP	–	Pairwise mutual projection

PSME	–	Poorest spatial mode elimination
QoS	–	Quality of service
RAS	–	Receive antenna selection
RBF	–	Random beamforming
RF	–	Radio frequency
RR-TCIBF	–	Round-robin transmit channel inversion beamforming
SAS	–	Single antenna selection
SDM	–	Space-division multiplexing
SINR	–	Signal-to-interference-plus noise ratio
SISO	–	Single-input single-output
SMS	–	Spatial mode selection
SNIP	–	Squared-normalized inner products
SNR	–	Signal-to-noise ratio
SSRM or SRM	–	Simultaneous scheduling and sum rate maximization
SVD	–	Singular value decomposition
TAS	–	Transmit antenna selection
TCIBF	–	Transmit channel inversion beamforming
TDM	–	Time-division multiplexing
THP	–	Tomlinson-Harashima pre-coding
TZFBF	–	Transmit zero-forcing beamforming
USEL	–	User selection
ZF	–	Zero-forcing
ZMCSCG	–	Zero mean, centrally symmetrical, complex Gaussian

Chapter 1

INTRODUCTION

The wireless communications industry has witnessed phenomenal worldwide growth since the early 1990s. Cellular communications networks in particular have become pervasive and the high subscriber rates testify to the fact that wireless communications is a reliable and viable transport mechanism for voice and data. This widespread success has given impetus to the development of newer wireless systems and standards for many other types of telecommunication traffic. In fact, cellular telephones began to evolve into something more than the wireless version of the telephone invented in the 19th century when the short message service was introduced. From this beginning of mobile convergence, the introduction of wireless access to the Internet is driving digital convergence in all conceivable areas, including digital audio broadcasting, digital video broadcasting, peer-to-peer multimedia communications, multimedia messaging service (MMS), 3D audio and video on demand [1]. The convergence is accelerated with the push towards the vision of a highly developed ubiquitous information society that incorporates paradigms such as ubiquitous networks, pervasive computing, and ambient intelligence. The content-rich lifestyle envisaged under these visions and paradigms will not only place heavy demands on the wireless communications infrastructure but also demands for more frequency spectrum, which is a very limited resource. In the cellular context, content-rich applications will result in high data-rate transmissions especially on downlinks, that is, data delivered from base stations to user terminals. A similar push has also been happening in the domain of military communication systems. It has been more than a decade since the first calls have been issued for the development of military communication systems that enable network-centric operations. There is a desire to realize the full potential of a robust ubiquitous military network that provides the benefits of Internet-like connectivity and flexibility. The development of such a network is still fraught with fundamental challenges like severe spectrum limitations [2].

It is commonly recognized that the "sweet-spot" of the radio spectrum is approximately between 300 and 5000 MHz where the bandwidth and propagation characteristics are favorable for the delivery of broadband and mobile wireless services. The demand for this spectral range is very high as more and more wireless services are being introduced or proposed. Although more bandwidth is available in the higher frequency bands, they tend to suffer from higher propagation losses and place limitations on coverage and reach. High-gain, directional antennas may be used in some cases to address the range issue. However, for cellular systems the use of highly directional antennas is impractical. Although the cellular concept allows for frequency re-use, the growing user density and the push for content-rich applications has spurred research efforts to extract as much as possible from every piece of spectral real estate.

In this respect, new ground in information and communications theory was broken in the mid 1990s with important breakthroughs beginning with the introduction of spatial multiplexing by Paulraj and Kailath in 1994 [3] and the subsequent fundamental research done at Bell Labs (for example, [4] and [5]). Under scatter-rich propagation conditions, huge data-rate gains over conventional point-to-point wireless systems were obtained by employing multiple-antenna arrays at both the transmitter and receiver [5]-[6]. Such multiple-antenna wireless systems have become commonly referred to as "multiple-input multiple-output" or MIMO systems. The block diagram of a typical point-to-point MIMO system is shown in Figure 1.1 where the transmitter has M antennas and the receiver has N antennas.

The high data-rate gains are obtained by separating signals that are transmitted in the same frequency channel from spatially separated transmitters. This creates parallel channels within the same channel bandwidth and the resulting capacity increase is linear with the array size rather than the conventional logarithmic increase with increasing signal-to-noise ratio. Under a scatter-rich environment where each path between a pair of transmit- and receive antennas experiences independent and identically distributed (i.i.d.) Rayleigh fading, the single-user capacity C_{SU} of the $M \times N$ point-to-point MIMO system in Figure 1.1 can be written as [5], [7]

$$C_{\mathrm{SU}} = \log_2 \det\left(\mathbf{I} + (\mathrm{SNR})\mathbf{H}\mathbf{H}^H\right) \quad \text{bits/sec/Hz}, \tag{1.1}$$

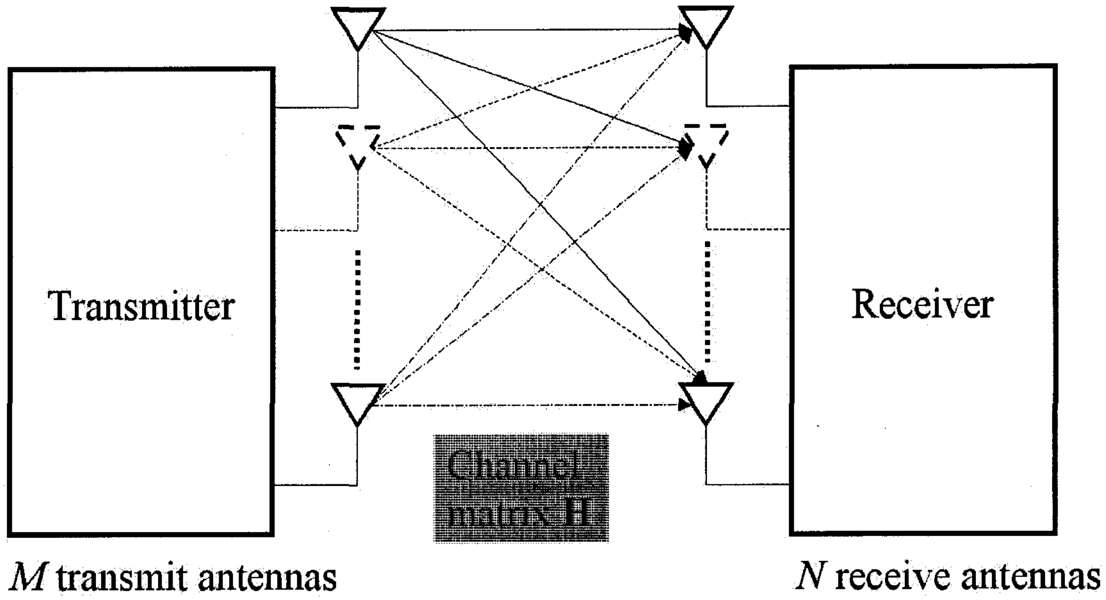


Figure 1.1: Block diagram of a single-user, point-to-point MIMO system

where **H** is the $N \mathrm{x} M$ channel matrix. When the signal-to-noise ratio (SNR) is high, the average or ergodic single-user capacity $\bar{C}_{\mathrm{SU}}$ is

$$\bar{C}_{\mathrm{SU}} = \min\left(M,N\right)\log_2 \mathrm{SNR} + O(1) \quad \text{bits/sec/Hz}, \tag{1.2}$$

where $O(1)$ represents a term with negligible contribution. Equation (1.2) shows that the capacity scales with $\min(M,N)$ bits/sec/Hz whenever the SNR increases by 3dB.

Departing from the traditional approach where antenna arrays are used in tandem with processing algorithms to mitigate the ill-effects of multipath propagation, MIMO systems exploit the multipath environment and excel under scatter-rich environments. MIMO systems can also be used to provide diversity gain rather than capacity gain. This is typically achieved by designing codes that improve link reliability by spreading data over both space- and time dimensions and is commonly referred to as space-time coding. In this dissertation, we will be focusing on the issues relating to the utilization of MIMO systems for spatial multiplexing to achieve capacity gains.

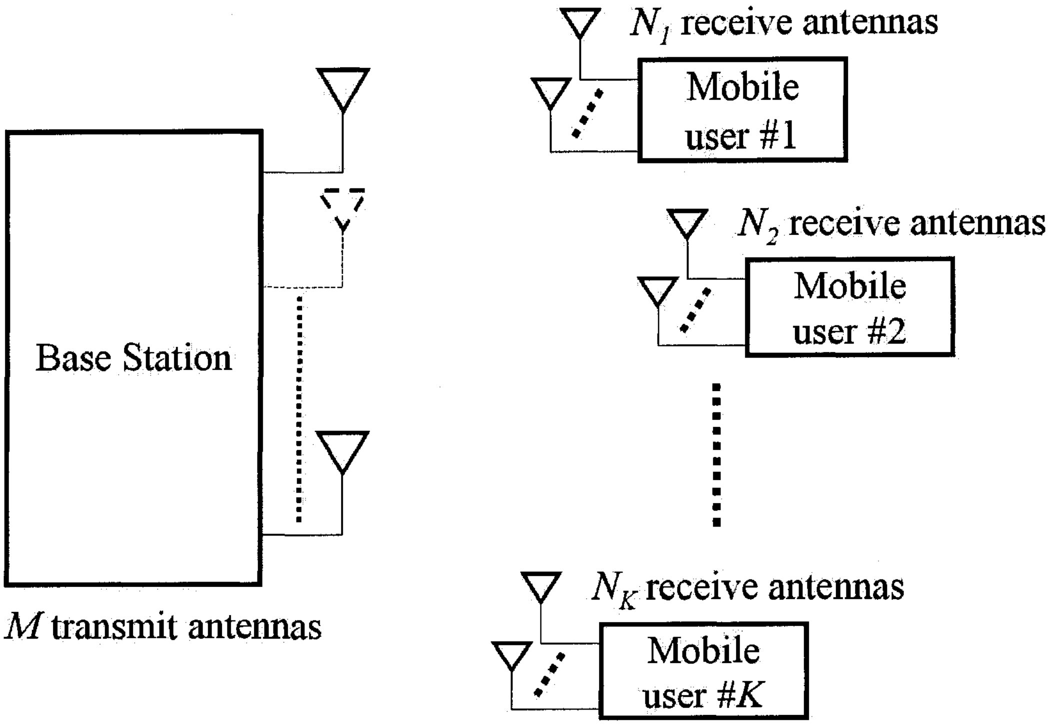


Figure 1.2: Block diagram of a multi-user MIMO system

To serve multiple users in cellular systems, traditional multiplexing methods for the downlink (from base station to the users) and multiple access methods for the uplink (from the users back to the base station) are based on one or a combination of the following, namely, time-division, frequency-division and code-division. MIMO systems have opened the spatial dimension for point-to-point systems and their consideration for the multi-user environment is a natural extension. It may appear that combining a MIMO spatial multiplexing scheme, that is, space-division multiplexing (SDM) with time-division multiplex (TDM) is a viable configuration to serve multiple users in the cellular downlink shown in Figure 1.2. This is however of limited utility primarily because of physical limitations at the user terminals. To see why this is so, consider first that the capacity of a point-to-point MIMO system scales almost linearly with $\min(M,N)$ over the single-input single-output (SISO) case, where M and N are the number of transmit and receive antennas, respectively [6]. The TDM regime will constrain the base station to see only one user at a time and the scaling with $\min(M,N)$ applies, where N is the number of antennas at each user terminal. Although it is possible to equip the base station with more antennas, mobile or nomadic terminals have limited space and power. For example, most existing mobile platforms have only one antenna and the configuration will only result in logarithmic capacity gain. Consequently, serving multiple users via a combination of MIMO techniques and TDM will not help achieve capacity that scales with *M*, the number of base station antennas.

Still, it appears reasonable that by exploiting the differences in spatial signatures under a scatter-rich environment, MIMO systems can create multiple channels in the spatial domain to serve multiple users simultaneously from a single base station using the same frequency channel and within the same time slot. For point-to-point MIMO systems, achieving a capacity that scales with $\min(M, N)$ requires channel state information (CSI) to be available at the receiver, but is not necessary at the transmitter. The channel state information comprises details on the channel gain and phase shift from each transmit antenna to each receive antenna. For the multi-user situation, spatial multiplexing to multiple users from a single base station is possible when full CSI is also available at the transmitting base station. This stems mainly from the fact that the transmitter is then able to account for the inter-user interference among the users when catering for the channel rate requirement of each user. CSI at the base station is not only useful in achieving the required SNR at a desired user but also in reducing the resultant interference at other points of the system. A survey on the various schemes to accomplish this will be given in the next section. CSI at the base station may be obtained via (a) the use of training or pilot data in the uplink for time-division duplex systems or (b) feedback of each user's channel estimates done using training data in the downlink for frequency-division duplex systems. Acquiring CSI at the transmitting base station using either method is challenging and incurs substantial overhead but appears to be justifiable due to sum rate gains in the multi-user environment.

A performance parameter of interest is the sum rate, which is the sum of all channel rates achieved in each of the parallel data channels created by spatial multiplexing. The maximum achievable sum rate of a multi-user MIMO system is referred to as its sum capacity. Sum rate maximization or achieving the sum capacity is of interest when high overall system throughput is desired. Note that achieving sum capacity often results in uneven resource allocation among the users served, that is, some users end up with high channel rates exceeding their requirements while others may be in deficit. An alternative to sum rate maximization is an exercise to meet a certain level of Quality of Service (QoS) at each user. This may be expressed in terms of a minimum data rate, maximum packet delay, etc. The problem of meeting QoS constraints at the minimum transmission power is commonly referred to as the power-control or interference-balancing problem. It must be recognized that meeting each user's QoS needs will likely cause a departure from the best achievable sum rate. Both issues will be addressed in this dissertation. Although more effort is placed on the issue of sum rate maximization, both issues are addressed jointly, for example, a mechanism for meeting the user-QoS requirements while minimizing the sum rate loss will be proposed.

Before giving a survey on the various spatial multiplexing schemes, it is noteworthy that spatial multiplexing could also be used to enhance future broadband systems at the infrastructure level besides serving multiple users at the "last-mile" stage. For example, multi-hop relaying is considered one of the most promising technologies in broadband systems that enables cost-effective enhancement of coverage, user throughput and system capacity [8]. The primary reason is that many broadband radio interfaces for next-generation mobile networks such as 3G LTE (Long Term Evolution) and mobile WiMAX 802.16e and beyond will be characterized by very limited range due to the push for higher data rates over higher transmission frequencies, which have higher propagation channel losses. The high data rate demand coupled with high channel losses translates to lower energy in every bit of information sent, which poses more challenges for reliable data recovery. For example, 3G systems are expected above the 2 GHz band while WiMAX 802.16e is destined to become the time-division duplex technology for the 2.5 and 3.5 GHz bands. To get a feel of the impact, good quality indoor coverage in a suburban setting would take nearly four times as many sites to deploy at 2 GHz than at 1 GHz, and 10 times as many for 3.5 GHz [8]. To ensure that the quality of end-user experience is not dependent on terminal location within a cell, a high density of base stations will be needed, which will drive up the cost of deployment. To address this problem, the use of multi-hop relays appears to be a viable solution [9]. In particular, wireless relay nodes do not need wired-backbone access and help reduce capital expenditures. They facilitate faster network rollout and enable adaptive traffic capacity engineering. They are also well suited for ad hoc deployments such as emergency and disaster scenarios as well as military settings. Since relays normally work in the half-duplex mode, the burst rate in each half will be high if a high overall system throughput is desired. For example, the burst rate in each relay hop is roughly double that of the overall system throughput for a two-hop relay system. The problem is compounded if a base station has to communicate with several relay nodes in different area sectors. Serving these relay nodes via time-division multiplexing coupled with half-duplex operation will result in very high burst rates that may exceed individual channel capacities. Toward this end, spatial multiplexing using MIMO technologies in the relay downlink between a base station and its neighboring wireless relay nodes is particularly attractive since reliance on time-division multiplexing can be avoided or at least reduced. Further cost reduction may be realized from the fact that MIMO systems operate best in a scatter-rich environment. This means that the relay nodes need not be located on high tower structures to get good line-of-sight with a base station. Additionally, there are two other reasons that make fixed wireless relays favorably disposed for the application of MIMO techniques as compared to mobile nodes. First, spatial multiplexing relies on the availability of accurate and timely channel state information. It is easier to meet this requirement with relay nodes because they are usually static when deployed. Second, the form factor of relay terminals is not normally constrained by space and power as compared to mobile terminals, which are constrained by ergonomics. This means for example that the possibility of incorporating more antennas and RF chains at a relay terminal is much higher than for a mobile terminal.

1.1 A Brief Survey of Spatial Multiplexing Schemes

Spatial multiplexing methods for multi-user MIMO downlinks may broadly be classified under two categories, namely, the coding approach to avoid inter-user interference or the signal processing approach. The methods may be further classified according to a number of criteria like (a) whether they attempt to approach the sum capacity bound, (b) the extent of inter-user interference elimination, (c) the number of user terminal antennas, (d) whether they utilize the user terminal processing capabilities, (e) whether they achieve the minimum QoS requirements, and (f) the number of data streams transmitted to each user [10]. As an example, a spatial multiplexing scheme that maximizes the sum capacity may not always lead to a desirable solution as it favors users with higher SNR and leaves weaker users with little or no throughput.

Information theory literature considers multi-user downlinks where transmissions originate from a base station as a "broadcast channel" (BC). It is referred to as a "vector broadcast channel" or as a MIMO broadcast channel when the MIMO structure is included. When the transmitter and receivers have full channel state information, the MIMO broadcast channel is considered as non-degraded. It has been proven in [11] – [14] that the maximum sum rate, that is, the sum capacity of non-degraded MIMO broadcast channels is achieved by a coding technique proposed by Costa called "writing on dirty paper" [15]. This is commonly referred to as dirty-paper coding (DPC) and a tutorial exposition is available in [16] from which the basic idea is outlined here. Suppose a signal S is to be sent in the presence of interference I and noise W. If the interference I is known beforehand to the transmitter that is sending S, Costa presented the surprising result that the capacity of this system is the same as if there were no interference present. This concept of writing on dirty paper implies designing a code that avoids the known interference I. In the context of multi-user downlinks, the base station has knowledge of the signals to be sent to each user. If the channel to each user is also known at the base station, the interference arising from the non-desired signals arriving at each user will then be known and dirty-paper methods can then be used to avoid mutual interference. It has been shown that the sum capacity C_{sum} as well as other combinations of individual channel rates are achievable, that is, all points of the rate region for different QoS are achievable using the dirty-paper coding approach [17]. For the case where all K user terminals have only one antenna, the sum-capacity expression is [12]

$$C_{\text{sum}} = \sup_{\mathbf{D} \in \mathbf{A}} \log_2 \det\left(\mathbf{I} + \mathbf{H}\mathbf{D}\mathbf{H}^H\right) \quad \text{bits/sec/Hz}, \tag{1.3}$$

where **A** is the set of all $K \times K$ non-negative diagonal matrices with $\text{trace}(\mathbf{D}) \leq P$, where P is the total transmitted power. The multi-user sum capacity expression in (1.3) is similar to the single-user capacity expression in (1.1) and hence it is easy to see that the expected sum capacity $\bar{C}_{\text{sum}}$ also scales linearly with $\min(M, K)$ under the same conditions as for $\bar{C}_{\text{SU}}$. When the number of users K is very large, it has been shown that the asymptotic expected sum capacity of a MIMO-BC system using DPC scales with the number of transmit antennas M as

$$M \log \log KN, \tag{1.4}$$

where N is the number of receive antennas per user [18].

Unfortunately, dirty-paper coding involves complex nonlinear designs that incur heavy computational load, which may present difficulties for real-time implementation at least in the near term. DPC is sometimes referred to as interference-dependent coding because the transmitted code is a nonlinear function of the information symbols and the interference environment. As such, DPC requires new code designs, which makes it incompatible with current communication standards and protocols and complicates its adoption. There are various works that seek to achieve simplified DPC techniques and one such example referred to as vector perturbation is found in [19].

There is a lower-complexity alternative to the coding approach for achieving spatial multiplexing in multi-user MIMO systems and this is the signal processing approach for which there is extensive industry experience and support. In this non-coding approach, the signal for an intended user is treated as noise when it arrives as interference to other users. A multi-user system that adopts this approach is referred to as a degraded-broadcast channel in information theory literature. In particular, linear processing techniques like transmit beamforming and receive beamforming can achieve spatial multiplexing with reduced complexity. In contrast to DPC, beamforming involves choosing appropriate transmit- and receive vectors and the process is independent of the signaling and coding protocols used. Hence, the integration of such linear processing techniques into current systems is much less complex than DPC. One simple linear transmit beamforming technique is zero-forcing beamforming (ZFBF), which is widely considered due to its relative simplicity. The transmit processor is chosen such that all inter-user interference will be reduced to zero. However, like zero-forcing receivers, which suffer from noise enhancement, ZFBF suffers from transmission power increase when signals arriving at different users are highly correlated. Another class of schemes in the signal processing approach employs non-linear techniques. An example of this is based on the Tomlinson-Harashima pre-coding (THP) technique [20], [21] that was originally developed for the pre-equalization of inter-symbol interference in dispersive channels. Compared to zero-forcing beamforming, transmit processing using THP techniques has better success at limiting the power increase when pre-eliminating the inter-user interference. More details on this approach can be found in literature, for example, [22].

To achieve the maximum sum rate, optimal beamforming requires interference balancing or equivalently, signal-to-interference-plus-noise ratio (SINR) balancing across all active users. Significantly, it has also been shown in [18] that the expected sum rate for optimal beamforming scales as for DPC, that is, with $M \log\log KN$, when the number of users K is very large. Nevertheless, optimal beamforming is still inferior when compared to the DPC approach, which achieves sum capacity for MIMO-BC. Although less complex than DPC, optimal beamforming involves SINR balancing and presents a non-convex problem that still entails high computational complexity [23], [24]. A sub-optimal beamforming method with lower complexity is zero-forcing beamforming, which does away with SINR balancing by enforcing zero co-channel interference (CCI) among the active users. In effect, this results in the creation of orthogonal parallel single-user or point-to-point MIMO channels that are free from mutual interference and coding for each user can be done independently of others.

The lower complexity of zero-forcing beamforming (ZFBF) is accompanied by a performance penalty in that its expected sum rate $\overline{R}_{\text{sum}}^{\text{ZFBF}}$ does not increase linearly with M. This is shown in [25] and more details are given in Chapter 2. However, this setback can be overcome when a large user pool is available from which user selection becomes possible. It is shown in [26] that $\overline{R}_{\text{sum}}^{\text{ZFBF}}$, the expected sum rate for ZFBF with *single*-antenna user terminals ($N = 1$) approaches that of DPC in the limit of large K and when judicious user selection is applied. This means that $\bar{R}_{\text{sum}}^{\text{ZFBF}} \approx \bar{C}_{\text{sum}}^{\text{DPC}}$ when $K \to \infty$ and the average ZFBF sum rate scales linearly with *M*, and from (1.4), it scales as $M \log\log K$. This is due to the existence of multi-user diversity, and the channel gains of the best users are roughly $\log K$ times the average channel gain, as stated in (1.4). When CSI is available at the transmitter, it can choose a group of users with high channel gains whose channel directions are closely matched to the zero-forcing beam directions.

Motivated by the possibility for ZFBF to scale like (1.4) and by its low complexity, our focus is on finding solutions to address the shortcomings of ZFBF in line with the intent of this dissertation to enhance the feasibility of deploying multi-user MIMO systems. To be commensurate with its low complexity, emphasis on efficiency is given when developing algorithms for the performance enhancement of ZFBF systems. For example, attention is paid to enhancing the performance of low-complexity, user-selection methods when developing algorithms to enable scaling with *M* as closely as possible.

When the users or nodes are equipped with multi-antenna terminals, a class of ZFBF that makes use of block diagonalization has been proposed in [27]–[29]. The motivation behind block diagonalization comes from the recognition that enforcing ZFBF between antennas of the same terminal is sub-optimal because those antennas can coordinate their processing for better performance. It is therefore better to enforce zero co-channel interference between user terminals only and this gives rise to the block diagonalized approach of achieving space-division multiplexing (SDM). It is recognized in [27]–[29] that block diagonalization effectively creates parallel single-user MIMO channels and hence optimal processing techniques that were traditionally proposed for single-user MIMO channels could be used. This includes schemes like layered space-time coding and a beamforming scheme that is based on the singular-value decomposition (SVD) of single-user MIMO channels. The SVD-based beamforming scheme is optimal for single-user MIMO channels [30] and a description of its use in BD systems can be found in [27]. It has also been recognized that in the presence of large user pools, judicious user selection is beneficial for *multi*-antenna terminals served by spatial multiplexing [31]. The benefits of user selection are also recognized for block diagonalized systems, for example, [32] and [33].

In this dissertation, it is shown that sole reliance on user selection is insufficient when maximizing the ZFBF sum rate for block diagonalized systems. Instead, an additional level of selection, commonly known as receive antenna selection (RAS) is also needed. The combination of RAS and user selection helps ZFBF systems with multi-antenna terminals to approach scaling with M faster than dependence on user selection alone. Another important impact of this combination is the possibility of scaling with M with smaller user-pool sizes than when relying on user selection alone. This will help realize performance improvements under realistic operating conditions and avoid a dependence on the existence of large user pools. It is shown in Chapter 5 that the user selection process is in fact subsumed under the RAS process when optimal selection is desired. Note that the RAS and user selection processes become identical for ZFBF systems serving single-antenna terminals.

The RAS concept was originally developed for single-user MIMO links where hardware cost reduction was sought while preserving as much of the performance as possible. Since the cost of antennas is generally lower than the attendant RF-chains, the idea is to cater more antennas than RF-chains. The antennas are spatially spread out and a subset of the best antennas is selected for connection with the RF-chains via RF switches. It is shown in [34] that judicious receive antenna selection done over this configuration will help achieve the diversity order of a fully equipped system, that is, a system where each antenna is served by its own RF-chain. It must be pointed out however that the switching loss incurred between the antennas and the RF chains may be high. This has negative impact on the signal-to-noise ratio and must be accounted for in the link budget. A feasibility study can then be made by weighing the degradation against the potential diversity gain. It is noteworthy however that progress in technologies such as micro-electromechanical systems (MEMS) may help reduce the losses incurred by RF switches and make the RAS concept more viable.

Differing from the conventional approach to receive antenna selection (RAS), it is shown in this dissertation that RAS is necessary for sum rate maximization even when

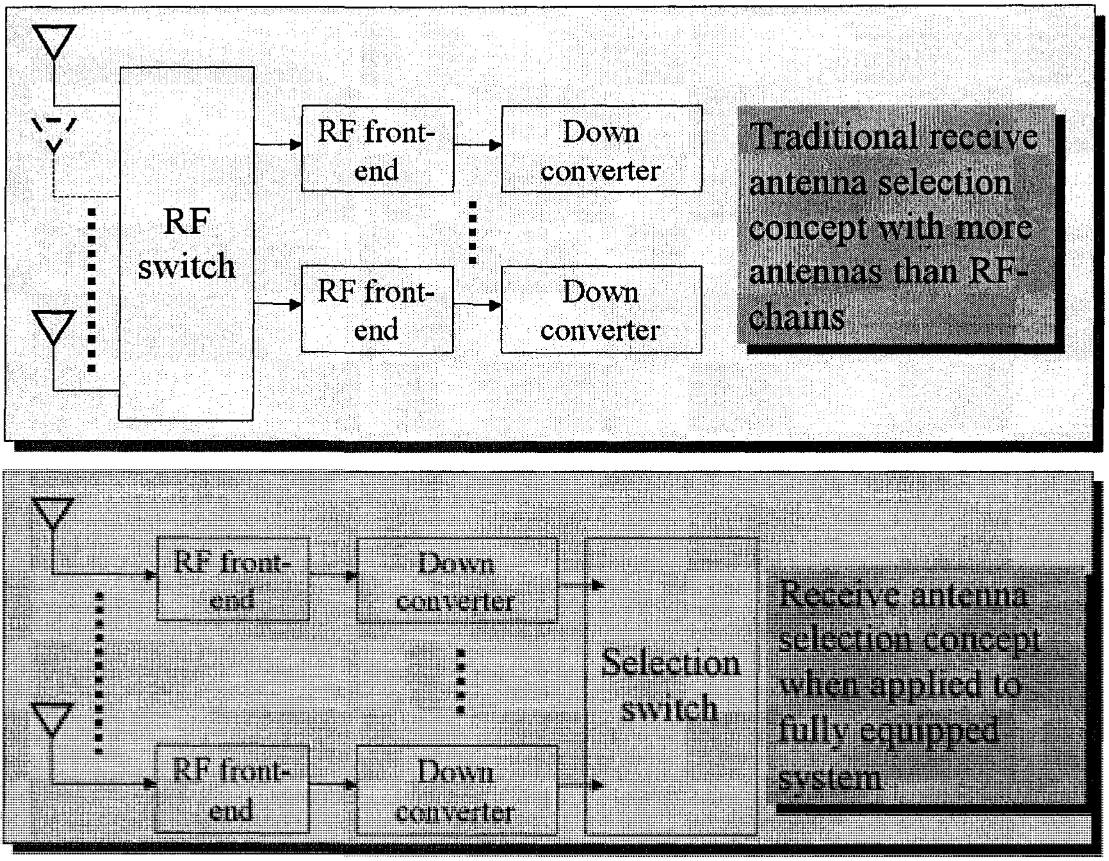

RF switch
RF front-end
Down converter
RF front-end
Down converter
Traditional receive antenna selection concept with more antennas than RF-chains
RF front-end
Down converter
Selection switch
RF front-end
Down converter
Receive antenna selection concept when applied to fully equipped system

Figure 1.3: Comparison between traditional RAS and RAS for fully equipped system

each user terminal is *fully* equipped. This is true despite a drop in the upper bound on the broadcast sum capacity. To highlight some numerical results, significant improvements to the average sum rate of between ~40% to ~50% are obtained when RAS is performed on an 8-user block-diagonalized system, where each user is equipped with 4 antenna-RF chains (note that user selection is not done). It is interesting to note that users with reduced antenna-array sizes may also enjoy channel rate increases. More details on this are given in Section 1.3 and in Chapter 3. It is also important to note that issues associated with switching losses between the antennas and RF chains do *not* exist in this case because each terminal is fully equipped, that is, each antenna is accompanied by an RF-chain. Selection can then be done after down-conversion so that possible impact on the link budget is typically negligible. A comparison between the original RAS scheme for single-user MIMO systems and the RAS scheme proposed for block-diagonalized systems is shown in Figure 1.3. Analysis of the mechanisms behind the impact of RAS, its extent and its consideration together with user selection is given in Chapter 3. For systems that use receive-weight matrices to control the number of data streams, that is, the number of spatial modes, it is shown that spatial mode selection (SMS) is also needed to help maximize the sum rate. This process is analogous to the receive antenna selection (RAS) process and its details are given in Section 1.3.

As noted earlier, the sum rate maximization process often results in uneven resource allocation among the users served, that is, some users end up with high channel rates exceeding their requirements, while others may be in deficit. On the other hand, it must be recognized that meeting each user's QoS needs will likely cause a departure from the best achievable sum rate. Both issues are jointly addressed in this dissertation and algorithms for meeting the user-QoS requirements while minimizing the ZFBF sum rate loss are proposed. A related problem exists for multi-antenna terminals that are capable of receiving more than one data stream. Here, the resource allocation problem entails (a) finding an appropriate number of antennas/modes to be activated at each user terminal to meet each individual demand and (b) solving a combinatorial problem that may arise when subsets of antennas/modes are to be chosen at some users. A challenge emerges here because resource allocation done at any one user will have an impact on all other users. To address these issues, efficient resource allocation algorithms are developed in this thesis that meet individual channel-rate requirements while minimizing the individual- and the overall sum rate losses.

The use of receive antenna selection in block diagonalized systems was previously mentioned in [35], [36] and [37]. An equivalent was proposed in [38], where beam ordering and selection were introduced for BD. However, [36] and [38] did not give detailed explanations of why receive antenna selection (RAS) benefits BD-SDM systems. They did not cover schemes that use receive-weight matrices for spatial mode allocation. The RAS algorithms in [36] and [38] use a single-antenna selection approach that often results in the scheduling of many users, especially when intra-terminal correlation is high. This gives rise to low individual channel rates, which may be insufficient to meet the individual rate demands. Although the algorithm in [36] provides very good sum rate performance, it incurs high complexity due to the iterative use of BD pre-coding with single-antenna selection. A block antenna/mode selection approach is introduced in this thesis to help overcome problems faced by the single-antenna selection approach. In addition, the issue of resource allocation to meet individual rate demands was also not well addressed in [36], [37] and [38].

Another major hindrance to the adoption of spatial multiplexing is the need for channel state information (CSI) at the base station, which can incur an enormous amount of overhead when the user pool is large. Zero-forcing beamforming systems require timely and accurate channel estimates for good performance. The problem is compounded when exploitation of multi-user diversity via judicious user/antenna/mode selection is desired and optimal selection is performed by the base station based on the channel matrices of all users under consideration. This has motivated much research effort to find ways of reducing the feedback overhead. In general, selection algorithms with better sum rate performance still require the full channel matrix of each user under consideration at the base station. There are two broad approaches to mitigate the overhead-reduction problem associated with CSI feedback to the base station, namely, (a) limited-bandwidth CSI feedback and (b) partial CSI feedback.

This dissertation focuses on the latter case of partial CSI feedback during the user/antenna/mode selection process and during the beamforming process. An example where a high degree of CSI feedback reduction is achieved during user selection is found in [39] where the authors propose an orthogonal random beamforming (RBF) scheme. The RBF scheme achieves the optimal DPC sum rate asymptotically when the user pool is very large. This is possible due to existence of multi-user diversity that enables the matching of users even with randomly chosen beam directions. However, RBF has slow convergence in K, the user-pool size and requires the presence of very large user-pool sizes to be effective. Results in [26] show poor performance for the RBF scheme for practical values of K, for example, $K < 100$, where a base station with four transmit antennas is serving single-antenna terminals. Hence, focus is given in this dissertation to ZFBF systems and the methods for achieving partial CSI feedback.

A straightforward method for reducing CSI feedback during user selection in ZFBF systems is to base the selection metric on the channel gain of each user. The reduction is due to fact that each channel gain value may be transmitted as a scalar to the base station, which is much less than transmitting the full channel matrix of each user. The base station chooses users with the best gains and the full channel matrix data are required only from these users. However, channel-gain based user selection often results in poor ZFBF performance because the chosen channel directions may not line up well with the zero-forcing directions. The impact of antenna/mode selection on ZFBF systems allows the development of a method to mitigate this situation and significant performance improvement is obtained with a low additional complexity and overhead. Results in Chapter 5 show significant reduction in the performance gap between channel-gain based user selection and more complex algorithms. This makes the deployment of channel-gain based user selection more feasible in practice.

A variant of the channel-gain method for CSI feedback reduction has been proposed in [40]. It is a polling method that picks the next user who has the highest projection magnitudes in the null space of a currently chosen user group. User selection decisions are again made on a single scalar feedback from each user while full CSI feedback is required only from the chosen users, thus contributing to feedback reduction. Again, the method proposed for channel-gain based user selection may be used here for sum rate improvement.

In relation to CSI feedback reduction during the ZFBF beamforming process, the analysis and results in Chapter 3 will show a possible method that is based on localized antenna/mode selection done at each user terminal, without the involvement of the base station. Localized antenna/mode selection can contribute to better sum rates but is sub-optimal compared to coordinated selection done by the base station. However this approach may be considered if CSI feedback reduction for the purpose of zero-forcing beamforming is of paramount concern.

Taken together, the practical feasibility of deploying ZFBF systems is raised by proposing an efficient streamlined process that integrates the sum rate maximization (via RAS and user selection) and the resource allocation processes to meet individual QoS needs while minimizing individual- and sum rate losses, along with reduction in the CSI feedback requirement during user selection or beamforming.

In the following sections, an overview of zero-forcing beamforming (ZFBF) methods for multi-user downlinks and the related issues is given. To emphasize the point that the zero-forcing operation is accomplished at the base station, the term "transmit zero-forcing beamforming" or TZFBF is used. Broadly, TZFBF methods may be grouped under two categories, namely, those for single-antenna terminals and those for multi-antenna terminals. An outline of the research effort and contributions made is then given against this background.

1.2 TZFBF for Single-Antenna Terminals

For single-antenna terminals, transmit zero-forcing beamforming (TZFBF) is easily implemented by a process known as channel inversion using a linear algebra operation known as the pseudo-inverse. In this case, the pseudo-inverse of the channel matrix is taken [41], [42] and is used to pre-code the data streams before transmission. The system may then be referred to as "transmit channel-inversion beamforming" or TCIBF. Transmit channel-inversion beamforming creates parallel channels that are orthogonal to each other, that is, the co-channel interference among them is forced to zero. When the channel is scatter-rich, a maximum of M such channels could be created to serve M users, where M is the number of base station transmit antennas. Depending again on the channel conditions, each user channel will experience a certain channel gain. In general high channel gains will result when the users are located in such a way that the multipath signals arriving at their terminals are uncorrelated with each other. Conversely, the channel gains will be low when the signals arriving at different users are correlated. This latter case tends to occur when the users are not sufficiently dispersed over a geographic area. Low channel gain requires more transmission power to maintain an adequate throughput to the affected user.

In practice however, the radio frequency (RF) amplification stage of any transmission system has a maximum rating, that is, typical systems are power constrained. Faced with a power constraint, an optimal way of dividing this power among a given set of parallel channels with different gains to result in the best sum rate may be done using a well known method called *waterfilling* [43]. Waterfilling across the orthogonalized channels can be done since CSI is available at the transmitter from which the channel gains can be computed after beamforming. It is also known however that using waterfilling alone does not achieve the best possible sum rate and that serving a subset of $< M$ users may result in better sum rates [11]. The mechanisms underlying this behavior will be dealt with at length in Chapter 3. Optimal selection of this subset requires an exhaustive search to find the one that results in the highest sum rate. The exhaustive search involves two processes, first choosing a candidate user subset and then performing a TCIBF rate evaluation for that subset. Each rate evaluation entails a channel inversion exercise since the TCIBF channel gains are needed during the rate evaluation. Since the optimal subset may be anywhere from one to M users (that is, TCIBF does not scale linearly with M), the exhaustive search will require a total of $\sum_{i=1}^{M}\binom{M}{i}$ rate evaluations, which has an exponential complexity order $\in O\left(2^{M}\right)$. This complexity grows rapidly with M and presents an unacceptable burden when implementing large systems.

To alleviate this, an algorithm is developed in this thesis to perform subset selection using only a maximum of M rate evaluations. The algorithm is arrived at after analyzing the underlying causes for poor channel gains in TCIBF. The algorithm is referred to as Joint Rate Evaluation and User Selection (JREUS) since it avoids the typical arrangement that entails *separate* user-subset selection and rate evaluation processes. Significantly, the user selection function in JREUS introduces negligible additional complexity to the original TCIBF rate evaluation process. It exhibits near-optimal performance over a wide range of channel and SNR conditions, and outperforms existing algorithms in [26], [44] – [50] under practical conditions. Alternatives with slightly better performance than JREUS are also derived by analyzing the factors that cause sub-optimal performance in JREUS. Further complexity reduction is realized via a recursive-inverse algorithm that avoids the need to perform each channel inversion afresh as the candidate users are considered in turn.

1.2.1 User Selection for TCIBF

The potential pool of users in a multi-user MIMO system normally exceeds the ability of a base station to support them simultaneously. Let us assume S is the potential user pool with K users, where $K \gg M$. A subset $S_r \subset S$ of at most M users must be chosen to meet the base station's constraint when TCIBF is used. Judicious user selection rather than random selection can help TCIBF to approach the DPC sum capacity asymptotically when the number of users is high [26], [51]. This is due to the existence of multi-user diversity [52]–[53], which may be exploited via judicious user selection. In other words, a large user pool presents the transmitter with a higher chance of choosing a group of users with high channel gains whose channel directions are matched to the zero-forcing beam directions [26]. This lowers the likelihood of signal attenuation due to poor channel gains in TCIBF, which occurs when channel inversion is performed on a chosen user subset whose associated channel is poorly conditioned.

To maximize the TCIBF sum rate, the optimal active user subset $S_r \subset S$ of size $K_r \leq M$ must again be found via an exhaustive search involving $\sum_{i=1}^{K_r=M}\binom{K}{i}$ rate evaluation steps. Clearly, the exhaustive search complexity becomes impractical when the user pool is large and this has attracted much effort to develop efficient user selection algorithms using various approaches, for example, [26], [44] – [50]. These algorithms employ methods like orthogonal complement projection, "greedy" TCIBF pre-coding, pair-wise metrics based on correlation, cosine and squared normalized inner products, channel gains and combinations thereof. In accordance with [18], a well designed user selection algorithm should achieve TCIBF sum rates that scale with $M \log\log K$.

1.2.1.1 Reducing CSI Feedback Requirement During User Selection

Since user selection is done over the entire pool of K users, all schemes in [26], [44] – [50] require the full channel state information (CSI) of all users at the base station, with the exception of schemes that rely on channel gains as the selection metric. This incurs very significant overhead when K is large and presents a major hindrance not only to the adoption of TCIBF, but also to the implementation of spatial multiplexing using MIMO systems in general. This is an interesting research topic and various approaches have been proposed to address this issue. To re-iterate, the focus of this thesis is on partial CSI feedback schemes.

In general, channel-gain based selection is attractive because selection can be done using only a scalar value from each user. A gain threshold may also be enforced to further reduce the amount of feedback to the base station. Full CSI is then obtained from the chosen users to implement the desired spatial multiplexing scheme, for example, TCIBF. However, channel-gain based selection has poor sum rate performance because it does not take the users' channel directions into account. This also results in a lower probability of scheduling M users even when K is large. As such, it fails to achieve sum rates that scale with $M \log\log K$ as given in [18]. This is also true in general for schemes that use pair-wise decision metrics. To address this, a scheme that strives to achieve scaling as (1.4) is developed. By incorporating JREUS in tandem with the algorithms in [26], [44] – [49], their performances are improved by providing additional opportunities for exploiting multi-user diversity. For convenience, the scheme is referred to as "simultaneous scheduling and sum rate maximization" or SSRM for short. Numerical results show that the poorer performing algorithms achieve significant sum rate improvements that are accompanied by higher probabilities of scheduling M users. Hence the SSRM scheme helps these low performing algorithms to approach a performance that scales with $M \log\log K$ [18]. For example, the expected sum rate performance of channel-gain based selection is improved from within about 24% to within about 9% of the best algorithm found in [50]. The probability of scheduling M users is simultaneously raised from a low percentage to almost 100% when K is reasonably large. This improves the feasibility of employing channel-gain based selection in practice, which is attractive given its lower CSI feedback demand and lower computational complexity.

For those spatial multiplexing schemes that can scale with M even under poor channel conditions, the SSRM approach of using the JREUS algorithm in tandem with existing user selection algorithms can help such schemes to scale closing with $\log K$ when $K > M$ than sole reliance on existing user-selection schemes alone. One example of such a scheme is regularized channel inversion, which possesses linear sum rate growth with M [25]. Regularized channel inversion, which is equivalent to using a minimum mean-squared error (MMSE) criterion to design the beamformer weights, helps address the poor performance of TCIBF under poor channel conditions, for example, when the channel is rank deficient due to users in close geographical proximity or due to a scatter-poor environment. Despite this, SSRM can be used to help pick better users from the potential user pool to help closer scaling with $\log K$.

1.2.1.2 Another Class of User Selection Algorithms

Next, [54] shows that the multi-user broadcast sum capacity is upper-bounded by the equivalent single-user, point-to-point MIMO capacity where all receiver antennas can cooperate. This motivates the evaluation of receive antenna selection (RAS) algorithms developed for single-user MIMO channels for the purpose of user selection in the multi-user setting [34], [55] – [56]. This approach is developed in Chapter 4 of this thesis. Briefly, RAS algorithms are developed to improve the performance of point-to-point MIMO systems in terms of capacity and diversity. The key idea is to provide more receive antennas than RF chains and selecting the best subset of antennas to be used. This is similar to the multi-user case where choosing a user subset is akin to choosing a subset of receive antennas. Specifically, it is found that the incremental antenna selection algorithm in [55], which is based on the maximization of point-to-point MIMO capacity, has lower implementation complexity and a performance that is close to the best performing user selection scheme in [50]. This opens up a different class of user selection algorithms that are based on RAS algorithms.

1.2.2 Resource Allocation in TCIBF

Regarding resource allocation versus the QoS requirements of each user, the sum rate maximization algorithms proposed in this thesis provide a systematic basis for allocation when coupled with power allocation methods. This is because algorithms like JREUS involve user ranking, which can then be used as a basis for resource allocation. Very importantly, these algorithms will help resource allocation methods to minimize the sum rate loss during allocation. The details are presented in Chapter 4.

1.2.3 Impact of Transmit Antenna Selection in TCIBF

Next, it is known that transmit antenna selection (TAS) methods provide diversity benefits through the provision of more transmit antennas, beyond the required M transmit-chains. This is applicable to both single-user as well as multi-user MIMO systems. It is also clear from [18] that TAS is not useful for fully equipped systems where all transmit antennas are accompanied by an RF chain. This is because the multi-user sum rates of optimal DPC and optimal beamforming scale as $M \log\log KN$ and reducing M will reduce the sum rate. Despite this, it is shown that transmit antenna selection (TAS) on a *fully* equipped system operating in a *full*-rank channel provides a means of increasing the sum rate in some cases when *sub*-optimal user selection (USEL) algorithms are used. The mechanism works by assisting the USEL search path to get out of a local maximum. The proposed method requires further USEL to follow any prior TAS process and the restoration of any transmit antennas that were removed. An analysis is provided to give insight into the proposed method. The analysis and scheme are applicable to *any* sub-optimal USEL algorithm and guidelines on decoupled search strategies are given. The analysis also affirms the statement that given channel state information at the transmitter, TAS alone does *not* help improve the sum rate of TZFBF, regardless of the channel condition, signal-to-noise ratio and USEL method employed. This means that joint exhaustive USEL-TAS searches are *not* needed to achieve the optimal sum rate and instead, *only* exhaustive USEL is needed. More details on this topic are presented in Chapter 4.

1.3 TZFBF for Multi-Antenna Terminals

When each user terminal has multiple antennas, creating parallel channels with zero co-channel interference at the same terminal is sub-optimal since each terminal is able to

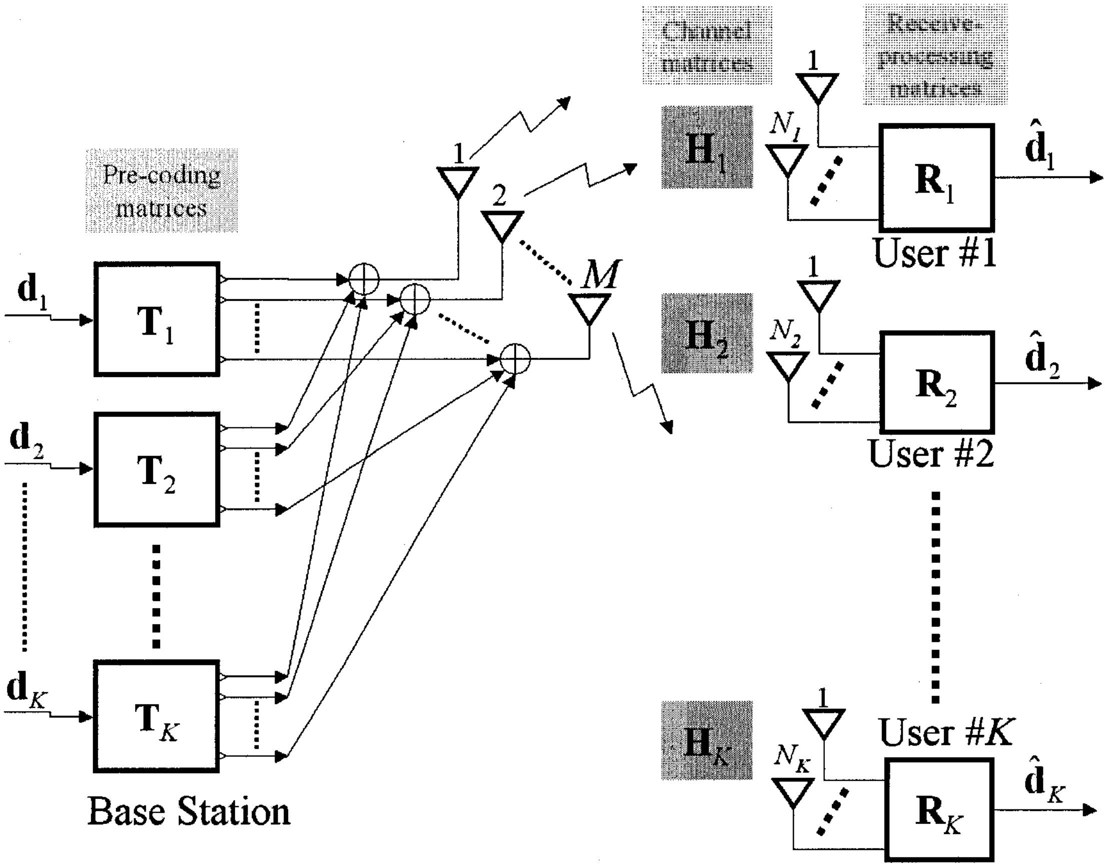


Figure 1.4: Block diagram of typical block diagonalized system

coordinate the processing of its receivers. Zero-forcing between antennas of the same terminal presents constraints to the solution set and it is therefore better to impose orthogonality between user terminals only. In this way, better techniques such as layered space-time coding can then be used for the task of separating different data streams assigned to each user that result in higher sum rates. Alternatively, singular value decomposition (SVD)-based beamforming with waterfilling can be used when coordination between the base station and users is possible. Enforcing zero co-channel interference between users is commonly referred to as block diagonalization (BD) and examples of BD schemes are found in [27], [28] and [29].

Figure 1.4 shows the general block diagram of a space-division multiplexing (SDM) system that employs block diagonalization. As shown, the typical BD system makes use of transmit- and receive- matrices, viz., $\mathbf{T}_j$ and $\mathbf{R}_j$, for each user *j*. In this dissertation, they will be referred to as transmit/receive-weight matrices or as pre-coding/decoding matrices, respectively. The simplest BD system does not make use of receive-weight matrices and outputs of the antenna RF-chains are used directly for receiver processing. This configuration is referred to as direct-BD for convenience. An example that enables direct-BD is when layered space-time coding is used so that layer- or data-stream separation can be done at each user terminal without the use of receive-weight matrices. In contrast, the SVD-based beamforming scheme mentioned above requires receive-weight matrices to enable access to each spatial mode. Since the block diagonalization process yields parallel single-user MIMO channels, the use of SVD-based beamforming with each user is the optimal solution for achieving the best sum rate, as shown in [27].

1.3.1 Spatial Mode Allocation in BD Systems

For direct-BD, each data stream, that is, spatial mode allocated to a user, requires a corresponding antenna-RF chain at that user's terminal. Along with power control, a dynamic spatial mode allocation strategy may be implemented in accordance with each user's QoS requirement. This means that users with low QoS demands do not need to activate all antenna-RF chains and vice versa. Given limited transmission resources, this strategy enables the scheduling of more users compared to a regime where selection is done only at the user level, that is, all antennas of each chosen user are activated.

However, this QoS-dependent strategy gives rise to a resource allocation problem that entails decisions on (a) the number of antenna-RF chains needed at each user and (b) the specific combination of antennas for activation at each terminal to help ensure high throughput in direct-BD. To illustrate the latter point, suppose a terminal equipped with 4 antenna-RF chains has been allocated 2 data streams or spatial modes. A decision is then needed for the choice of 2 antennas out of $\binom{4}{2}=6$ possible combinations. These two decisions cannot be made in isolation at each terminal since a choice made at one user impacts the rate of other users in the BD context. Such allocations must therefore be made at the base station where it can be done with the aim of minimizing rate losses at the individual- and the overall sum rate levels. Note that any sum rate maximization regime may result in some users having inadequate channel rates and others having excess rates. On the other hand, resource allocation exercises to meet the rate requirement of each user usually cause a departure from the maximum sum rate. The challenge is therefore in finding ways of implementing resource allocation while minimizing rate losses.

The mechanism for spatial mode activation in direct-BD is sub-optimal because the unused antenna-RF chains could contribute to better diversity performance. To address this, schemes such as the Coordinated Transmit-Receive (CTR) [27] and the iterative null space directed SVD (Nu-SVD) [29] use appropriately dimensioned receive weight matrices that reflect the number of spatial modes to be activated at each user terminal. In this way, no receive antenna-RF chains are dropped during mode allocation and better performance results because diversity is preserved. Block diagonalization is performed on projected virtual channels, which are made of each user's channel matrix combined with its associated receive-weight matrix. We will refer to this method as virtual-channel BD. However, a similar resource allocation problem remains and entails decisions on (a) the number of modes needed at each user and (b) the specific choice of modes for activation at each terminal to help ensure high throughput in virtual-channel BD. Note that all references made to "antennas" for direct-BD systems in this dissertation are also applicable to the "modes" in virtual channel BD systems. The contributions made in this thesis towards this issue are listed in Section 1.4.

1.3.2 User Selection in BD Systems

When the user pool is large, BD systems can enjoy better sum rates by exploiting multi-user diversity via judicious user selection (see Section 1.1). Optimal user selection for BD sum rate maximization requires an exhaustive search to find the best user subset. Consider a base station with M antennas serving S, a pool of K terminals, each with an arbitrary number of N_j antennas. Let K_r be the number of active users allowed in a direct-BD and it is determined by the following pre-coding dimensionality constraint

$$M - \sum_{i=1,i\neq j}^{K_r} N_i > 0, \quad \forall j. \tag{1.5}$$

For virtual-channel BD, (1.5) is also applicable where N_j is replaced by m_j, the number of modes activated at each terminal.

Finding the optimal active subset of K_r users that satisfies (1.5) using exhaustive search becomes impractical for large user pools and this motivates the development of user-selection algorithms. Most, such as those in [26], [44] – [50], were proposed for single-antenna terminals. There is relatively little work done for multi-antenna USEL (examples are [33] and [57]). The incremental selection algorithm in [33] performs BD pre-coding on each potential user by considering each in turn against the currently selected subset and selects the user that contributes the largest sum rate gain. This is computationally heavy since the evaluation of each candidate user requires a BD pre-coding exercise, which involves singular-value decomposition (SVD) or its equivalent to finding the projection null spaces. It is actually the multi-antenna terminal version of the algorithm in [50]. The algorithms proposed in [57] are computationally less complex by making their decisions based on pair-wise metrics such as angles and correlations between channel row vectors. The contributions made in this thesis towards this issue are discussed in the next section.

1.3.3 Antenna/Mode Selection in BD Systems

Antenna selection has been proposed to save cost for single-user MIMO systems where a limited number of analog RF chains are adaptively switched to a subset of available antennas (see [34] and references therein). This approach is attractive because it retains the diversity benefits of a system that has a high spatial degree of freedom [34] without the need to match every antenna with an analog RF chain. Although the full system capacity is not achieved, identifying the best subset of antennas for each channel realization helps in attaining a large fraction of capacity. Optimal antenna subset selection

Table 1.1: Sum rate improvement due to RAS/SMS – A snapshot

User	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**Total**
Direct-BD without RAS									
#Ants	4	4	4	4	4	4	4	4	
Rate	12.2	12.5	11.7	10.5	12.4	10.8	12.7	11.0	93.8
Direct-BD with RAS									
#Ants	3	2	3	4	2	3	4	3	
Rate	16.8	11.4	16.9	22.7	12.0	16.4	23.9	17.2	**137.4**
Nu-SVD without SMS									
#Modes	4	4	4	4	4	4	4	4	
Rate	12.2	12.5	11.7	10.5	12.4	10.8	12.7	11.0	93.8
Nu-SVD with SMS									
#Modes	3	2	3	3	3	3	4	4	
Rate	17.6	11.9	18.2	17.4	17.4	16.4	23.9	23.1	**145.9**

requires an exhaustive search and this has motivated the development of many selection algorithms to help lower computational complexity, for example, [34], [55] and [56]. In [58], this concept of antenna selection is extended to a multi-user MIMO downlink setting where each user terminal is equipped with more receive antennas than analog chains. The authors in [58] proposed antenna selection algorithms with the same objective of diversity gain enhancement.

In this dissertation, it is shown analytically and numerically that receive antenna selection (RAS) is a necessary part of maximizing the achievable sum rate for multi-user MIMO wireless downlinks that use block diagonalized space-division multiplexing. This is true even though RAS reduces the broadcast sum capacity when all user terminals are fully equipped, that is, all receive antennas are equipped with analog RF chains. The need for RAS holds true even when optimal processing such as SVD-based beamforming with waterfilling is used in each of the parallel single-user MIMO channels created via block diagonalization for a fully equipped system. When virtual-channel BD schemes such as CTR [27] and Nu-SVD [29] are used to provide a means of spatial mode allocation while preserving diversity, applying RAS to the projected virtual channels is equivalent to spatial mode selection (SMS).

An example to illustrate the benefits of RAS/SMS is given in Table 1.1 using direct-BD and Nu-SVD. A system serves $K = 8$ users, each equipped with $N_j = 4$ antennas, via spatial multiplexing from a base station that is equipped with $M = \sum_{j=1}^{K_r} N_j = 32$ antennas. The fairly large value of $N_j = 4$ is chosen to better illustrate the effect of RAS/SMS. The user channel characteristics are assumed to be homogeneous and identical, that is, the users are located in such a way that their average signal-to-noise ratios (SNR) and fading characteristics are identical. Specifically, each path between a transmit- and a receive antenna is assumed to experience Rayleigh fading and one particular channel realization is used in this example. Table 1.1 shows the individual and overall sum channel rates (in bits/sec/Hz) with and without RAS/SMS. To avoid exhaustive search, a RAS algorithm known as "Maximum Determinant Ranking" or MDR (details given in a later section) is used. As shown in Table 1.1, RAS/SMS has substantial impact on the system sum rates with improvements of ~46% and ~56% for direct-BD and Nu-SVD, respectively.

It is interesting to note that in many cases, users with reduced antenna array sizes or reduced spatial mode sets enjoy rate increase as well, for example, users #1 and #3. Note also that the rate loss for users #2 and #5 in the direct-BD scheme is not large despite having 2 antennas removed. The same is also true in Nu-SVD where user #2 has 2 modes removed. The illustration demonstrates the *mutual* benefit when judicious RAS/SMS is performed across the entire system, which improves the sum rates achievable within the BD context. This is true despite sum capacity and individual capacity loss due to RAS/SMS on a fully equipped system. To further highlight some numerical results, a substantial improvement to the averaged sum-rate of ~40% to ~42% is obtained when RAS is performed on the above 8-user direct-BD system, while a range of ~47% to ~53% is obtained when SMS is performed on the same 8-user Nu-SVD system.

The use of RAS in BD systems was previously mentioned in [35], [36] and [37]. An equivalent was proposed in [38], where beam ordering and selection were introduced for BD systems. However, [36] – [38] did not give detailed explanations of why receive antenna selection (RAS) benefits BD systems. They did not cover schemes that use receive weight matrices for spatial mode allocation and the issue of resource allocation to meet individual rate demands was also not well addressed.

Differing from the typical RAS process in single-user MIMO systems, it is necessary in BD systems to discriminate between intra-terminal and inter-terminal antennas, because zero-forcing is done across terminals via BD pre-coding whereas intra-terminal antenna cooperation is possible. In general, this necessitates a selection metric that is based on the rate contribution of each antenna within the BD context. This approach is adopted in the RAS algorithms in [36] – [38] and [58].

Optimal sum rate maximization involves exhaustive search over the entire potential user pool to find the optimal antenna- or spatial-mode subset. Hence, optimal *user* selection for BD sum rate maximization is subsumed within the RAS/SMS process for multi-antenna terminals. Both user- and antenna/mode selection processes become identical for the case of single-antenna terminals. The RAS algorithms in [36] and [38] perform RAS and user selection jointly by considering one candidate antenna at a time. This single-antenna selection approach gives rise to two key problems:

(a) Higher computational loads may be incurred since a decision metric must be generated for each and every antenna under consideration. For example, although the algorithm in [36] provides very good sum rate performance, it incurs high complexity due to the repeated use of BD pre-coding to consider each candidate antenna one at a time.

(b) The single-antenna selection approach often results in the scheduling of many users, especially when intra-terminal correlation is high at many terminals. As transmit resources are spread over many users, it results in low individual channel rates, which may be insufficient to meet the individual rate demands.

To help reduce the complexity mentioned in (a), the following two approaches are considered:

(i) Methods to reduce the number of selection metrics needed, that is, a number that is less than the total number of receive antennas or modes across all users under consideration.

(ii) Alternative selection metrics that require less computational effort, that is, not using the BD sum rate contribution of each antenna or mode as a decision metric.

For point (i), the concept of "block antenna selection (BAS)" and "block mode selection (BMS)" is introduced. Note that the block approach still allows for RAS/SMS together with user selection. In BAS/BMS, selection is done on a subset basis instead of a single-antenna selection (SAS) basis. In this way, the user selection process is also subsumed under a BAS/BMS process.

For point (ii), it is shown that existing RAS algorithms meant for single-user MIMO systems can be modified for BAS/BMS. These are computationally more efficient than [36], [33] and [58] because decision via repeated BD pre-coding is not required. In this way, decremental BAS/BMS, which has potential for better performance than incremental BAS/BMS, is also possible since the BD pre-coding constraint no longer applies. To highlight, a decremental user selection algorithm based on a decremental RAS algorithm from [34] achieves better performance than [33].

The block antenna/mode selection approach will also help address the issue of scheduling too many users as pointed out in (b) above. This is clear since the antennas/modes are chosen as subset blocks from user terminals. In the extreme case, the block sizes may equal the maximum antenna/mode dimensions of each terminal and this corresponds to the usual user selection approach, like the algorithms in [33] and [57]. In this case, RAS/SMS may follow after user selection (USEL) is done, that is, the user-selection and RAS/SMS processes are decoupled.

It is important to note that the RAS/SMS process allows BD systems to scale closer with $M \log\log KN$ under small user-pool sizes than relying on user selection alone. Firstly, the performance gap from the optimal sum capacity level is narrowed when

RAS/SMS is employed in cases where no additional user selection is possible (that is when $KN = M$). When $KN > M$ and user selection is possible, the RAS/SMS process may free transmit resources that allow the consideration of more users. Iterating this way, scaling with $M \log\log KN$ in (1.4) may be achieved with smaller user groups than otherwise possible. This iterative scheme is particularly beneficial for channel-gain based user selection schemes and numerical results show significant improvements. In this way, RAS/SMS contributes not only to better sum rates, it also allows the use of channel-gain based user selection schemes, which have lower processing complexity and significantly lower CSI feedback requirements.

1.3.4 An Integrated Process for BD Systems

It appears at this stage that the various processes associated with the use of block diagonalized space division multiplexing (BD-SDM) are to be considered separately. The main processes are:

(a) Sum rate maximization via receive antenna/mode selection.

(b) Sum rate maximization via user selection.

(c) Resource allocation to meet each user's channel-rate requirement.

(d) Channel state information (CSI) feedback requirement reduction, that is, achieving partial CSI feedback from each user terminal to the base station.

In line with the push for efficiency, effort is made in this thesis to identify the tasks that are common between processes, intermediate results that may be shared or inferences that may be made. An integrated, streamlined process is then proposed. Beginning with the task of sum rate maximization, the methods for joint antenna/mode selection and user selection have been discussed in Section 1.3.3. The RAS/SMS process invariably ranks the antennas or modes under consideration. Such rankings are useful during the resource allocation phase because they help in decisions that involve trading off the next worst antenna or mode in order to benefit the rest. Importantly, this enables a systematic means for resource allocation with rate loss minimization. Integration with two partial CSI feedback schemes will also be discussed, namely those associated with channel-gain based user selection schemes and the one found in [40].

1.4 Key Contributions

It is shown that receive antenna selection (RAS) or spatial mode selection (SMS) is necessary for maximizing the achievable sum rate of MIMO wireless downlinks that use block diagonalized space-division multiplexing (BD-SDM). This allows BD-SDM systems to scale closer with $M \log\log KN$ under small user-pool sizes than relying on user selection alone. This means that the use of RAS/SMS helps BD-SDM systems to better approach the DPC sum capacity rather than the traditional method of relying on user selection alone. RAS/SMS algorithms for which user selection is a subset of, are proposed to help realize this scaling. For systems with single-antenna terminals, the RAS/SMS and user selection processes are identical. For BD systems, RAS/SMS not only contributes to better sum rates, it also allows the use of channel-gain based user selection schemes, which have lower processing complexity and significantly lower CSI feedback requirements.

A detailed analysis on the joint impact of receive antenna selection and user selection upon block diagonalized systems is given. The key contribution is in the form of a novel lower bound on the expected BD sum rate that takes RAS and SNR into account. It provides a means of estimating performance without the need for time consuming Monte Carlo simulations and is easily extended to channel-inversion systems for single-antenna terminals. The approach is extended to provide an upper bound on the expected BD sum rate that takes user selection, RAS and SNR into account. It demonstrates the inter-play of the various mechanisms and provides a means of estimating the upper bound performance that includes (a) the expected BD sum rates versus the number of user subsets considered, (b) the expected number of antennas or modes to be activated, and (c) the SNR level. Again, this upper bound analysis is easily extended to channel-inversion systems for single-antenna terminals.

A systematic method is developed for resource allocation to meet the channel rate requirements at each user while minimizing losses at the individual- and sum rate levels. The rate-loss minimization approach provides a systematic approach to address the impact on all other users when resource allocation is done at any one user. In addition, method does away with the need to make *a priori* decisions on the number of antennas/modes at each terminal. It also solves the combinatorial problem that presents itself when subsets of antennas/modes are to be chosen at some users. A streamlined process that simultaneously reduces CSI feedback requirement while achieving sum rate maximization via RAS/SMS and user selection, and systematic rate-loss minimizing resource allocation is proposed.

Further details on the key contributions are categorized and listed under those for single-antenna terminals (Section 1.4.1) and those for multi-antenna terminals (Section 1.4.2).

1.4.1 Transmit ZFBF for Single-Antenna Terminals

For this case where transmit channel inversion beamforming (TCIBF) is used, the following key contributions are made:

a. An analysis of conditions for sum rate increase during user selection in TCIBF when the maximum number of users is $K = M$, where M is the number of transmit antennas.

b. An efficient, near-optimal user selection algorithm that implements joint rate evaluation and user selection (JREUS). JREUS requires a maximum of only M steps for its decision metric computations compared to $\in O(M^2)$ steps for most existing algorithms or $\in O(2^M)$ steps for an exhaustive search. This is accompanied by an analysis to show the sub-optimality of JREUS and alternative strategies to improve performance.

c. A lower bound on the expected TCIBF sum rate taking the impact of user selection when the user pool is $K \leq M$. The lower bound is parameterized by the different levels of SNR and different user-subset sizes. Estimates of the expected sum rates are given with respect to the number of users scheduled under different levels of SNR. Estimates of the average number of users to be scheduled for sum rate maximization can then be made for different SNR levels. Comparisons with numerical results show that the bound is fairly tight and therefore useful in practice.

d. An upper bound on the expected TCIBF sum rate taking the impact of user selection when the user pool is $K > M$. The upper bound is parameterized by the different levels of SNR, different user-subset sizes and the number of user subsets considered, which reflects the potential user-pool size. The upper bound provide ballpark estimates of the expected sum rates are given with respect to the number of users scheduled under different levels of SNR and indicative user-pool sizes. Estimates of the average number of users to be scheduled for sum rate maximization can then be made for different SNR levels and user-pool sizes.

e. Adaptation and evaluation of receive-antenna selection (RAS) algorithms designed for single-user MIMO systems for use as user-selection algorithms in TCIBF. Results show near-optimal performance and lower complexity than the best user-selection algorithm in [50], which is computationally heavy because of repeated TCIBF pre-coding during selection.

f. In general, user selection algorithms with lower computational complexity or partial CSI feedback perform poorer in terms of the achieved sum rate and the number of scheduled users. As such, these algorithms fail to achieve sum rates that scale with $M \log\log K$. A scheme that strives toward the scheduling of M users is developed to enhance the performance of such algorithms by incorporating JREUS to work in tandem. Significant performance improvements are gained by providing additional opportunities for exploiting multi-user diversity. For convenience, the scheme is referred to as "simultaneous scheduling and sum rate maximization" or SSRM for short.

g. A method for simultaneous proportionally fair scheduling and partial CSI feedback is proposed.

h. The proposed user selection algorithms provide a systematic means for resource allocation when meeting QoS requirements for each user.

i. An analysis to show that using transmit-antenna selection (TAS) alone for TCIBF does not help increase the sum rate regardless of the channel conditions, SNR levels or user-selection (USEL) algorithms used. When sub-optimal USEL algorithms are used however, TAS followed by further USEL may achieve higher sum rates by helping the search path to get out of a local maximum. This is true even for a fully equipped system operating in a full-rank channel. In accordance with the analysis, restoring the transmit antenna removed will always result in higher sum rates. Guidelines are given on how USEL should be conducted together with TAS.

1.4.2 Transmit ZFBF for Multi-Antenna Terminals

For this case where block diagonalized (BD) beamforming is used to implement space-division multiplexing (BD-SDM), the following key contributions are made:

a. A novel lower bound on the expected BD-SDM sum rate that takes antenna/mode selection into account when the user pool does not exceed the pre-coding constraint. The lower bound is parameterized by the different levels of SNR and different number of antennas chosen for each user, which corresponds to spatial mode allocation. Estimates of the expected sum rates are given with respect to the number of antennas/modes scheduled under different levels of SNR. Estimates of the average number of antennas/modes to be scheduled for sum rate maximization can then be made for different SNR levels. Comparisons with numerical results show that the bound is fairly tight and therefore useful in practice.

b. An upper bound on the expected BD-SDM sum rate to demonstrate the joint impact of user selection when the user pool exceeds the pre-coding constraint. The interplay of various mechanisms is demonstrated because the upper bound is parameterized by the different levels of SNR, different numbers of user terminal antennas/modes and the number of user subsets considered, which reflects the potential user-pool size. This upper bound can provide ballpark estimates of the expected sum rates with respect to the number of antennas/modes scheduled under different levels of SNR and indicative user-pool sizes. Estimates of the average number of antennas/modes to be scheduled for sum rate maximization can then be made for different SNR levels and user-pool sizes.

c. Efficient and near-optimal algorithms for RAS/SMS are developed for the case where the total number of receive antennas or spatial modes is within the block diagonalization pre-coding constraint. The algorithms provide spatial channel ranking and can therefore be used to provide a systematic method for resource allocation to meet the individual QoS needs of the scheduled group. This rate-loss minimization approach provides a systematic approach to address the impact on all other users when resource allocation is done at any one user. In addition, method does away with the need to make *a priori* decisions on the number of antennas/modes at each terminal. It also solves the combinatorial problem that presents itself when subsets of antennas/modes are to be chosen at some users.

d. Efficient algorithms for joint user selection and RAS/SMS to maximize the sum rate. To allow joint selection, the concepts of "block antenna selection (BAS)" and "block mode selection (BMS)" are introduced, which account for differences in intra- and inter-terminal processing in block diagonalized systems. A novel approach is based on the modification of existing RAS algorithms is proposed. It has good performance and low complexity, which is realized by avoiding repeated use of BD pre-coding during selection. It allows for decremental selection, which has potential for better performance than incremental selection. An equivalent method for "simultaneous scheduling and sum rate maximization" or SSRM is developed to allow scaling with $M \log\log KN$ (1.4). This method gives significant sum rate improvement for channel-gain based user selection, which have lower processing complexity and significantly lower CSI feedback requirements during user selection.

1.5 Thesis Organization

1.5.1 Outline of Chapter Two

Chapter 2 provides descriptions of the system model, operating conditions and assumptions. It begins with the general formulas for the MIMO broadcast channel sum capacity and proceeds to the sum rate expressions when beamforming is used. It then covers sum rates for the specific case where zero-forcing beamforming is used. Details on the implementation of zero-forcing schemes, that is, block-diagonalized (BD) beamforming for terminals with multiple antennas and the transmit channel inversion beamforming (TCIBF) for terminals with single antennas will be given. Variants of block-diagonalized systems will also be discussed. Key issues relating to resource allocation and channel state information (CSI) feedback will also be outlined.

1.5.2 Outline of Chapter Three

In line with the intention of enhancing the feasibility of fielding multi-user MIMO systems, this chapter starts by highlighting the issues affecting the sum rate of zero-forcing beamforming systems. It describes the mechanisms that affect the sum rates for a given channel and then seeks to find methods for sum-rate improvement. In particular, the impact of antenna selection and user selection and the underlying mechanisms governing their behavior will be studied in detail. Where possible, expressions to quantify their impact will be given on an ergodic basis and on an asymptotic basis, for example, when the user pool becomes very large.

It is known that the sum rate performance of a block diagonalized (BD) system is lower bounded by its equivalently sized transmit channel-inversion beamforming (TCIBF) system. Given this, Chapter 3 begins by analyzing the conditions and mechanisms that contribute to poor performance in transmit channel-inversion beamforming (TCIBF). It then describes the impact of user de-selection on the TCIBF sum rate when the number of users K is limited to $K \leq M$, where M is the number of transmit antennas. Two lower bounds are then derived for the ergodic TCIBF sum rate under a Rayleigh fading channel. The first lower bound captures the effect of user de-selection and SNR upon the expected TCIBF sum rate. This expression allows ergodic sum rate estimations with respect to the number of users scheduled at various levels of SNR. The expression is similar to that in [59] but unlike the approach in [59], it is *not* derived under the assumption of a large system, that is, $K \to \infty$ and $M \to \infty$, with $M / K = \beta$ where $\beta \leq 1$ is a constant. This implies applicability of such lower bounds, including the one in [59], to systems of practical sizes, which is confirmed via numerical results.

A second tighter lower bound is derived next by dropping the use of Jensen's inequality. The effect of user de-selection is captured by the parameters of unordered eigenvalues of Wishart matrices. Numerical results show that this second lower bound provides more accurate estimates than the first. This approach is then extended to formulate an upper bound that demonstrates the joint effect of user selection when $K >> M$ by incorporating methods from order statistics. Finally, an approach is given to demonstrate the scaling of TCIBF with $M \log\log K$. The approach uses techniques from extreme value theory and differs from that in [26].

Turning next to block diagonalized (BD) systems, a novel approach for lower bounding the expected sum rate BD systems operating in Rayleigh fading channels is presented. It jointly captures the effect of receive antenna selection (RAS) (or equivalently, spatial mode selection (SMS)) and different SNR levels. It uses the parameters of unordered eigenvalues of Wishart matrices to capture the effects of RAS/SMS. Estimates given by this lower bound compares favorably with numerical results. Next, an upper bounding approach to capture the joint effects of user selection, RAS/SMS and different SNR levels is outlined. It combines the use of unordered eigenvalues of Wishart matrices together with methods from order statistics. Since block diagonalized (BD) systems perform better than an equivalent TCIBF system, it should approach scaling with $M \log\log K$ at a faster rate than TCIBF systems.

1.5.3 Outline of Chapter Four

Chapter 4 focuses on the development of selection algorithms for zero-forcing beamforming for single-antenna terminals, that is, algorithms for transmit channel-inversion beamforming (TCIBF). It begins with a user selection algorithm for the case when $K \leq M$ and avoids the typical arrangement that entails *separate* user selection and TCIBF rate evaluation processes. It is therefore referred to as Joint Rate Evaluation and User Selection (JREUS) because user selection is made possible during TCIBF rate evaluation. It incurs a maximum of M steps, which is lower than the $\in O(2^M)$ steps needed for exhaustive search and is also lower than many existing algorithms, which $\in O(M^2)$ steps. The factors behind the sub-optimality of JREUS are then examined and alternative algorithms with higher complexity are proposed for better performance.

User selection algorithms for the case when $K > M$ are developed next. A brief survey of existing algorithms is given and a new class of user selection algorithms that is based on receive antenna selection (RAS) algorithms is introduced. Focus is then made to help algorithms with lower complexity in terms of lower computational loads and lower CSI feedback requirement to perform better by helping them to scale closer with $M \log\log K$. Essentially, this requires the use of JREUS in tandem with any user selection algorithm of choice. This helps in dropping users that contribute to poorer sum rates and create opportunities for the scheduling of better users. This process is referred to as "scheduling and sum rate maximization" or SSRM for convenience.

Schemes for the required channel state information (CSI) feedback reduction, scheduling fairness and resource allocation are addressed next. This is followed by an analysis on the impact of transmit antenna selection on the TCIBF sum rate. The numerical results of various schemes are presented at the end of this chapter.

1.5.4 Outline of Chapter Five

Chapter 5 deals mainly with the selection and allocation algorithms for sum rate maximization in block-diagonalized systems. Selection algorithms refer to those for receive antenna selection (RAS), spatial mode selection (SMS) and user selection. For receive antenna selection, a scheme known as "maximum determinant ranking" or MDR is derived based on the JREUS algorithm from Chapter 4. For spatial mode selection, a simple but near-optimal scheme is developed for the Nu-SVD block diagonalization method developed in [29]. It is based on the "poorest spatial mode elimination" approach and is referred to as the PSME algorithm.

The concept of block antenna selection or block spatial mode selection is developed next and the accompanying algorithms are proposed. To highlight, one class of block selection algorithms that is based on existing receive antenna selection algorithms has low complexity while providing good performance. The lower complexity is due mainly to the fact that repeated BD sum rate evaluations are avoided while the good performance results from minimizing the capacity loss of an equivalent single-user system. By combining block selection together with RAS/SMS, a scheme similar to the SSRM (scheduling and sum rate maximization) scheme in Chapter 4 helps BD systems to scale closer to $M \log\log KN$.

Next, the resource allocation problem to satisfy possibly different user QoS requirements is discussed, and an algorithm is proposed to help meet individual channel rate demands while minimizing rate losses at the individual- and sum rate levels. This is followed by a scheme that is proposed to address the issue of CSI feedback reduction. The chapter ends by proposing an overall streamlined process that covers sum rate maximization via user selection and RAS/SMS, resource allocation with rate loss minimization and partial CSI feedback.

1.5.5 Outline of Chapter Six

In this concluding chapter, a summary of the main issues and contributions is given. Some ideas on possible avenues for future work are then discussed.

Chapter 2

System Models and Assumptions

2.1 Multi-user MIMO Downlinks

The focus is on the multi-user MIMO downlink of a base station (BS) serving S, a group of K geographically distributed users via spatial multiplexing that is achieved using linear pre- and post-processing at the transmitter and receivers. The base station has M antennas and transmit-chains while each user j has one or more antennas N_j, each coupled with a receive-RF chain. Unless otherwise specified, any reference to receive antennas in this dissertation will imply an antenna that is equipped with an analog-RF chain. The system is therefore considered fully equipped and the total number of antenna-RF chains at the user receivers is $N = \sum_{j=1}^{K} N_j$. Each user's channel sub-matrix is denoted as $\mathbf{H}_j \in \mathbb{C}^{N_j \mathrm{x} M}$ and the overall channel matrix is denoted as $\mathbf{H} \in \mathbb{C}^{N \times M}$ where $\mathbf{H} = [\mathbf{H}_1^T\, \mathbf{H}_2^T \cdots \mathbf{H}_K^T]^T$. The entries h_{ij} of $\mathbf{H}$ represents the channel complex gain between the j^{th} transmit antenna and the i^{th} user terminal antenna. They are all independent and identically distributed, each with a complex Gaussian distribution. A scatter-rich environment is assumed and consequently, $\mathbf{H}_j$ and $\mathbf{H}$ are of full rank. In addition, it is assumed that the user channels experience i.i.d. blockwise flat fading that is constant over a block and varies independently from block to block. The information-theoretic assumption of infinitely long code-block lengths is applicable assuming each fading block is sufficiently long. More details on the channel model used are given in Section 2.2. The base station caters a transmit vector $\mathbf{s}_j \in \mathbb{C}^{M \mathrm{x} 1}$ for each user j and the received signal vector $\mathbf{y}_j \in \mathbb{C}^{N_j \mathrm{x} 1}$ at user j is given by

$$\mathbf{y}_j = \mathbf{H}_j \sum_{i=1}^{K} \mathbf{s}_i + \mathbf{n}_j, \tag{2.1}$$

where $\mathbf{n}_j \in \mathbb{C}^{N_j \mathrm{x} 1}$ has variance $\mathbb{E}\{\mathbf{n}_j \mathbf{n}_j^H\} = \sigma^2 \mathbf{I}_{N_j}$ for all receivers.

The sum capacity for a system described by (2.1) has been formulated using the dirty-paper coding (DPC) framework for the case of Gaussian noise, e.g., [12] and [60]. In fact, it was shown in [17] that all points of the rate region is achievable using DPC. Dirty paper coding employs a multi-user encoding strategy that is based on interference pre-subtraction [15]. Briefly, the base station (transmitter) first picks a codeword for user #1. This is followed by choosing a codeword for user #2, which is done with full (non-causal) knowledge of the codeword for user #1. Hence, the codeword of user #1 can be pre-subtracted such that user #2 does not have interference arising from user #1's codeword. Similarly, the codeword for user #3 is chosen so that it does not have interference arising from the codewords of user #1 and #2. This process is implemented until all K users are coded. In this way, the achievable channel rate R_j^{DPC} for each user is

$$R_j^{\mathrm{DPC}} = \frac{\log_2 \left| \mathbf{R}_{n_j n_j} + \mathbf{H}_j \left(\sum_{i=1}^{j} \mathbf{R}_{s_i s_i} \right) \mathbf{H}_j^H \right|}{\log_2 \left| \mathbf{R}_{n_j n_j} + \mathbf{H}_j \left(\sum_{i=1}^{j-1} \mathbf{R}_{s_i s_i} \right) \mathbf{H}_j^H \right|}, \qquad (2.2)$$

where $\mathbf{R}_{s_j s_j} = \mathbb{E}\{\mathbf{s}_j \mathbf{s}_j^H\}$ are covariance matrices for each transmitted data vector $\mathbf{s}_i$, $\mathbf{R}_{n_j n_j} = \sigma_n^2 \mathbf{I}_{N_j}$ is the receiver noise covariance and $|\bullet| = \det(\bullet)$. To achieve sum capacity, optimal values for $\mathbf{R}_{s_j s_j}$ must be found and the resulting sum capacity $C_{\mathrm{sum}}^{\mathrm{DPC}}$ is

$$C_{\mathrm{sum}}^{\mathrm{DPC}} = \max_{\left(\mathbf{R}_{s_1 s_1}, \cdots, \mathbf{R}_{s_K s_K}\right) : \sum_{j=1}^{K} \mathrm{tr}\left(\mathbf{R}_{s_j s_j}\right) \le P} \sum_{j=1}^{K} R_j^{\mathrm{DPC}}, \qquad (2.3)$$

where $\sum_{j=1}^{K} \mathrm{tr}\left(\mathbf{R}_{s_j s_j}\right) \le P$ is the constraint on the transmitted power. Combining (2.2) and (2.3), the sum capacity may be written as

$$\begin{gathered} C_{\mathrm{sum}}^{\mathrm{DPC}} = \max_{\left(\mathbf{R}_{s_1 s_1}, \cdots, \mathbf{R}_{s_K s_K}\right)} \sum_{j=1}^{K} \log_2 \left| \mathbf{I} + \left(\sigma_n^2 \mathbf{I}_{N_j} + \sum_{i=1}^{j-1} \mathbf{H}_j \mathbf{R}_{s_i s_i} \mathbf{H}_j^H \right)^{-1} \mathbf{H}_j \mathbf{R}_{s_j s_j} \mathbf{H}_j^H \right|, \\ \text{s.t. } \sum_{j=1}^{K} \mathrm{tr}\left(\mathbf{R}_{s_j s_j}\right) \le P. \end{gathered} \qquad (2.4)$$

However, dirty paper coding involves complex nonlinear designs that incur heavy computational loads, which may present difficulties for real-time implementation at least in the near term. There is a lower-complexity alternative to the coding approach for achieving spatial multiplexing in multi-user MIMO systems and this is the signal processing approach for which there is extensive industry experience and support. In this non-coding approach, the signal for an intended user is treated as noise when it arrives as interference to other users. A multi-user system that adopts this approach is referred to as a degraded-broadcast channel in information theory literature. In particular, linear processing techniques like transmit beamforming and receive beamforming can achieve spatial multiplexing with reduced complexity, which in line with the intent of this dissertation. In contrast to DPC, beamforming involves choosing appropriate transmit and receive vectors and the process is independent of the signaling and coding protocols used. Hence, the integration of such linear processing techniques into current systems is relatively less complicated than DPC.

The following system model delineates a sub-optimal method of spatial multiplexing via *linear* processing at the transmitter and receivers. Specifically, spatial multiplexing is achieved via beamforming using linear pre-processors ($\mathbf{T}_j$) and post-processors ($\mathbf{R}_j$) at the transmitter and receivers respectively. It is assumed that the base station has complete channel knowledge to compute all $\mathbf{T}_j$ and $\mathbf{R}_j$ matrices.

Consider first the case where only pre-processing matrices $\mathbf{T}_j$ are used. For each user *j*, a data vector of arbitrary dimension $\mathbf{d}_j \in \mathbb{C}^{m_j \mathrm{x} 1}$ is pre-coded by $\mathbf{T}_j \in \mathbb{C}^{M \mathrm{x} m_j}$ to result in transmission vectors $\mathbf{s}_j = \mathbf{T}_j\mathbf{d}_j \in \mathbb{C}^{M \mathrm{x} 1}$. The received signal vector $\mathbf{y}_j \in \mathbb{C}^{N_j \mathrm{x} 1}$ at user *j* is given by

$$\mathbf{y}_j = \mathbf{H}_j \sum\nolimits_{i=1}^{K} \mathbf{T}_i \mathbf{d}_i + \mathbf{n}_j. \tag{2.5}$$

All m_j data streams in $\mathbf{d}_j$ are i.i.d. $\sim \mathbb{C}\mathcal{N}(0,\gamma_i)$ with covariance matrix $\mathbf{R}_{d_j d_j} = \mathbb{E}\{\mathbf{d}_j \mathbf{d}_j^H\} \;= \mathrm{diag}(\gamma_1,\cdots,\gamma_{m_j})$, where $\gamma_j = \mathbb{E}\{|d_j|^2\}$. This implies that all data streams are independently coded. To constrain the total transmit power, $\mathrm{tr}(\mathbf{R}_{ss}) \le P$ must again be satisfied, where $\mathbf{R}_{ss} = \mathbb{E}\{\mathbf{s}\mathbf{s}^H\}$. To find the capacity from the viewpoint of a user j, (2.5) is first re-written as

$$\mathbf{y}_j = \underbrace{\mathbf{H}_j\mathbf{T}_j\mathbf{d}_j}_{\text{desired signal}} + \underbrace{\mathbf{H}_j\sum\nolimits_{i=1,i\neq j}^{K}\mathbf{T}_i\mathbf{d}_i}_{\text{interference}} + \underbrace{\mathbf{n}_j}_{\text{noise}}, \tag{2.6}$$

The interference term arising from transmissions to the other $K-1$ users may be assumed to be asymptotically Gaussian distributed. This is reasonable since the differential entropy $H(\mathbf{y}_j)$ is maximized when $\mathbf{y}_j$ is a zero mean, centrally symmetrical, complex Gaussian (ZMCSCG) random variable. This in turn implies that $\mathbf{d}_j$ must also be ZMCSCG. Since the sum of Gaussian random variables is also Gaussian, it is therefore reasonable to assume that the interference term is asymptotically Gaussian distributed. Let

$$\mathbf{z}_j = \mathbf{H}_j\sum\nolimits_{i=1,i\neq j}^{K}\mathbf{T}_i\mathbf{d}_i + \mathbf{n}_j = \mathbf{H}_j\tilde{\mathbf{T}}_j\tilde{\mathbf{d}}_j + \mathbf{n}_j, \tag{2.7}$$

where $\tilde{\mathbf{T}}_j = [\mathbf{T}_1 \cdots \mathbf{T}_{j-1}\ \mathbf{T}_{j+1} \cdots \mathbf{T}_K]$ and $\tilde{\mathbf{d}}_j = [\mathbf{d}_1 \cdots \mathbf{d}_{j-1}\ \mathbf{d}_{j+1} \cdots \mathbf{d}_K]$. Given this, the capacity from the viewpoint of a user j for a fixed channel $\mathbf{H}_j$ is

$$C_j^{\text{BF}} = \max_{\mathbf{R}_{y_j}} \log\frac{\det(\mathbf{R}_{y_jy_j})}{\det(\mathbf{R}_{z_jz_j})}, \tag{2.8}$$

where $\mathbf{R}_{y_jy_j} = \mathbb{E}\{\mathbf{y}_j\mathbf{y}_j^H\}$ and $\mathbf{R}_{z_jz_j} = \mathbb{E}\{\mathbf{z}_j\mathbf{z}_j^H\}$. Hence

$$C_j^{\text{BF}} = \max_{\mathbf{T}_j,\,\mathbf{R}_{d_jd_j}} \log\det\left(\mathbf{I}_{N_j} + \frac{\mathbf{H}_j\mathbf{T}_j\mathbf{R}_{d_jd_j}\mathbf{T}_j^H\mathbf{H}_j^H}{\mathbf{R}_{z_jz_j}}\right). \tag{2.9}$$

The sum rate R_{Sum}^{BF} of the entire system comprising K users is then

$$R_{Sum}^{\text{BF}} = \max_{\substack{\mathbf{T}_j,\,\mathbf{R}_{d_jd_j},\,j=1,\cdots,K \\ \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss})\leq P}} \sum\nolimits_{j=1}^{K} \log\det\left(\mathbf{I}_{N_j} + \frac{\mathbf{H}_j\mathbf{T}_j\mathbf{R}_{d_jd_j}\mathbf{T}_j^H\mathbf{H}_j^H}{\mathbf{R}_{z_jz_j}}\right). \tag{2.10}$$

Finding the optimal sum rate is not an easy task as it involves SINR balancing via a set of $\mathbf{T}_j$ and $\mathbf{R}_{d_jd_j}$ matrices.

When post-processing matrices $\mathbf{R}_j \in \mathbb{C}^{m_j \times N_j}$ are used, estimates of the transmitted data symbols at user j are derived as

$$\hat{\mathbf{d}}_j = \mathbf{R}_j(\mathbf{H}_j \sum_{i=1}^{K} \mathbf{T}_i \mathbf{d}_i + \mathbf{n}_j). \tag{2.11}$$

In this case, the expressions for $\mathbf{z}_j$ and R_{Sum}^{BF} are modified as

$$\mathbf{z}_j = \mathbf{R}_j\left(\mathbf{H}_j \sum_{i=1, i \neq j}^{K} \mathbf{T}_i \mathbf{d}_i + \mathbf{n}_j\right) = \mathbf{R}_j \mathbf{H}_j \tilde{\mathbf{T}}_j \tilde{\mathbf{d}}_j + \mathbf{R}_j \mathbf{n}_j, \tag{2.12}$$

$$R_{Sum}^{\mathrm{BF}} = \max_{\substack{\mathbf{T}_j,\, \mathbf{R}_{d_j d_j},\, j=1,\cdots,K \\ \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss}) \leq P}} \sum_{j=1}^{K} \log\det\left(\mathbf{I}_{N_j} + \frac{\mathbf{R}_j \mathbf{H}_j \mathbf{T}_j \mathbf{R}_{d_j d_j} \mathbf{T}_j^H \mathbf{H}_j^H \mathbf{R}_j^H}{\mathbf{R}_{z_j z_j}} \right). \tag{2.13}$$

Again, finding the optimal sum rate for beamforming is a difficult non-convex optimization problem [23] as it involves SINR balancing via a set of $\mathbf{T}_j$, $\mathbf{R}_j$ and $\mathbf{R}_{d_j d_j}$ matrices. Simpler but sub-optimal schemes may be obtained via the zero-forcing beamforming approach and details are given later in Section 2.3.

2.2 Channel Model Used

In general, channel models may be classified into two broad categories, namely, physical models and analytical models. Physical channel models characterize an environment on an electromagnetic wave propagation basis by describing the multipath propagation between the location of the transmit array and the location of the receive array. In contrast, analytical channel models characterize the transfer function of the channel between the individual transmit and receive antennas in a mathematical/analytical way without explicitly accounting for wave propagation. The individual transfer functions are usually expressed in the form of a MIMO channel matrix. Analytical models are very popular for synthesizing MIMO channel matrices in the context of system and algorithm development and verification [61].

Analytical models can be further subdivided into propagation-motivated models and correlation-based models [61]. The first subclass models the channel matrix via propagation parameters whereas the second subclass characterizes the MIMO channel matrix statistically in terms of the correlations between the matrix entries. A very popular correlation-based model is the spatially independent and identically-distributed (i.i.d.) flat-fading channel, which is commonly referred to as a *scatter-rich* narrowband channel. Other popular correlation-based analytical channel models are the Kronecker model and the Weichselberger model.

In this thesis, the following assumptions are made in relation to the channels and the channel state information available at the base station:

(a) The channels are assumed to be narrowband with Rayleigh flat-fading and therefore completely characterized in terms of their spatial structure [61]. Note that this assumption is valid within the sub-channels of broadband systems that use OFDM schemes.

(b) The channels are quasi-stationary where each user channel experience i.i.d. block-wise flat fading that is constant over a block and varies independently from block to block. This is a so-called "quasi-static channel", a common simplification to facilitate analysis. This models the *slow fading* situation where the delay requirement of the application is short compared to the channel coherence time. Essentially, the channel is random but is constant at least for the duration of a block plus the time delays associated with channel state information (CSI) feedback. This allows channel matrices to be written without the need to account for changes with time.

(c) Together with (b), the definition of ergodic MIMO capacities or sum rates is permissible when (1) the BS transmitter can adapt its transmission strategy in accordance with the instantaneous channel state so as to maximize the instantaneous transmission rate, and (2) the information-theoretic assumption of infinitely long code-block lengths is applicable assuming each block is sufficiently long.

(d) The channels used are strictly non line-of-sight, that is, the line-of-sight component is zero and the *K*-factor is zero.

(e) The Kronecker model is used to capture the presence of spatial correlations. This model is popular given its simplicity arising from separable transmit and receive correlation that allows for independent array optimization. Details on this model are given in the following section.

(f) A single-cell system is assumed in this work and inter-cell interference is not accounted for. However, the results are applicable to cellular systems with network coordination, where inter-cell interference can be mitigated.

(g) The channel state information available at the base station is assumed to be timely and accurate. As mentioned in (b) above, the timeliness must be taken with respect to the channel coherence time. No attempt has been made in this thesis to study the impact of inaccurate, delayed or erroneous CSI.

2.2.1 Further Details on the Channel Model Used

The following description is mainly taken from [61] and [69]. In general, various narrowband analytical models are based on a multi-variate complex Gaussian distribution of the MIMO channel coefficients, that is, Rayleigh or Ricean fading. With the assumption of zero line-of-sight component, we consider only the case with non line-of-sight components characterized by the Gaussian matrix **H**. In the most general form, the zero-mean multi-variate complex Gaussian distribution of $\mathbf{h} = \operatorname{vec}\{\mathbf{H}\}$ is given by

$$f(\mathbf{h}) = \frac{1}{\pi^{nm} \det\{\mathbf{R_H}\}} \exp\left(-\mathbf{h}^H \mathbf{R}_\mathbf{H}^{-1} \mathbf{h}\right), \tag{2.14}$$

where $\mathbf{R_H} = \mathbb{E}\left(\mathbf{h}\mathbf{h}^H\right)$ and $\operatorname{vec}\{\mathbf{A}\} = \left[\mathbf{a}_1^T \cdots \mathbf{a}_m^T\right]^T$ where $\mathbf{A} = \left[\mathbf{a}_1 \cdots \mathbf{a}_m\right]$.

$\mathbf{R}_{\mathbf{H}_w}$ is known as the full correlation matrix and describes the spatial MIMO channel statistics. It contains the correlations of all channel matrix elements. Realizations of MIMO channels with the distribution of (2.14) can be obtained by

$$\mathbf{H} = \text{unvec}\{\mathbf{h}\} \quad \text{with } \mathbf{h} = \mathbf{R}_{\mathbf{H}}^{1/2}\mathbf{h}_w, \tag{2.15}$$

where $\mathbf{R}_{\mathbf{H}}^{1/2}$ denotes an arbitrary matrix square root ($\mathbf{R}_{\mathbf{H}}^{1/2}\mathbf{R}_{\mathbf{H}}^{H1/2} = \mathbf{R}_{\mathbf{H}}$), and $\mathbf{h}_w$ is an $nm\,\text{x}\,1$ vector with i.i.d. Gaussian elements with zero mean and unit variance.

For the classical i.i.d. model, $\mathbf{R}_{\mathbf{H}} = \rho^2\mathbf{I}$, that is, all elements of $\mathbf{H}$ are uncorrelated and hence statistically independent and have equal variance ρ^2. Physically, this corresponds to a spatially white MIMO channel, which occurs only in rich scattering environments. The i.i.d. model is parameterized only by ρ^2 and is often used for theoretical considerations like the information theoretic analysis of MIMO systems

Next, the Kronecker model assumes that spatial transmit and receive correlation are separable, which is equivalent to restricting to correlation matrices that can be written a Kronecker product

$$\mathbf{R}_{\mathbf{H}} = \mathbf{R}_t \otimes \mathbf{R}_r, \tag{2.16}$$

where $\mathbf{R}_{\text{Tx}}$ and $\mathbf{R}_{\text{Rx}}$ are the transmit and receive correlation matrices. They are deterministic, positive-definite Hermitian matrices and are given by

$$\left.\begin{aligned} \mathbf{R}_{\text{Tx}} &= \mathbb{E}\left(\mathbf{H}^H\mathbf{H}\right) \\ \mathbf{R}_{\text{Rx}} &= \mathbb{E}\left(\mathbf{H}\mathbf{H}^H\right) \end{aligned}\right\} \tag{2.17}$$

It can be shown that under the assumption in (2.16), (2.15) then simplifies to the Kronecker model

$$\mathbf{h} = \left(\mathbf{R}_t \otimes \mathbf{R}_r\right)^{1/2}\mathbf{h}_w \;\Leftrightarrow\; \mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}, \tag{2.18}$$

where $\mathbf{H}_w = \text{unvec}\left(\mathbf{h}_w\right)$ is an i.i.d. unit-variance MIMO channel matrix.

It is assumed in this thesis that the base station antennas are well spaced enough to allow $\mathbf{R}_t = \mathbf{I}$ and the users are well separated enough to consider only the intra-terminal antenna correlation. A constant correlation model may be used when the receive antennas are in close spatial proximity. Let r_{ij} be the entries of $\mathbf{R}_r$, then $r_{ii} = 1$ and $r_{ij} = \varphi$ for the constant correlation model where φ is the correlation between any two antennas at each user terminal. For arrays with better linear spacing, an exponential correlation model may be used where each element r_{ij} in $\mathbf{R}_r$ is $r_{ij} = \varphi^{|i-j|}$ where φ is the maximum correlation between any two antennas at each user terminal.

When switches are used for antenna selection or RF-chain selection (after the down-conversion stage) at the receiver, the coupling factor of the switches may also be included in the receive correlation matrices $\mathbf{R}_r$.

2.3 Transmit Zero-forcing Beamforming Methods

As stated in Section 2.1, finding the optimal sum rate for beamforming is a difficult non-convex optimization problem [23] as it involves SINR balancing via a set of $\mathbf{T}_j$, $\mathbf{R}_j$ and $\mathbf{R}_{d_j d_j}$ matrices. Simpler but sub-optimal schemes may be obtained via the zero-forcing beamforming approach, which enforces zero co-channel interference (CCI) among active users. For single-antenna terminals, zero-forcing beamforming may be easily implemented using the pseudo-inverse of the channel matrix. This creates orthogonal channels to each user where data to each user may then be encoded independently. For multi-antenna terminals, the pseudo-inverse method may also be used. This creates parallel orthogonal channels not only between users, but also within each user terminal. This approach is sub-optimal however, as it is better to impose orthogonality between users only, because antennas located at the same terminal can cooperate effectively. In this way, techniques such as layered space-time coding or singular value decomposition (SVD)-based beamforming with waterfilling could then be considered at each user terminal for better performance. This is commonly referred to as block diagonalization (BD) [27], [28], [29]. Since block diagonalization is also applicable to single-antenna terminals, it has a more generalized form and its description is given first.

Having given the assumptions regarding the channel, it is appropriate to highlight that the applicability of beamforming methods in practice is contingent on the availability of timely and accurate channel state information (CSI) at the base station transmitter. In general, this means that the channel coherence time should be longer than the data block duration plus the feedback delays associated with the CSI feedback. This is often referred to as a slow-fading channel.

2.3.1 Block Diagonalized Zero-forcing Beamforming

Block diagonalization enforces zero co-channel interference (CCI) among users with multi-antenna terminals. To accomplish this, the pre-coding matrices $\mathbf{T}_j$ are chosen so that each user's beam is forced to lie within the nullspace of a composite channel matrix that comprises all other users' channel matrices. In other words, the zero co-channel interference (CCI) constraint forces $\mathbf{T}_j$ to lie in the null space of $\tilde{\mathbf{H}}_j$ [27], where

$$\tilde{\mathbf{H}}_j = [\mathbf{H}_1^T \ \cdots \ \mathbf{H}_{j-1}^T \ \mathbf{H}_{j+1}^T \ \mathbf{H}_K^T]^T \tag{2.19}$$

and K is the number of users. In this way, K parallel single-user MIMO channels are created because

$$\mathbf{H}_j\mathbf{T}_i = 0 \quad \forall i \neq j. \tag{2.20}$$

Note that for a group of K users, block diagonalization is possible when the following pre-coding constraint is met

$$(M - \textstyle\sum_{i=1, i\neq j}^{K} N_i) > 0, \ \forall j, \tag{2.21}$$

where M is the number of transmit antennas at the base station and N_i is the number of antennas at each user terminal.

The simplest from of block diagonalization makes use of pre-coding matrices $\mathbf{T}_j$ only, without the use of receive-processing matrices $\mathbf{R}_j$. To recover the data streams at each user terminal, schemes such as V-BLAST may be used in each block-diagonalized channel. For convenience, this form of block diagonalization is referred to as "direct-BD" since the pre-coding matrices are derived directly from the channel matrices $\mathbf{H}_j$, without the involvement of the receive-processing matrices $\mathbf{R}_j$. In practice, direct-BD is desirable in situations where resources for the transmission of $\mathbf{R}_j$ from the base station to each user is limited or unavailable. The sum rate expression for direct-BD can be obtained using (2.10) and given $\mathbf{H}_j\mathbf{T}_i = 0 \ \forall i \neq j$ in (2.20), the interference term in (2.7) reduces to zero so that $\mathbf{z}_j = \mathbf{n}_j$ and $\mathbf{R}_{z_jz_j} = \sigma^2\mathbf{I}_{N_j}$. The sum rate for direct-BD (DBD) is then

$$R_{Sum}^{\text{DBD}} = \max_{\substack{\mathbf{T}_j,\, \mathbf{R}_{d_jd_j},\, j=1,\cdots,K \\ \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss})\leq P}} \sum\nolimits_{j=1}^{K} \log\det\left(\mathbf{I}_{N_j} + \mathbf{H}_j\mathbf{T}_j\mathbf{R}_{d_jd_j}\mathbf{T}_j^H\mathbf{H}_j^H / \sigma^2\right). \tag{2.22}$$

One means of finding $\mathbf{T}_j$ is via SVD($\tilde{\mathbf{H}}_j$)

$$\underbrace{\tilde{\mathbf{H}}_j}_{\left(\sum_{i=1,i\neq j}^{K} N_i \,\mathrm{x}\, M\right)} = \tilde{\mathbf{U}}_j\tilde{\mathbf{\Sigma}}_j[\quad \underbrace{\tilde{\mathbf{V}}_j^{(1)}}_{\left(M \,\mathrm{x} \sum_{i=1,i\neq j}^{K} N_i\right)} \quad \underbrace{\tilde{\mathbf{V}}_j^{(0)}}_{\left(M \,\mathrm{x} \left(M - \sum_{i=1,i\neq j}^{K} N_i\right)\right)} \quad]^H. \tag{2.23}$$

Since $\tilde{\mathbf{V}}_j^{(0)}$ forms an orthonormal basis for the row or left null space of $\tilde{\mathbf{H}}_j$, its column vectors can therefore be used as part of the pre-coding matrix $\mathbf{T}_j$ of user *j*, i.e., $\mathbf{T}_j = \tilde{\mathbf{V}}_j^{(0)}\mathbf{P}_j$, where $\mathbf{P}_j$ is the other part of the pre-coding matrix to be determined. This form of $\mathbf{T}_j$ makes (2.22) realizable because

$$\mathbf{H}_r\mathbf{T}_r = \begin{bmatrix} \mathbf{H}_1\tilde{\mathbf{V}}_1^{(0)}\mathbf{P}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2\tilde{\mathbf{V}}_2^{(0)}\mathbf{P}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_K\tilde{\mathbf{V}}_K^{(0)}\mathbf{P}_K \end{bmatrix}, \tag{2.24}$$

where $\mathbf{H}_r = [\mathbf{H}_1^T\ \mathbf{H}_2^T \cdots \mathbf{H}_K^T]^T$ and $\mathbf{T}_r = [\mathbf{T}_1\ \mathbf{T}_2 \cdots \mathbf{T}_K]$. Note that each pre-coded channel of the form $\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$ may be thought of as a *projected* channel $\mathbf{H}_{P_j} = \mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$ with dimensions

$$N_j \,\mathrm{x}\, (M - \sum\nolimits_{i=1,\, i\neq j}^{K} N_i). \tag{2.25}$$

Note that the pre-coding constraint in (2.21) is simply derived from (2.25) and states that the number of columns in each projected channel must be > 0 to realize block diagonalization in a group of chosen users.

The block orthogonalization process has created K single-user MIMO channels and the optimal solution for $\mathbf{P}_j$ is then clear via [30], i.e., using $\mathrm{SVD}(\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)})$, set $\mathbf{P}_j = \mathbf{V}_j^{(1)}$, where

$$\mathbf{H}_{p_j} \triangleq \mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j\begin{bmatrix} \mathbf{\Sigma}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}[\mathbf{V}_j^{(1)} \quad \mathbf{V}_j^{(0)}]^H. \tag{2.26}$$

When the use of receive-processing matrices $\mathbf{R}_j$ is possible, the optimal solution as in any single-user MIMO channel is obtained by setting $\mathbf{R}_j = \mathbf{U}_j$, where $\mathbf{U}_j$ is obtained from (2.26). In this case (2.12) becomes $\mathbf{z}_j = \mathbf{U}_j\mathbf{n}_j$ so that $\mathbf{R}_{z_jz_j} = \sigma^2\mathbf{I}_{N_j}$ since $\mathbf{U}_j$ is unitary and the sum rate in (2.13) becomes

$$\begin{aligned} R_{Sum}^{\mathrm{BD}} &= \max_{\substack{\mathbf{T}_j,\, \mathbf{R}_{d_jd_j},\, j=1,\cdots,K \\ \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss})\le P}} \sum\nolimits_{j=1}^{K} \log\det\left(\mathbf{I}_{N_j} + \mathbf{R}_j\mathbf{H}_j\mathbf{T}_j\mathbf{R}_{d_jd_j}\mathbf{T}_j^H\mathbf{H}_j^H\mathbf{R}_j^H/\sigma^2\right) \\ &= \sum\nolimits_{j=1}^{K} \log\det\left(\mathbf{I}_{N_j} + \mathbf{R}_{d_jd_j}\mathbf{\Sigma}_j^2/\sigma^2\right), \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{d_jd_j}) \le P, \end{aligned} \tag{2.27}$$

where the power constraint $\mathrm{tr}(\mathbf{R}_{d_jd_j}) \le P$ is obtained because $\mathbf{T}_j$ is assumed to be unitary as is commonly implemented for block diagonalized systems.

As shown in (2.27), direct access to the projected channels' spatial modes is achieved and waterfilling can be done to maximize each user's throughput. Note however that all $\mathbf{R}_j = \mathbf{U}_j$ matrices must be transmitted from the base station to each user because only the base station possesses the complete channel state information needed to design all $\mathbf{T}_j$ and $\mathbf{R}_j$ matrices. Note also that (2.27) may be used as a means of computing the best sum rate of a given direct-BD system.

As shown above, each user j has a number of spatial modes created after block diagonalization. For direct-BD, each spatial mode allocated to a user will require a corresponding antenna-RF chain at that user's terminal. The maximum possible number of spatial modes that could be allocated to a user is therefore limited by number of antenna-RF chains available at its terminal. Along with power control, a dynamic spatial-mode allocation strategy may be done in accordance with each user's channel-rate requirement.

This enables the scheduling of more users compared to a fixed allocation regime, since the base station transmission resources are limited. In this way, some terminals may deactivate one or more antenna-RF chains when the number of spatial modes allocated to them is less than the antenna-RF chains available. This mechanism of spatial mode control is sub-optimal because the un-used antenna-RF chains could contribute to better diversity performance.

To address this, the Coordinated Transmit-Receive (CTR) [27] and the iterative null space directed SVD (Nu-SVD) [29] schemes perform block diagonalization on a projected virtual channel $\mathbf{H}_e$, which is defined as $\mathbf{H}_e \triangleq [[\mathbf{R}_1^H \mathbf{H}_1]^T \cdots\cdots [\mathbf{R}_K^H \mathbf{H}_K]^T]^T$. The BD-SDM process in CTR and Nu-SVD create zero co-channel interference after the receive-processing matrices at the user receivers instead of just after the antennas for direct-BD. The receive-processing matrices $\mathbf{R}_j$ are appropriately dimensioned according to the desired number of spatial modes to be activated for each user j. For CTR, the $\mathbf{R}_j$ matrices are labeled as $\mathbf{W}_j$ in [27]. Since all antennas remain in use, both CTR and Nu-SVD provide better diversity performance than the simple scheme of de-activating antennas according to the number of modes needed at each user.

2.3.2 Channel Inversion Zero-forcing Beamforming

The block diagonalization procedure described above requires $2K$ SVD operations. When user terminals are equipped with only one antenna, a simpler way of achieving zero-forcing beamforming (ZFBF) is available via pre-coding with the right-side Moore-Penrose pseudo-inverse. Let $\mathbf{h}_j \in \mathbb{C}^{1\mathrm{x}M}$ be a vector with complex entries that represent the channel between the base station and user j and let $\mathbf{H} = [\mathbf{h}_1^T\ \mathbf{h}_2^T \cdots \mathbf{h}_K^T]^T$, $K = N$ be a matrix that contains a concatenation of all channel vectors $\{\mathbf{h}_1, \cdots \mathbf{h}_K\}$. The right-side Moore-Penrose pseudo-inverse of $\mathbf{H}$ is $\mathbf{H}^+ \in \mathbb{C}^{M\mathrm{x}N}$ where

$$\mathbf{H}^+ = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}. \tag{2.28}$$

Since $(\mathbf{HH}^H)^{-1}$ exists only if $N \leq M$, a pre-coding constraint in the form of $K \leq M$ exists for single-antenna user terminals. When the potential pool of users S contains $K > M$ users, a subset $S_r \subset S$ of users where $|S_r| = K_r \leq M$, must be chosen. Henceforth, it is assumed that a subset $S_r \subset S$ of $K_r = M$ users is chosen and the associated overall channel matrix is $\mathbf{H}_r$. Let the following vectors represent concatenation of quantities across the system:

$$\left.\begin{array}{ll} \text{Overall receive vector} & \mathbf{y} = [y_1, y_2, \cdots y_K]^T \\ \text{Overall data vector} & \mathbf{d} = [d_1, d_2, \cdots d_K]^T \\ \text{Overall noise vector} & \mathbf{n} = [n_1, n_2, \cdots n_K]^T \end{array}\right\} \tag{2.29}$$

The overall transmit vector is then $\mathbf{s} = \mathbf{H}_r^+ \mathbf{d}$ and the overall receive vector becomes

$$\mathbf{y} = \mathbf{H}_r \mathbf{s} + \mathbf{n} = \mathbf{H}_r \mathbf{H}_r^+ \mathbf{d} + \mathbf{n} = \mathbf{d} + \mathbf{n}. \tag{2.30}$$

From (2.22), the sum rate R_{Sum}^{CI} of such a system with channel-inversion pre-coding is

$$R_{Sum}^{\mathrm{CI}} = \sum\nolimits_{i=1}^{K_r} \log_2\left(1 + \gamma_i / \sigma^2\right), \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss}) \leq P, \tag{2.31}$$

where $\mathbf{H}_r \mathbf{R}_{ss} \mathbf{H}_r^+ = \mathbf{H}_r \mathbf{H}_r^+ \mathbf{R}_{d_i d_i} (\mathbf{H}_r \mathbf{H}_r^+)^H = \mathrm{diag}(\gamma_1, \cdots, \gamma_{K_r})$. For convenience, this zero-forcing beamforming scheme will be referred to as "transmit channel-inversion beamforming" or TCIBF. Note the power constraint

$$\mathrm{tr}(\mathbf{R}_{ss}) = \mathrm{tr}(\mathbf{H}_r^+ \mathbf{R}_{d_j d_j} \mathbf{H}_r^{+H}) = \sum\nolimits_{i=1}^{K_r} (\gamma_i / b_i) \leq P, \tag{2.32}$$

$$\text{where } \{1/b_i\} = [(\mathbf{H}_r \mathbf{H}_r^H)^{-1}]_{i,i}. \tag{2.33}$$

The optimal choice for $\{\gamma_i : i = 1, \cdots, K_r\}$ is via waterfilling where,

$$\gamma_i = (\mu_T b_i - \sigma^2)_+ , \; i = 1, \ldots\ldots, K_r - p + 1, \tag{2.34}$$

$$\mu_T = \frac{P}{K_r - p + 1}\left(1 + \frac{\sigma^2}{P} \sum_{i=1}^{K_r - p + 1} \frac{1}{b_i}\right), \tag{2.35}$$

where $(x)_+ = \max(0,x)$ and μ_T is the water-level. From (2.35), it can be seen that the b_i values are in effect the orthogonalized channel gains. Variable p is first set to 1 and subsequently incremented whenever negative power allocations occur. The data stream with the most negative power allocation is assigned zero power and removed from further consideration. The process is repeated until $\gamma_i \geq 0\ \forall i = 1,\cdots,K_r$. Users with zero-power assignment are effectively de-selected and a subset of $S_a \subset S_r$ active users remains, where $|S_a| = K_a$. However, relying on the $(x)_+$ power allocation policy alone as in (2.34) does not help optimize the TCIBF sum rate because it is also dependent on the overall channel matrix $\mathbf{H}_r$, whose impact can be seen from (2.4). Low b_i values will constrain γ_i to low values so that the power constraint in (2.32) can be met. This is signal attenuation in effect and results in low TCIBF sum rates. Note that orthogonalized channel gains b_i may also be viewed in terms of projections as [11]

$$b_i = \left\| \mathbf{h}_i . \mathrm{null}\left(\tilde{\mathbf{H}}_{ri}\right) \right\|^2, \tag{2.36}$$

where $\tilde{\mathbf{H}}_{ri} = \left[\mathbf{h}_1^T \cdots \mathbf{h}_{i-1}^T\ \mathbf{h}_{i+1}^T \cdots \mathbf{h}_{K_r}^T\right]^T$ and null($\mathbf{A}$) is an orthonormal basis of $\mathbf{A}$'s null space.

Chapter 3

AN ANALYSIS OF ZERO-FORCING BEAMFORMING SYSTEMS

In line with the intention of enhancing the feasibility of fielding multi-user MIMO systems, this chapter starts by highlighting the issues affecting the sum rate of zero-forcing beamforming systems. It describes the mechanisms that affect the sum rates for a given channel and then seeks to find methods for sum-rate improvement. In particular, the impact of antenna selection and user selection and the underlying mechanisms governing their behavior will be studied in detail. Where possible, expressions to quantify their impact will be given on an ergodic basis and on an asymptotic basis, for example, when the user pool becomes very large.

3.1 Transmit Channel Inversion Beamforming (TCIBF) for Single-Antenna Terminals

3.1.1 Factors Affecting TCIBF Sum Rates

Recall from Chapter 2 that the TCIBF sum rate is given by

$$R_{Sum}^{\mathrm{CI}} = \sum\nolimits_{i=1}^{K_r} \log_2\left(1+\gamma_i/\sigma^2\right), \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss}) \le P, \tag{3.1}$$

and the power constraint is

$$\mathrm{tr}(\mathbf{R}_{ss}) = \mathrm{tr}(\mathbf{H}_r^{+}\mathbf{R}_{dd}\mathbf{H}_r^{+H}) = \sum\nolimits_{i=1}^{K_r}(\gamma_i/b_i) \le P, \tag{3.2}$$

$$\text{where } \{1/b_i\} = [(\mathbf{H}_r\mathbf{H}_r^H)^{-1}]_{i,i}. \tag{3.3}$$

The optimal choice for $\{\gamma_i : i=1,\cdots,K_r\}$ is via waterfilling where,

$$\gamma_i = (\mu_T b_i - \sigma^2)_+ \,,\ \ i=1,\ldots\ldots,K_r-p+1, \tag{3.4}$$

$$\mu_T = \frac{P}{K_r - p + 1}\left(1 + \frac{\sigma^2}{P}\sum_{i=1}^{K_r - p+1}\frac{1}{b_i}\right), \tag{3.5}$$

where $(x)_+ = \max(0, x)$ and μ_T is the water-level. From (3.5), it can be seen that the b_i values are in effect the orthogonalized channel gains.

However, relying on the $(x)_+$ power allocation policy alone as in (3.4) does not help optimize the TCIBF sum rate because it is also dependent on the overall channel matrix $\mathbf{H}_r$, whose impact can be seen from (3.2). Low b_i values will constrain γ_i to low values so that the power constraint can be met. This is signal attenuation in effect and results in low TCIBF sum rates. This means that the mechanism of allocating zero power to those channels with low gains does not help to improve the sum rate significantly, especially when the channel conditions are poor. To see the reason behind this phenomenon, an expansion using $\text{SVD}(\mathbf{H}_r)$ on (3.3) is helpful

$$\begin{aligned} 1/b_i &= [(\mathbf{H}_r\mathbf{H}_r^H)^{-1}]_{i,i} = [(\mathbf{U}\Sigma\Sigma\mathbf{U}^H)^{-1}]_{i,i} \\ &= \sum\nolimits_{j=1}^{K_r} u_{ij}u_{ij}^{*}\lambda_j^{-1}, \quad i = 1, \ldots., K_r, \end{aligned} \tag{3.6}$$

where λ_j are the eigenvalues of $\mathbf{H}_r\mathbf{H}_r^H$. It is seen from (3.6) that each $1/b_i$ value is influenced by the *entire* set of eigenvalues $\{\lambda_j : j = 1, \cdots, K_r\}$. Low λ_j's result in large $1/b_i$ values $\forall i$, which then constrains γ_i to low values and results in a low sum rate. Since the converse is true, it motivates the search for methods to increase the λ_j values. To begin, the orthogonalized channel gains b_i may be viewed in terms of projections as [11]

$$b_i = \left\|\mathbf{h}_i.\text{null}\left(\tilde{\mathbf{H}}_{ri}\right)\right\|^2, \tag{3.7}$$

where $\tilde{\mathbf{H}}_{ri} = \left[\mathbf{h}_1^T \cdots \mathbf{h}_{i-1}^T \ \mathbf{h}_{i+1}^T \cdots \mathbf{h}_{K_r}^T\right]^T$ and null(**A**) is an orthonormal basis of **A**'s null space. Hence b_i will be high when the channel row vectors are close to being orthogonal and low when highly correlated channel row vectors are present. This points to the first approach for improving the λ_j values, namely, given *S*, a pool of $K > M$ users, judicious selection must be made so that $\mathbf{H}_r$ is populated by row vectors from $K_r = M$ users that are close to being orthogonal.

Despite this, sometimes a subset $S_r \subset S$ of $K_r = M$ is not sufficiently orthogonal and higher b_i values can be obtained by *removing* one or more highly correlated row vectors. This is still in line with (3.7) but done at the expense of a reduced active user subset where $1 \le K_r < M$. The impact of removing row vectors, which may be interpreted as removing users or as removing receive antennas from the overall system, can be seen starting with (3.6). Let $S_s \subset S_r$ be a subset with one user *removed* from S_r and let $\mathbf{H}_s \subset \mathbf{H}_r$ be the associated channel matrix. Let $\lambda_{\max}(\mathbf{Z}_s) \ge \lambda_2(\mathbf{Z}_s) \ge \cdots \ge \lambda_{\min}(\mathbf{Z}_s)$ and $\lambda_{\max}(\mathbf{Z}_r) \ge \lambda_2(\mathbf{Z}_r) \ge \cdots \ge \lambda_{\min}(\mathbf{Z}_r)$ be the ordered eigenvalues of $\mathbf{Z}_s = \mathbf{H}_s\mathbf{H}_s^H$ and $\mathbf{Z}_r = \mathbf{H}_r\mathbf{H}_r^H$, respectively. With $K_r \le M$, it can be shown that

$$\lambda_{\max}(\mathbf{Z}_r) \ge \lambda_{\max}(\mathbf{Z}_s) \ge \lambda_2(\mathbf{Z}_r) \ge \lambda_2(\mathbf{Z}_s) \ge \cdots\cdots\cdots \ge \lambda_{\min}(\mathbf{Z}_s) \ge \lambda_{\min}(\mathbf{Z}_r). \qquad (3.8)$$

The proof for (3.8) is given in Appendix A. Hence the extremal eigenvalues of $\mathbf{Z}_s$ lie between those of $\mathbf{Z}_r$. In particular, the minimum and the lower eigenvalues of $\mathbf{Z}_s$ are *increased.* If increases in the lower eigenvalues outweigh the decreases in the higher values, then lower $\{1/b_i\}$ values may result. Note that (3.8) is true regardless of the transmit- and receive-correlation matrices $\mathbf{R}_r^{1/2}$ and $\mathbf{R}_t^{1/2}$ that may be associated with $\mathbf{H}_j = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$ (see Chapter 2 on channel model used). Further insights are given in the next section where the eigenvalue distributions of $\mathbf{Z}_r = \mathbf{H}_r\mathbf{H}_r^H$ are examined.

3.1.1.1 Case When Receive Antennas are Uncorrelated

In this case, we set the correlation matrices $\mathbf{R}_r^{1/2} = \mathbf{I}_N$ in $\mathbf{H}_r = \mathbf{R}_r^{1/2}\mathbf{H}_w$ so that $\mathbf{H}_r = \mathbf{H}_w \in \mathbb{C}^{N \times M}$. The base station antennas are assumed to be well spaced enough to allow $\mathbf{R}_t^{1/2} = \mathbf{I}_M$. The i.i.d. complex Gaussian entries of $\mathbf{H}_r$ are scaled as $\mathcal{N}_c(0,1)$ so that

$\mathbf{H}_r\mathbf{H}_r^H$ is a complex Wishart matrix **W**, i.e., a random complex Hermitian square $N \times N$ matrix. In general, **W** is defined as

$$\mathbf{W} = \begin{cases} \mathbf{H}_r\mathbf{H}_r^H & N < M \\ \mathbf{H}_r^H\mathbf{H}_r & N \geq M. \end{cases} \tag{3.9}$$

Wishart matrices are parameterized by

$$\begin{aligned} m &= \max\{N, M\} \\ n &= \min\{N, M\}, \end{aligned} \tag{3.10}$$

and often written as $\mathbf{W}(n,m)$. The eigenvalues of **W** are random variables and can be described by their joint probability density function. The joint density of the ordered eigenvalues of $\mathbf{W}(n,m)$ is [6]

$$p_\lambda(\lambda_1, \cdots \lambda_n) = K_{n,m}^{-1} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{m-n} \prod_{i<j} (\lambda_i - \lambda_j)^2, \tag{3.11}$$

where $K_{n,m}$ is a normalizing factor that can be found in [63]. The unordered eigenvalues have the joint density

$$p_\lambda(\lambda_1, \cdots \lambda_n) = (n! K_{n,m})^{-1} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{m-n} \prod_{i<j} (\lambda_i - \lambda_j)^2. \tag{3.12}$$

The distribution of one of the unordered eigenvalues is [6]

$$p_\lambda(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \varphi_i(\lambda)^2 \lambda^{m-n} e^{-\lambda}, \tag{3.13}$$

where

$$\varphi_{k+1}(\lambda) = \left[\frac{k!}{(k+m-n)!} \right]^{1/2} L_k^{m-n}(\lambda), \quad k = 0, \ldots\ldots, n-1 \tag{3.14}$$

and $L_k^{m-n}(\lambda)$ is the associated Laguerre polynomial of order k where

$$L_k^{m-n}(\lambda) = \frac{1}{k!} e^{\lambda} \lambda^{n-m} \frac{d^k}{d\lambda^k} (e^{-\lambda} \lambda^{m-n+k}) \tag{3.15}$$

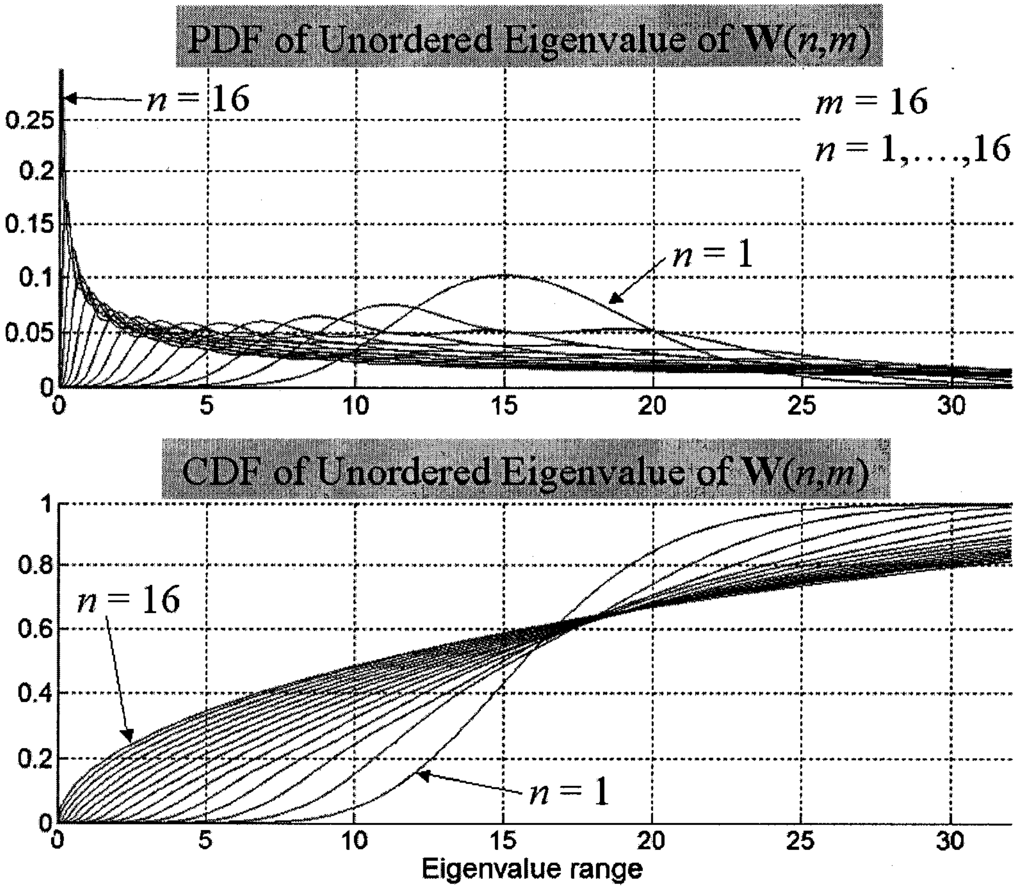


Figure 3.1. PDF and CDF of unordered eigenvalue of $\mathbf{W}(n,m)$

$$=\sum_{j=0}^{k}\binom{k+m-n}{k\text{-}j}\frac{(-\lambda)^j}{j!} \tag{3.16}$$

The impact of user/antenna selection on the orthogonalized channel gains b_i can be examined by studying its impact on a *representative* unordered-eigenvalue distribution. Removing users and their corresponding row vectors from $\mathbf{H}_r$ on a *random* selection basis is equivalent to reducing the parameter n in $\mathbf{W}(n,m)=\mathbf{H}_r\mathbf{H}_r^H$. Figure 3.1 shows the unordered-eigenvalue probability distribution function in (3.13) and the accompanying cumulative distribution function of an example with $\mathbf{H}_r \in \mathbb{C}^{16x16}$, $\mathbf{W}(n,m=16)$, where n is decreased from 16 to 1. As shown, the spread of eigenvalues is widest when $n=m=16$

and in fact, its PDF exhibits a negative exponential behavior with a prominent low-value cluster. This low-value cluster diminishes rapidly even with $n = m-1$. From the CDF plots, it is clear that the eigenvalue range becomes narrower as the value of n is decreased, i.e., the lower-eigenvalue range is raised while the higher range is lowered, which is in line with (3.8). The eigenvalue variance versus parameter n is tabulated in Table 3.1.

Table 3.1: Eigenvalue variance σ_λ^2 versus parameter n in $\mathbf{W}(n,m) = \mathbf{H}_r\mathbf{H}_r^H$, $\mathbf{H}_r \in \mathbb{C}^{16x16}$.

n	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
σ_λ^2	247	238	224	208	192	176	160	144	128	112	96	80	64	48	32	16

It is important to note that the lower-eigenvalue range may be raised by orders of magnitude while the higher-eigenvalue range may remain largely in the same order of

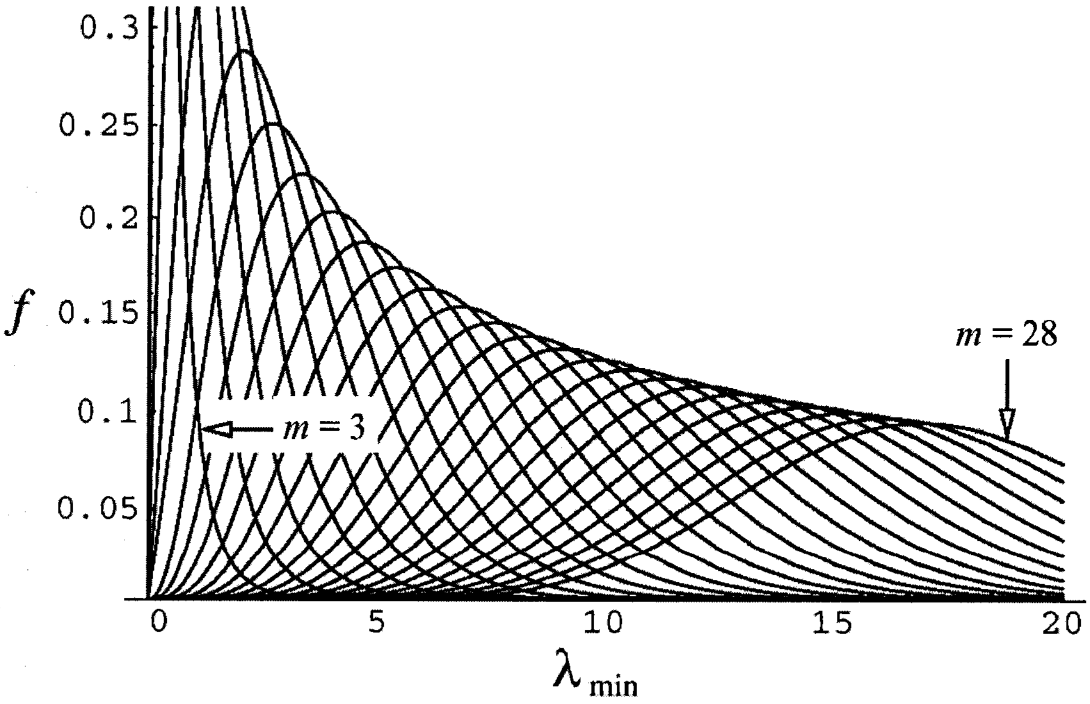


Figure 3.2. The PDF of $\lambda_{\min}$ for $\mathbf{W}(n=3,m)$, $m = 3,4,5,\ldots\ldots,28$. Adapted from [63]

magnitude as n is decreased. This mechanism supports the argument that the b_i values may be raised by reducing the number of rows in $\mathbf{H}_r$. Note again that this argument is statistically true even for this case where $\mathbf{H}_r = \mathbf{H}_w$, i.e., without the influence of the receive-correlation matrix $\mathbf{R}_r$.

In terms of the ordered eigenvalues, a view of the minimum eigenvalue ($\lambda_{\min}$) behavior may be seen from [63] where the minimum-eigenvalue PDFs of both real and complex Wishart matrices were derived. Both PDF expressions were of similar form and a sample $\lambda_{\min}$-PDF plot for a series of real Wishart matrices was given for $\mathbf{W}(3,3)$ to $\mathbf{W}(3,28)$. This plot is reproduced in Figure 3.2 and it is clear that the PDF of $\lambda_{\min}$ shifts to a higher range as n becomes increasingly less than m. Again, it is important to note that the lower-eigenvalue range may be raised by orders of magnitude.

3.1.1.2 Case When Receive Antennas are Correlated

In this case, the correlation matrix at the user terminal is $\mathbf{R}_r^{1/2} \neq \mathbf{I}_N$. When the constant correlation model or the exponential correlation model is used, the resulting correlation matrix $\mathbf{R}_r^{1/2} \in \mathbb{R}^{N \times N}$ can be determined using only the first row vector $\mathbf{r} = (r_{11}, r_{12}, \cdots, r_{1N})$ in $\mathbf{R}_r^{1/2}$. The matrix may then be viewed as a special case of a Hermitian Toeplitz matrix. The effect of the correlation value φ upon the eigenvalues of $\mathbf{R}_r^{1/2}$ may be seen via its extreme eigenvalue bounds. A simple algorithm in [64] may be used for this purpose where the proposed bounds pertain to matrices of any dimension and amount to computing inner products. The maximal eigenvalue $\overline{\lambda}$ of any Hermitian Toeplitz matrix is given by the inner product

$$\langle \tilde{\mathbf{r}}, \overline{\mathbf{w}} \rangle, \text{ where} \tag{3.17}$$

$$\tilde{\mathbf{r}} = (r_{11}, |r_{12}|, \cdots, |r_{1N}|), \tag{3.18}$$

$$\overline{\mathbf{w}} = (1, \overline{\lambda}_2, \cdots, \overline{\lambda}_N), \tag{3.19}$$

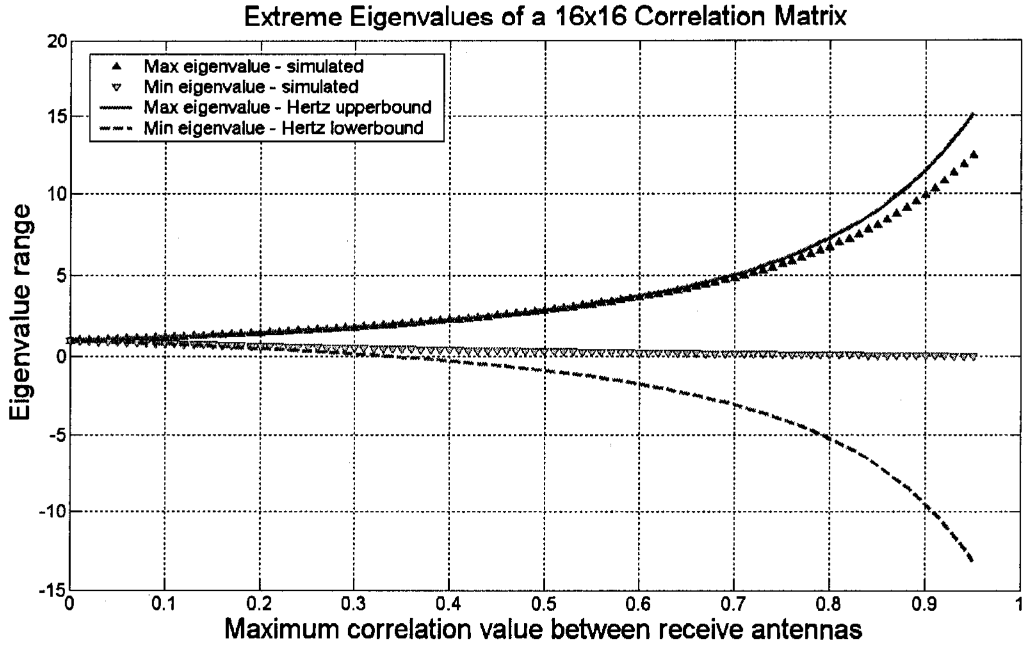


Figure 3.3. Extreme eigenvalues of a $\mathbf{R}_r^{1/2} \in \mathbb{R}^{16\mathrm{x}16}$ correlation matrix

$$\bar{\lambda}_i = 2\cos\left(\frac{\pi}{\lfloor (N-1)/(i-1) \rfloor + 2}\right), i = 2, \cdots, N, \tag{3.20}$$

where $\lfloor x \rfloor$ denotes the floor of x. The minimal eigenvalue $\underline{\lambda}$ of any Hermitian Toeplitz matrix is given by the inner product

$$\langle \tilde{\mathbf{r}}, \underline{\mathbf{w}} \rangle, \text{ where} \tag{3.21}$$

$$\underline{\mathbf{w}} = (1, \underline{\lambda}_2, \cdots, \underline{\lambda}_N), \tag{3.22}$$

$$\underline{\lambda}_i = -\bar{\lambda}_i, \quad i = 2, \cdots, N. \tag{3.23}$$

In Figure 3.3, an example with $\mathbf{R}_r^{1/2} \in \mathbb{R}^{16\mathrm{x}16}$ utilizing the exponential correlation model is given. The extreme eigenvalues versus the correlation φ are computed and compared with the bounds given by [64]. As shown, the upper bound is tight while the lower bound is good till about $\varphi = 0.32$. In any case, it is clear that the eigenvalue spread widens, as the inter-antenna correlation φ becomes higher.

Since $\bar{\mathbf{w}} = |\underline{\mathbf{w}}|$, the direct proportion between the eigenvalue spread and φ is due to $\tilde{\mathbf{r}}$, as can be seen from (3.18). This means that the removal of antennas that are highly correlated with other elements will help reduce the eigenvalue spread, i.e., raising the lower eigenvalues and lowering the higher eigenvalues. To examine the combined effect in $\mathbf{H}_r\mathbf{H}_r^H = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2}$, we note the following from [67]:

(a) Given two $(n \mathrm{x} n)$ positive definite matrices **A** and **B**, then

$$\lambda \text{ is eigenvalue of } \mathbf{BA} \Leftrightarrow \lambda \text{ is eigenvalue of } \mathbf{B}^{1/2}\mathbf{A}\mathbf{B}^{1/2} \tag{3.24}$$

(b) Given two $(n \mathrm{x} m)$ matrices **A** and **B** with rank $r = \min\{n, m\}$, let $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) \geq 0$, $\sigma_1(\mathbf{B}) \geq \cdots \geq \sigma_r(\mathbf{B}) \geq 0$ and $\sigma_1(\mathbf{BA}^H) \geq \cdots \geq \sigma_r(\mathbf{BA}^H) \geq 0$ be the singular values of **A**, **B** and $\mathbf{BA}^H$, respectively. Then

$$\sigma_{i+j-1}(\mathbf{BA}^H) \leq \sigma_i(\mathbf{A})\sigma_j(\mathbf{B}), \quad 1 \leq \{i, j\} \leq r \text{ where } (i+j) \leq (r+1) \tag{3.25}$$

Let $\mathbf{A} = \mathbf{H}_w\mathbf{H}_w^H$ and $\mathbf{B} = \mathbf{R}_r$, we can see from combining (3.24) and (3.25) that:

$$\lambda_{i+j-1}(\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2}) \leq \lambda_i(\mathbf{H}_w\mathbf{H}_w^H)\lambda_j(\mathbf{R}_r), \quad 1 \leq \{i, j\} \leq N \text{ where } (i+j) \leq (N+1) \tag{3.26}$$

Note that (3.26) applies because the singular values in (3.25) are non-negative. Equation (3.26) shows the trend for lower eigenvalues in $\mathbf{H}_r\mathbf{H}_r^H = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2}$ when the eigenvalues in either $\mathbf{H}_w\mathbf{H}_w^H$ and/or $\mathbf{R}_r$ become lower. For example, (3.26) shows that $\lambda_{\min}(\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2}) \leq \lambda_{\min}(\mathbf{H}_w\mathbf{H}_w^H)\lambda_1(\mathbf{R}_r)$ or $\lambda_{\min}(\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2}) \leq \lambda_1(\mathbf{H}_w\mathbf{H}_w^H)\lambda_{\min}(\mathbf{R}_r)$, where $\lambda_{\min}(\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2}) = \lambda_N(\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{H}_w^H\mathbf{R}_r^{1/2})$.

It is therefore seen from the above that two mechanisms for achieving higher b_i values are available, namely, (a) judicious user subset selection from a user pool to choose $S_r \subset S$ of $K_r = M$ users and (b) de-selecting users to result in a subset $S_s \subset S_r$ of $1 \leq K_r < M$ users. The sum rate expression in (3.1) may be updated to reflect the possibility of maximization via these two mechanisms as

$$R_{Sum}^{\mathrm{CI}}\left(\mathbf{H}_s\right)_{\max} = \max_{\mathbf{H}_s, \mathbf{R}_{ss}} \sum\nolimits_{i=1}^{K_s} \log_2\left(1+\gamma_i / \sigma^2\right), \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \leq P, \tag{3.27}$$

where $\mathbf{H}_s$ is the channel matrix associated with S_s and $\mathbf{R}_{ss}$ is found using waterfilling.

Note that (b) is equivalent to removing antennas that are highly correlated from within a chosen subset. Regarding point (b) however, the benefit of user removal upon the TCIBF sum rate remains to be seen, as it is known that Sato's cooperative upper bound [54], which represents the sum-rate upper bound for multi-user broadcast channels, will be reduced. This is addressed in the next section where an analysis on the conditions for sum-rate increase during user/antenna de-selection is given.

3.1.2 Sum Rate Impact of User/Antenna De-selection for $K \leq M$

Let S represent a given pool of $K \leq M$ users and assume that there is no possibility of replacing any user. Under poor channel conditions, the waterfilling process may assign zero power to users associated with low b_i values. Although this helps mitigate the signal attenuation problem in terms of meeting the power constraint in (3.1), it does not achieve the best sum rate since the poor channel condition still exerts its influence on the remaining b_i values as shown by (3.6). It has been recognized that judicious user-subset selection is needed for TCIBF sum-rate maximization, e.g., see [11] and [66]. However, there has been little analysis to show the underlying mechanisms behind the benefit of user-subset selection. To analyze the conditions under which TCIBF sum rate improvements would occur with user-subset reduction, the $1/b_i$ values may first be represented as

$$\{1/b_i : i = 1,\cdots,K_r\} = \{(A_{11})_r \Delta_r^{-1},........,(A_{K_rK_r})_r \Delta_r^{-1}\}, \tag{3.28}$$

where $(A_{ii})_r$ are cofactors associated with the principal diagonal elements h_{ii} in $\mathbf{H}_r\mathbf{H}_r^H$ and $\Delta_r = \det(\mathbf{H}_r\mathbf{H}_r^H)$. Each $(A_{ii})_r$ is found after eliminating row i and column i in $\mathbf{H}_r\mathbf{H}_r^H$, which corresponds to eliminating row i in $\mathbf{H}_r \in \mathbb{C}^{K_r \mathrm{x} M}$ to give sub-matrix $\mathbf{H}_s \in \mathbb{C}^{K_s \mathrm{x} M}$ where $K_s = K_r - 1$, i.e., eliminating one user i. For the general case starting with $K_r = M$, let $\gamma_i = (\mu_T b_i - \sigma^2)_{k+}$ represent the computation of γ_i after having removed k rows and having waterfilling re-applied. Note that $K_s = K_r - k$ for the general case and substituting (3.28) into (3.4) and (3.5), we obtain

$$\gamma_i = \left(\frac{\Delta_s P}{(A_{ii})_s K_a} \left(1 + \frac{\sigma^2}{\Delta_s P} \sum_{j=1}^{K_a} (A_{jj})_s \right) - \sigma^2 \right)_{k+}, \tag{3.29}$$

where $(A_{ii})_s$ and Δ_s are associated with $\mathbf{H}_s\mathbf{H}_s^H$ and $K_a = K_s - p + 1$. Using (3.29) in (3.27),

$$R_{Sum}^{\mathrm{CI}}(\mathbf{H}_s) = \max_{\mathbf{H}_s,\, \mathbf{R}_{ss},\ \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss}) = P} \left(\sum_{i=1}^{K_s} \log_2 \left(1 + \left(\frac{P\Delta_s}{\sigma^2 K_a (A_{ii})_s} \left(1 + \frac{\sigma^2}{\Delta_s P} \sum_{j=1}^{K_a} (A_{jj})_s \right) - 1 \right)_{k+} \right) \right). \tag{3.30}$$

Considering only those row eliminations that result in non-zero power allocations, i.e., $\gamma_i > 0$, $\forall i$, and $K_a = K_s$, (3.30) is re-written as

$$R_{Sum}^{\mathrm{CI}}(\mathbf{H}_s) = \max_{\mathbf{H}_s,\, \mathbf{R}_{ss},\ \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss}) = P} \left(\underbrace{K_s \log_2 \left(\frac{\Delta_s P}{\sigma^2} + \sum_{i=1}^{K_s} (A_{ii})_s \right)}_{Term\ I} - \underbrace{\log_2 \prod_{i=1}^{K_s} (A_{ii})_s}_{Term\ II} - \underbrace{K_s \log_2 K_s}_{Term\ III} \right). \tag{3.31}$$

It appears from (3.31) that $\Delta_s P / \sigma^2$ dominates under high SNR where higher Δ_s values will help increase $R_{Sum}^{\mathrm{CI}}(\mathbf{H}_s)$. This leads to a 2-part question, (i) does $\Delta_s > \Delta_r$ exist for $S_s \subset S_r$, and (ii) if it does, would it result in a higher sum rate?

To answer the first part, note that $\mathbf{H}_r\mathbf{H}_r^H$ is positive definite Hermitian and as shown in Appendix A, the inclusion principle [67] applies so that

$$\lambda_{\min}(\mathbf{H}_r\mathbf{H}_r^H) \le \lambda_{\min}(\mathbf{H}_s\mathbf{H}_s^H) \le \lambda_2(\mathbf{H}_r\mathbf{H}_r^H) \le \lambda_2(\mathbf{H}_s\mathbf{H}_s^H) \le \cdots \le \lambda_{\max}(\mathbf{H}_s\mathbf{H}_s^H) \le \lambda_{\max}(\mathbf{H}_r\mathbf{H}_r^H), \tag{3.32}$$

where $\lambda_i(\mathbf{B})$ represents the eigenvalue λ_i of matrix $\mathbf{B}$. Note that $\Delta_s > \Delta_r$ can happen for example when the lower eigenvalues of $\mathbf{H}_r\mathbf{H}_r^H$ transit from $\lambda_i(\mathbf{H}_r\mathbf{H}_r^H) < 1$ to $\lambda_i(\mathbf{H}_s\mathbf{H}_s^H) \ge 1$ after a row elimination in $\mathbf{H}_r$. Supposing $\mathbf{H}_s$ is a sub-matrix of $\mathbf{H}_r$ after a one-row reduction, then the following lemma applies:

> *Lemma 3.1*: The largest determinant $(\Delta_s)_{\max} = \max\left(\det(\mathbf{H}_s\mathbf{H}_s^H)\right)$ among all one-row reduced sub-matrices $\mathbf{H}_s$ of $\mathbf{H}_r$ is equal to the cofactor that is associated with the maximum $1/b_i$ value among $\{1/b_i : i = 1, \cdots, K_r\}$, where $\{1/b_i\} = [(\mathbf{H}_r\mathbf{H}_r^H)^{-1}]_{i,i}$.

Proof: In relation to (3.28), $\Delta_s > \Delta_r$ occurs whenever a cofactor $(A_{ii})_r$ is *larger* than its associated determinant Δ_r, i.e., when there are one or more $1/b_i > 1$. From (3.28), it is clear that $\max(1/b_i)$ is associated with the largest cofactor $\max\left((A_{ii})_r\right)$. Let $\max\left((A_{ii})_r\right) = (A_{kk})_r$ where $(A_{kk})_r$ is associated with the sub-matrix $\mathbf{A}_{kk}$ that arises from eliminating row k and column k in $\mathbf{H}_r\mathbf{H}_r^H$, which corresponds to removing row k in $\mathbf{H}_r$ to give $\mathbf{H}_s$. This will result in a $\mathbf{H}_s\mathbf{H}_s^H$ that possesses the largest determinant $\max\left(\det(\mathbf{H}_s\mathbf{H}_s^H)\right) = (\Delta_s)_{\max} = (A_{kk})_r$, which is associated with $\max\left(1/b_i\right) = 1/b_k$. □

To answer if $(\Delta_s)_{\max} > \Delta_r$ would lead to a higher sum rate, (3.31) can be approximated at high SNR as

$$R_{Sum}^{\mathrm{CI}}(\mathbf{H}_s) \approx K_s \log_2\left(\Delta_s \frac{P}{\sigma^2}\right). \tag{3.33}$$

Equation (3.33) assumes $(\Delta_s P/\sigma^2) > \sum_{i=1}^{K_s}(A_{ii})_s$ and that dropping *Term II* is reasonable via the inequality of arithmetic- and geometric- means (AM-GM inequality), where:

$$(K_s)^{-1}\sum_{i=1}^{K_s}(A_{ii})_s \geq \left(\prod_{i=1}^{K_s}(A_{ii})_s\right)^{1/K_s} \Rightarrow$$

$$\left(K_s \log_2 \sum_{i=1}^{K_s}(A_{ii})_s - K_s \log_2 K_s\right) \geq \log_2 \prod_{i=1}^{K_s}(A_{ii})_s, \tag{3.34}$$

with equality only when $\{A_{ii}\}_s = \{c\}$, where c is a constant. Given a random matrix $\mathbf{H}_s$, $\{A_{ii}\}_s \neq \{c\}$, and applying (3.34) to (3.31), we see that (*Term I*)>(*Term II*) for all SNR levels because $K_s \log_2 \sum_{i=1}^{K_s}(A_{ii})_s > \log_2 \prod_{i=1}^{K_s}(A_{ii})_s$ even when $\text{SNR} = 0$. Given this, the following equation (3.35) must hold for sum rate increases to occur with $\mathbf{H}_s \subset \mathbf{H}_r$, i.e., $R_{Sum}^{\text{CI}}(\mathbf{H}_s) - R_{Sum}^{\text{CI}}(\mathbf{H}_r) > 0$,

$$\Delta_s \frac{P}{\sigma^2} > \left(\Delta_r \frac{P}{\sigma^2}\right)^{K_r/K_s}. \tag{3.35}$$

The condition in (3.35) is not difficult to meet. Firstly, the existence of highly correlated pairs in $\mathbf{H}_r$ would render $\Delta_r \ll 1$. The removal of one row in such pairs will result in $\Delta_s > \Delta_r$. Next, the removal of a user with very low channel gain will also result in higher Δ_s since

$$\left(\prod_{i=1}^{K_r}\lambda_i(\mathbf{Z}_r) = \Delta_r\right) \leq \prod_{i=1}^{K_r}(z_{ii})_r, \tag{3.36}$$

where $\mathbf{Z}_r = \mathbf{H}_r\mathbf{H}_r^H$ and $(z_{ii})_r$ are elements of the principal diagonal in $\mathbf{Z}_r$ and represent the channel gains. The power exponent in (3.35) also poses no problems since if both K_r and K_s are large, then $K_r/K_s \to 1$. However, this also implies that achieving the maximum sum rate with a small user subset is less likely, unless the user channel vectors are all highly correlated.

The statistical analysis of Wishart matrices $\mathbf{W}(n,m)$ given earlier in Section 3.1.1.1 also applies here, i.e., it gives an explanation of why $\Delta_s > \Delta_r$ can occur when

$n < m$, which reflects user or antenna de-selection. With $\mathbf{W}(n,m) = \mathbf{W}_r = \mathbf{H}_r\mathbf{H}_r^H$ when $n = m$ and $\mathbf{W}(n,m) = \mathbf{W}_s = \mathbf{H}_s\mathbf{H}_s^H$ when $n < m$, the possibility for $\Delta_s > \Delta_r$ can be easily seen from Figures 3.1 and 3.2 where the probability distributions depart from the lower-eigenvalue regions when $n < m$. It is important to note that the lower-eigenvalue range may be raised by orders of magnitude while the higher-eigenvalue range may remain largely in the same order of magnitude as n is decreased. It is under such conditions that $\Delta_s > \Delta_r$ can occur when $n < m$. Note again that this argument is statistically true even for the case where $\mathbf{H}_r = \mathbf{H}_w$, i.e., when $\mathbf{R}_r^{1/2} = \mathbf{I}_N$ where all users are geographically dispersed and uncorrelated.

When correlation is present among the users, we let $\mathbf{H}_r = \mathbf{R}_{rr}^{1/2}\mathbf{H}_{wr}$ and $\mathbf{H}_s = \mathbf{R}_{rs}^{1/2}\mathbf{H}_{ws}$ where $\mathbf{R}_{rs}^{1/2}$ is appropriately matched to $\mathbf{H}_s$. Then, $\det(\mathbf{W}_r) = \left(\det(\mathbf{R}_r^{1/2})\right)^2 \det(\mathbf{Z}_{wr})$ and $\det(\mathbf{W}_s) = \left(\det(\mathbf{R}_{rs}^{1/2})\right)^2 \det(\mathbf{Z}_{ws})$, where $\mathbf{Z}_{wr} = \mathbf{H}_{wr}\mathbf{H}_{wr}^H$ and $\mathbf{Z}_{ws} = \mathbf{H}_{ws}\mathbf{H}_{ws}^H$. Then

$$\Delta_r = \det(\mathbf{W}_r) = \prod\nolimits_{i=1}^{K_r} \lambda_i^2(\mathbf{R}_r^{1/2}) \prod\nolimits_{i=1}^{K_r} \lambda_i(\mathbf{Z}_r), \text{ and} \tag{3.37}$$

$$\Delta_s = \det(\mathbf{W}_s) = \prod\nolimits_{i=1}^{K_r} \lambda_i^2(\mathbf{R}_{rs}^{1/2}) \prod\nolimits_{i=1}^{K_r} \lambda_i(\mathbf{Z}_s). \tag{3.38}$$

It was shown earlier in Section 3.1.1.2 that the removal of antennas that are highly correlated with other elements will help reduce the eigenvalue spread in $\mathbf{R}_{rs}^{1/2}$, i.e., raising the lower eigenvalues and lowering the higher eigenvalues. When combined with the reduction in eigenvalue spread for $\mathbf{H}_{ws}\mathbf{H}_{ws}^H$ (as shown above), $\Delta_s > \Delta_r$ can occur when $n < m$.

An example using $\mathbf{H}_r \in \mathbb{C}^{16\text{x}16}$ is given in Figure 3.4 where the cumulative distribution functions of Δ_r and Δ_s are compared for 50000 channel realizations and two levels of inter-user correlation, viz., 0.0 and 0.5. For each channel realization, the value of Δ_s is obtained in three ways: (a) a row in $\mathbf{H}_r$ is randomly de-selected to yield $\mathbf{H}_s$, (b) de-selecting the row that yields the highest Δ_s among all one-row reduced sub-matrices, and

(c) de-selecting one or more rows to find the highest possible Δ_s among all sub-matrices of $\mathbf{H}_r$. As shown, the probability of higher Δ_s is improved when judicious user de-selection is done. The results also show that more than one de-selection may be needed for higher Δ_s values. Table 3.2 shows the percentage of time that a method of de-selection is better than the case without de-selection. As shown, the percentage of time that judicious de-selection is useful is high, especially when inter-user correlation is high.

The above analysis of the conditions under which $\Delta_s > \Delta_r$ can occur, provides a basis for user/antenna selection algorithms with the objective of TCIBF sum-rate maximization. It is recognized from the onset that any sum-rate maximization algorithm that is based solely on finding $\max(\Delta_s)$ is sub-optimal since the cofactor terms in (3.31) are neglected. The various approaches to algorithm development will be discussed in detail in Chapter 4.

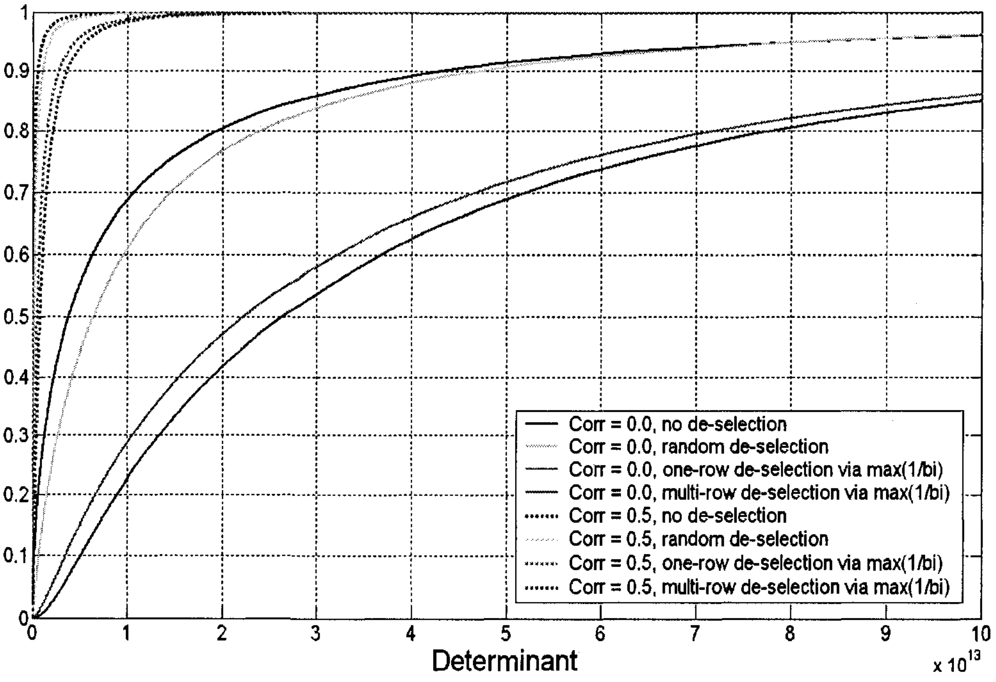


Figure 3.4. Cumulative distribution functions of Δ_r and Δ_s

Table 3.2: Percentage of time that a de-selection method is better than no de-selection

User De-selection Method	**Percentage of time better than:**	
	(a) No user de-selection	(b) Judicious one-user de-selection
Inter-user correlation = 0.0		
Random one-user de-selection	63.5%	n.a.
Judicious one-user de-selection	98.4%	n.a.
Judicious multi-user de-selection	98.4%	47.3%
Inter-user correlation = 0.5		
Random one-user de-selection	80.2%	n.a.
Judicious one-user de-selection	99.9%	n.a.
Judicious multi-user de-selection	99.9%	84.2%

3.1.3 Impact of User/Antenna De-selection on Ergodic TCIBF Sum Rate in Rayleigh Fading Channel

Beginning with a set S_r of $K_r = M$ users, the focus here is to examine the impact of subset selection (also referred to as user/antenna de-selection) on $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$, the ergodic TCIBF sum rate. Specifically, this refers to the case where a subset of users $S_s \subseteq S_r$ is varied in the range $|S_s| = K_s = 1,\cdots K_r$, under Rayleigh fading channel conditions while maintaining M transmit antennas. Using random subset selection, two lower bounds for $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$ will be presented, where the second bound is tighter than the first. It will be shown that the sum rates estimated using the second bound are close to the case when judicious subset selection is done. Importantly, it is shown that the maximum ergodic TCIBF sum rates usually occur at $K_s < M$ for practical SNR ranges and that random user/antenna de-selection provides fairly good performance compared to judicious user/antenna de-selection. It must be emphasized that this section focuses on subset selection from a set S_r of $K_r = M$ users, which must not be confused with user selection from a potential pool of $K > M$ users.

3.1.3.1 First Lower Bounding Approach for $\mathbb{E}\left(R_{Sum}^{\text{CI}}\right)$

As shown in the previous section, sum-rate maximization for TCIBF may be achieved for a set S_r of $K_r = M$ users by choosing a subset $S_s \subseteq S_r$ of $|S_s| = K_s = 1,\cdots K_r$ users. When judicious user subset selection (also referred to as judicious user/antenna de-selection) is done, the ergodic TCIBF sum-rate expression can be derived from (3.27) as

$$\mathbb{E}\left(R_{Sum}^{\text{CI}}\left(\mathbf{H}_s^{\text{opt}}\right)\right) = \mathbb{E}\left(\max_{\mathbf{H}_s,\mathbf{R}_{ss}} \sum\nolimits_{i=1}^{K_s} \log_2\left(1+\gamma_i/\sigma^2\right)\right), \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \le P. \qquad (3.39)$$

$$\overset{(a)}{\le} \max_{\mathbf{H}_s,\mathbf{R}_{ss}} \sum\nolimits_{i=1}^{K_s} \log_2\left(1+\mathbb{E}(\gamma_i)/\sigma^2\right), \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \le P, \qquad (3.40)$$

where $\mathbf{H}_s$ is the composite channel matrix of the subset S_s and the inequality (a) in (3.40) arises because log(•) is concave for which the following form of Jensen's inequality applies

$$\int_{-\infty}^{\infty} \varphi(x)\, f_X(x)\, dx \;\le\; \varphi\left(\int_{-\infty}^{\infty} x\, f_X(x)\, dx\right), \qquad (3.41)$$

where φ is a concave function and X is a random variable with a probability distribution function $f_X(x)$.

The γ_i values in (3.40) are influenced by λ_i, the eigenvalues of $\mathbf{W}(n,m) = \mathbf{W}_s = \mathbf{H}_s\mathbf{H}_s^H$, where n and m are the number of rows and columns in $\mathbf{H}_s$, respectively. In turn, the λ_i values are influenced by the number of users/antennas in S_s, which may be varied in the range $|S_s| = K_s = 1,\cdots K_r$. This changes $n = K_s$ in $\mathbf{W}(n,m) = \mathbf{W}_s = \mathbf{H}_s\mathbf{H}_s^H$ while keeping $m = M$. Note that varying n in $\mathbf{W}(n,m)$ amounts to random subset selection and this may be adopted to simplify the analysis. Adopting random subset selection instead of judicious subset selection, the ergodic sum rate may be examined with respect to user/antenna de-selection, that is, versus K_s. The ergodic TCIBF sum rate is then lower bounded as

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right) \geq \sum_{i=1}^{K_s} \log_2\left(1+\mathbb{E}(\gamma_i)/\sigma^2\right), \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \leq P. \tag{3.42}$$

Assuming that all K_s chosen users/antennas will be active on average, that is, $\mathbb{E}(\gamma_i) \neq 0\ \forall i = 1, \cdots K_s$ in (3.4), we put $K_a = (K_r - p + 1) = K_s$ in (3.4) and (3.5) so that

$$\mathbb{E}(\gamma_i) = \mathbb{E}\left(\mu_T b_i - \sigma^2\right), \ i = 1, \ldots\ldots, K_s, \tag{3.43}$$

$$\mu_T = \frac{P}{K_s}\left(1 + \frac{\sigma^2}{P}\sum_{i=1}^{K_s}\frac{1}{b_i}\right). \tag{3.44}$$

Hence (3.40) becomes

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right) \geq \sum_{i=1}^{K_s} \log_2\left(1+\mathbb{E}\left(\mu_T b_i - \sigma^2\right)/\sigma^2\right) \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \leq P$$

$$\geq \sum_{i=1}^{K_s} \log_2 \mathbb{E}\left(\mu_T b_i / \sigma^2\right) \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \leq P. \tag{3.45}$$

To simplify the analysis, we assume a high SNR regime and allow for equal power allocation among the users, i.e., $\mu_T \approx P/K_s$ in (3.44). This assumption further lower bounds $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right)$ so that

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right) \geq \sum_{i=1}^{K_s} \log_2\left(\frac{P}{\sigma^2 K_s}\mathbb{E}(b_i)\right) \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \leq P. \tag{3.46}$$

To find $\mathbb{E}(b_i)$, it is seen from (3.6) that

$$\mathbb{E}(1/b_i) = \mathbb{E}\left(\sum_{j=1}^{K_s} u_{ij} u_{ij}^* \lambda_j^{-1}\right), \quad i = 1, \ldots., K_s. \tag{3.47}$$

Now, it is known from [34] and references therein that the singular values of $\mathbf{H}_s$ and the left-hand singular vectors in $\mathbf{U}_s$ taken from $\operatorname{SVD}(\mathbf{H}_s) = \mathbf{U}_s \mathbf{\Sigma}_s \mathbf{V}_s^H$ are statistically independent when the entries of $\mathbf{H}_s$ are zero mean, centrally symmetrical, complex Gaussian (ZMCSCG) random variables. Hence

$$\begin{aligned}\mathbb{E}(1/b_i) &= \textstyle\sum_{j=1}^{K_s} \mathbb{E}(u_{ij}u_{ij}^*)\mathbb{E}(\lambda_j^{-1}), \quad i=1,....,K_s \\ &= \frac{1}{K_s}\textstyle\sum_{j=1}^{K_s} \mathbb{E}(\lambda_j^{-1}), \quad i=1,....,K_s \\ &= \frac{1}{K_s}\mathbb{E}\left(\textstyle\sum_{j=1}^{K_s} \lambda_j^{-1}\right), \quad i=1,....,K_s \end{aligned} \tag{3.48}$$

where $\mathbb{E}(u_{ij}u_{ij}^*) = 1/K_s \ \forall j$. It is noted from [70] that

$$\mathbb{E}\left(\text{trace}\left(\mathbf{W}_s^{-1}\right)\right) = \frac{K_s}{M-K_s} \quad \text{for } K_s < M. \tag{3.49}$$

Since $\text{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$ for a $(n \text{x} n)$ matrix $\mathbf{A}$ with eigenvalues $\lambda_1, \cdots, \lambda_n$ [67], (3.49) can be substituted into (3.48) to obtain

$$\mathbb{E}(1/b_i) = \frac{1}{K_s}\mathbb{E}\left(\text{trace}\left(\mathbf{W}_s^{-1}\right)\right) = \frac{1}{M-K_s}, \quad i=1,....,K_s. \tag{3.50}$$

The result in (3.50) is also shown in [65] and [66]. In [65], it is derived from the empirical distribution function of inverse eigenvalues of Wishart matrices in the Stieltjes domain. In [66], an expression similar to (3.49) was derived using the probability distribution function of the unordered eigenvalue of a Wishart matrix. Substituting (3.50) into (3.46)

$$\mathbb{E}\left(R_{Sum}^{\text{CI}}(K_s)\right) \geq K_s \log_2\left(\frac{P}{\sigma^2}\left(\frac{M}{K_s}-1\right)\right) \quad \text{s.t. } \text{tr}(\mathbf{R}_{ss}) \leq P. \tag{3.51}$$

A similar expression in (3.52) is derived in [59] by appealing to asymptotic parameters arising from large systems where $M \to \infty$ and $K_s \to \infty$ with M/K_s kept as a constant

$$\frac{M}{\beta}\log_2\left(1+\frac{P}{\sigma^2}(\beta-1)\right), \tag{3.52}$$

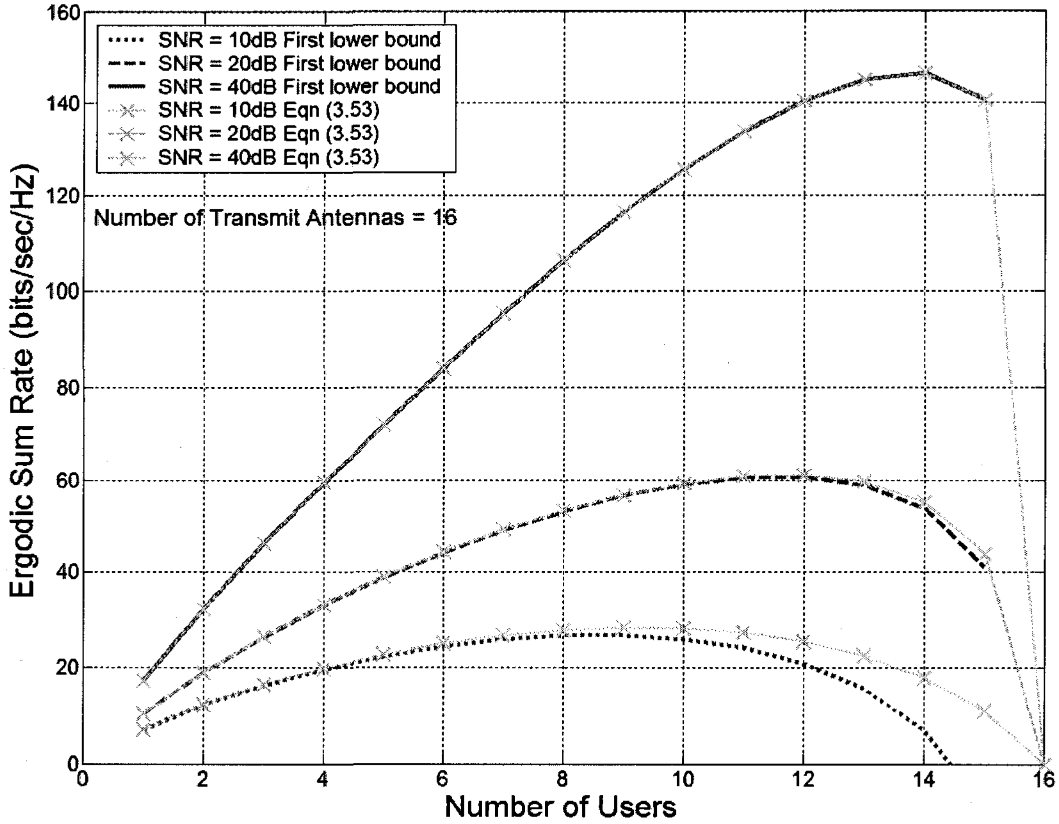


Figure 3.5. Ergodic TCIBF Sum Rate based on (3.51) and (3.53)

where $\beta = M / K$ and K is the number of active users, which is equal to K_s in our context. Using the notation in our context, (3.52) becomes

$$K_s \log_2\left(1+\frac{P}{\sigma^2}\left(\frac{M}{K_s}-1\right)\right), \tag{3.53}$$

which is almost identical to (3.51).

The derivation of (3.51) does not use an asymptotic approach and this makes it applicable for predicting the performance of systems of practical sizes. It also gives the assurance that (3.53) is also applicable to smaller user pools, as will be shown later. Note however that $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$ in (3.51) is undefined for the case where $M = K_s$, but should tend towards zero as shown by (3.53). To illustrate the behaviour of (3.51) and (3.53), an

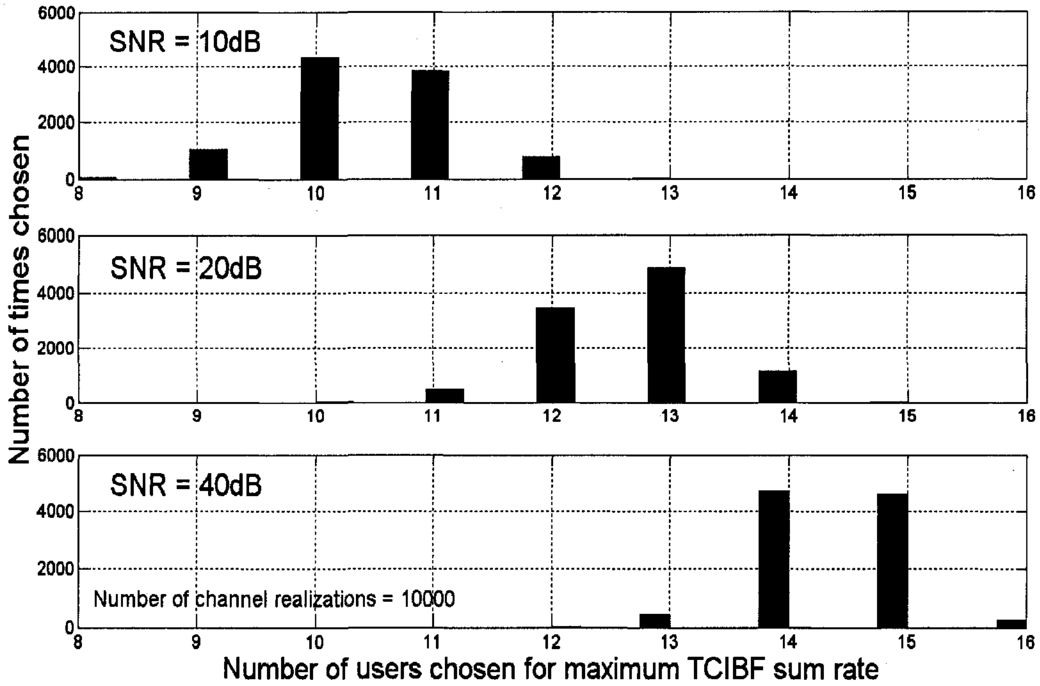


Figure 3.6. Histogram of number of users chosen for maximum TCIBF sum rate from numerical results

example of a TCIBF system with 16 users and three levels of $SNR = P/\sigma^2$ is given in Figure 3.5. The following points may be made from Figure 3.5:

a. The number of users/antennas that results in the maximum ergodic TCIBF sum rate is less than M, the number of transmit antennas.

b. A greater number of users/antennas must be de-selected when the SNR is low and vice versa.

This is compared to Monte Carlo simulation results where histograms of the best number of users for the highest ergodic sum rates are shown in Figure 3.6 for various SNR levels. To save computation time, the histogram is obtained using a sub-optimal user-selection algorithm. Waterfilling is done to maximize the sum rate. As shown, the simulation results coincide very well with the theoretical results given in Figure 3.5, that is, the number of users scheduled for maximum sum rate in Figure 3.6 are close to those in Figure 3.5. The ergodic sum rates obtained from simulation are shown in Table 3.3 and they show that the lower bound of (3.51) gives reasonable estimates, especially at high SNR.

Table 3.3: Simulated results for TCIBF ergodic sum rate with 16 users

SNR (dB)	**TCIBF Ergodic Sum Rate**		
	Simulated (bits/sec/Hz)	**Lower Bound in** (3.51)	
		(bits/sec/Hz)	**Difference (%)**
10dB	34.2	26.6	22.2%
20dB	70.8	60.7	14.3%
40dB	161.7	146.7	9.3%

The impact of user/antenna de-selection upon the ergodic TCIBF sum rate is therefore shown via the expression in (3.51). To help maximize the average throughput, the number users/antennas to be de-selected is higher when the SNR level is low and vice versa. To raise the probability of supporting *all* users in a chosen group while maintaining a high average throughput, a TCIBF system with high average SNR is needed. Assuming random subset selection, the maximum $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right)$ is therefore found by determining the best value of K_s when given a level of SNR and the number of transmit antennas M. This can be expressed as

$$\max\left(\mathbb{E}\left(R_{CI}(K_s)\right)\right) \geq \max_{K_s}\left(K_s \log_2\left(\frac{P}{\sigma^2}\left(\frac{M}{K_s}-1\right)\right)\right) \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \leq P. \qquad (3.54)$$

Since the expression for $\max\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right)$ in (3.54) is concave in K_s, a derivative can be taken to give the K_s to SNR relationship that yields the highest ergodic sum rates:

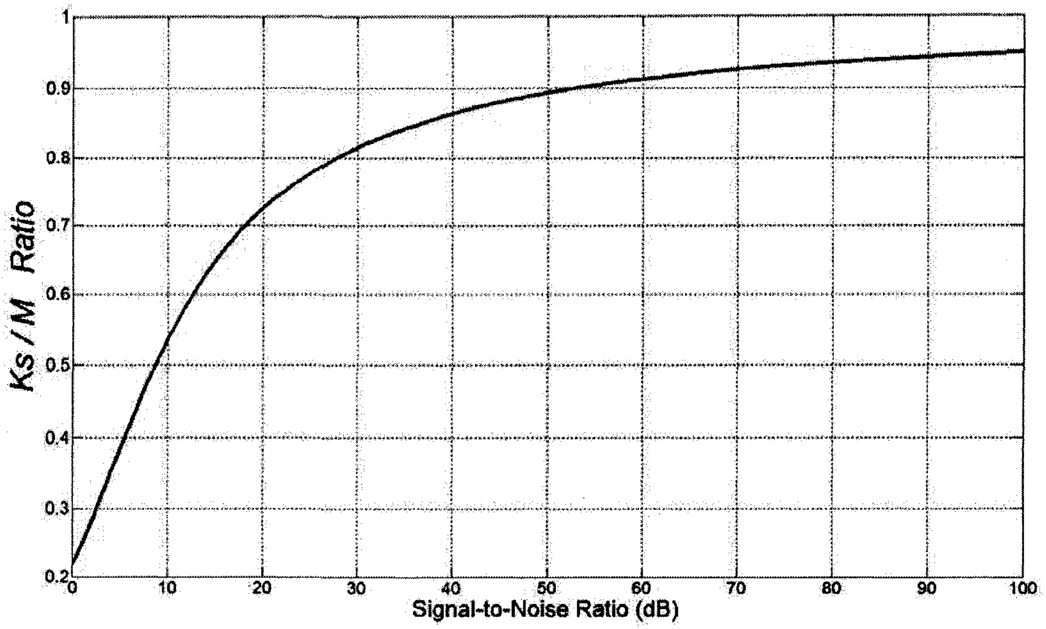


Figure 3.7. Plot of K_s / M ratio to SNR for maximum TCIBF sum rate

$$\left(\frac{K_s / M}{1-K_s / M}\right)\exp\left(\frac{1}{1-K_s / M}\right) = SNR. \tag{3.55}$$

The relationship is plotted in Figure 3.7, which corresponds to the peaks in Figure 3.5. Equation (3.55) is useful for estimating the average percentage of users served for any given SNR value in TCIBF systems. An expression similar to (3.55) can also be found in [59]. It is seen from Figure 3.7 that achieving a load factor of $K_s / M \to 1$ occurs only at very high SNR levels. This stringent requirement may be lowered in practice when the potential user pool size K is larger than M and the existence of multi-user diversity may be exploited via judicious user selection (discussed below).

3.1.3.2 A Tighter Second Lower Bounding Approach for $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$

A tighter lower bound is obtained by dropping the use of Jensen's inequality in (3.40). Maintaining the high SNR and random subset selection regime, then

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right) \geq \mathbb{E}\left(\sum_{i=1}^{K_s} \log_2\left(\frac{P}{\sigma^2 K_s} b_i\right)\right) \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P. \tag{3.56}$$

Considering TCIBF from the projection viewpoint as in (3.7), TCIBF may be considered as a special case of block diagonalization where all user terminals have only one antenna. Each user j will experience a projected channel $\mathbf{H}_{pj}$ with Wishart matrices $\mathbf{W}_{Pj}(n,m) = \mathbf{H}_{Pj}\mathbf{H}_{Pj}^H$ that have unordered eigenvalues λ with PDF $f_\lambda(\lambda) = n^{-1}\sum_{i=1}^{n} \varphi_i(\lambda)^2 \lambda^{m-n} e^{-\lambda}$. In this case, the values of n and m may be given using (3.103) below

$$\left.\begin{array}{c} n = 1 \\ m = \left(M + q - (K_r - 1)\right) \end{array}\right\} \tag{3.57}$$

It is important to note from (3.57) that the parameter m (the number of column vectors in $\mathbf{H}_{pj}$) is affected by the parameter q, which represents the number of users/antennas that are de-selected. Note that $K_r \leq M$ represents the initial number of users within the user set S_r. Note also that varying the parameter m implies random user/antenna de-selection. Using this approach, the ergodic TCIBF sum rate is then

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right) \geq \mathbb{E}\left(\sum_{i=1}^{K_s} \log_2\left(\frac{P}{\sigma^2 K_s} \lambda\right)\right) \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P, \text{ or}$$

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}(K_s)\right) \geq \int_0^\infty K_s \log_2\left(\frac{P}{\sigma^2 K_s} \lambda\right) f_\lambda(\lambda)\, d\lambda \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P. \tag{3.58}$$

Note that $K_s = K_r - q$ in (3.57).

Figure 3.8 compares ergodic TCIBF sum rates arising from the first lower bound of (3.51) with those arising from the second lower bound of (3.58). Table 3.4 follows up on Table 3.3 by comparing the simulated results with both bounds. As shown, the lower bound in (3.58) is tighter than that in (3.51). We note again that finding $\max\left(\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)\right)$ involves finding the optimum value of $K_s = K_r - q$ and (3.58) may be expressed as

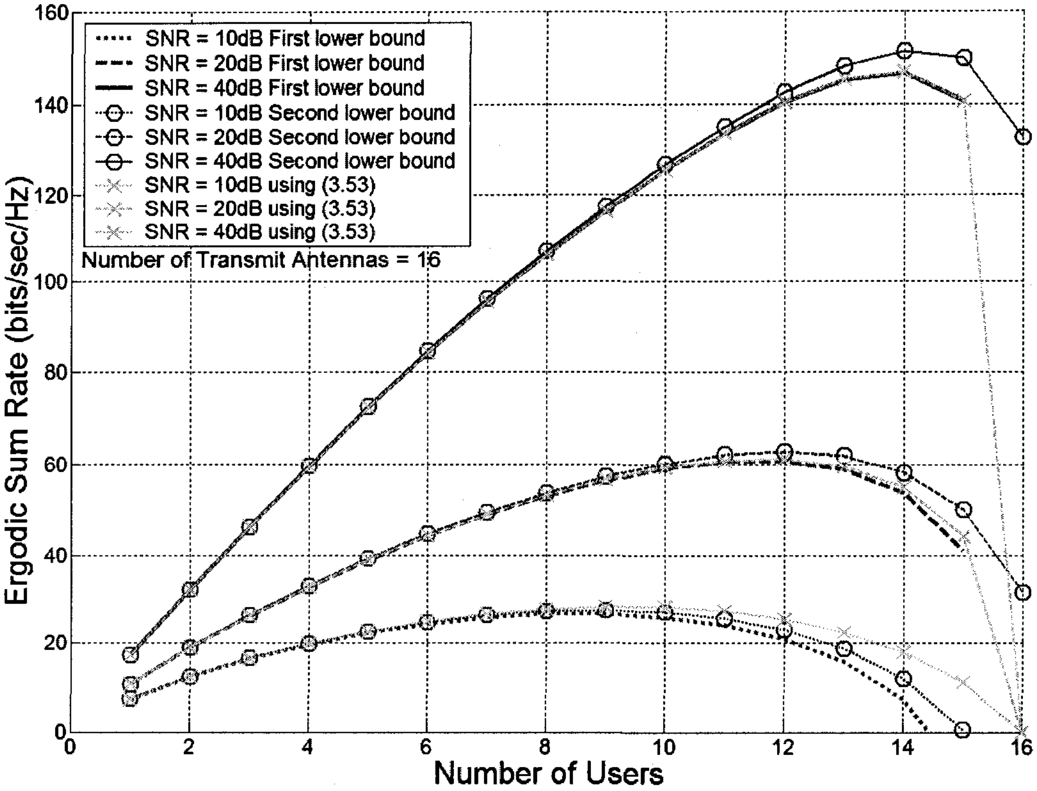


Figure 3.8. Comparing the two lower bounds in (3.51) and (3.58) with (3.53) by [59]

$$\max\left(\mathbb{E}\left(R^{\text{CI}}_{Sum}(K_s)\right)\right) \geq \max_{K_s} \int_0^\infty K_s \log_2\left(\frac{P}{\sigma^2 K_s}\lambda\right) f_\lambda(\lambda)\ d\lambda \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{dd}) \leq P. \quad (3.59)$$

Note again that the maximization in (3.59) with respect to K_s implies random selection.

In Figure 3.9, the two bounds of (3.51) and (3.58) are compared to numerical results. Two sets of numerical results are presented for the ergodic TCIBF sum rate and each set is obtained using 2000 channel realizations. In the first set, de-selection of the users/antennas is done randomly while judicious de-selection is done for the second set using a sub-optimal selection algorithm from Chapter 4.

As shown in Figure 3.9, the second lower bound of (3.58) is tight when compared to the numerical results where random de-selection is done. It is also fairly tight when compared against the case where judicious de-selection is done (see Table 3.4). Its performance is better than that of (3.53) [59] at high SNR levels (from 20dB onwards). The bound of (3.58) is therefore useful for design assessments where initial estimates of (a) the ergodic sum rates of TCIBF systems and (b) the optimum number of users for the best sum rates, are needed. Monte Carlo simulations may then follow for more accurate results when a system size is decided. As a side note, the performance of judicious de-selection will be significantly better than random de-selection under heterogeneous channel conditions.

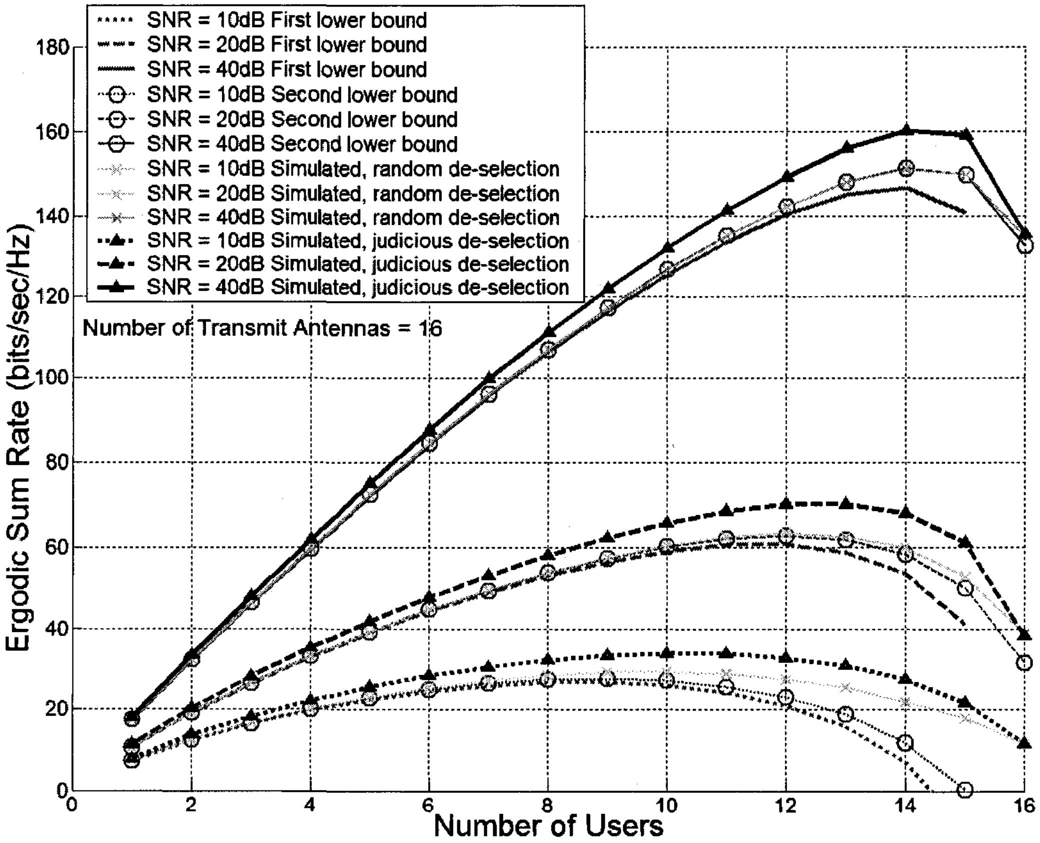


Figure 3.9. Comparing the two lower bounds in (3.51) and (3.58) with numerical results

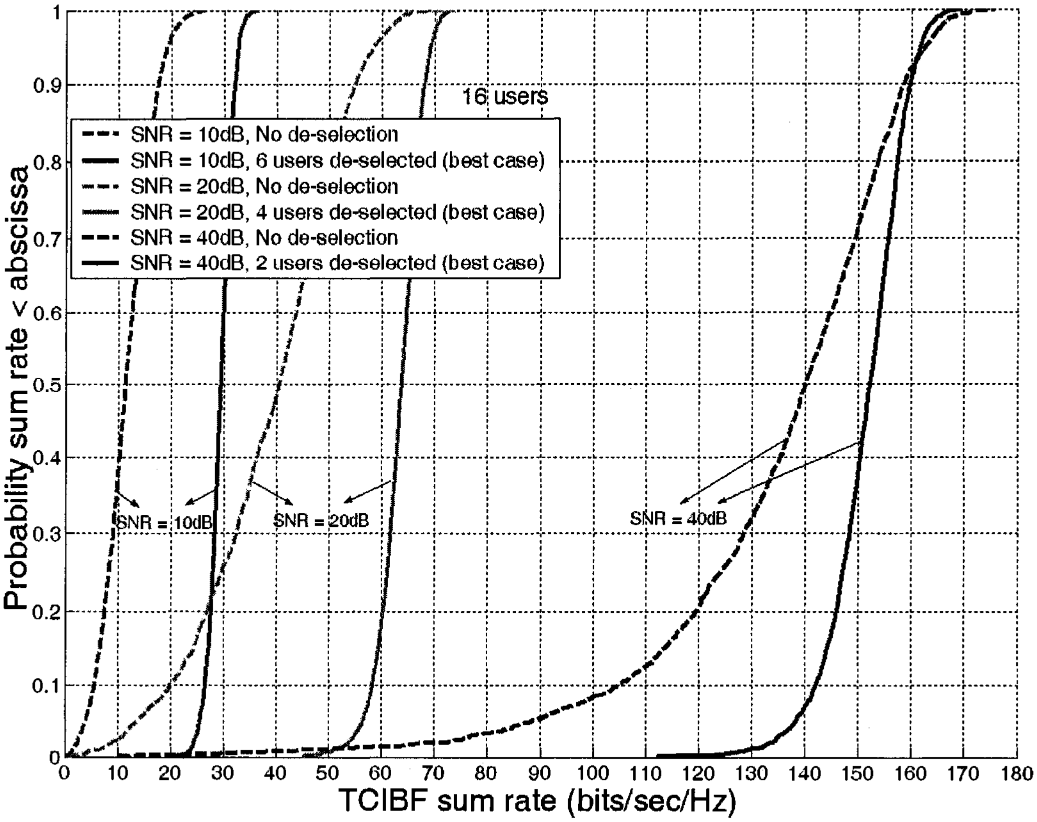


Figure 3.10. Numerical results of outage sum rates with and without user/antenna de-selection

When the transmitter does not have adaptive modulation, the outage capacity is the more appropriate measure of performance. The impact of random user/antenna de-selection on the outage sum rate is shown in Figure 3.10. A comparison is made between no de-selection and where de-selection is done to achieve the best sum rate (see Figure 3.9 for best points). As shown in Figure 3.10 and Table 3.5, improvements to the 10%-outage sum rate are higher in percentage terms compared to improvements to the average sum rate. The improvements are especially significant when the SNR levels are low.

Table 3.4: Simulated results compared with the two lower bounds.

SNR (dB)	**TCIBF Ergodic Sum Rate (bits/sec/Hz)**				
	Simulated (bits/sec/Hz)	**1st Lower Bound in (3.51)**		**2nd Lower Bound in (3.58)**	
		(bits/sec/Hz)	**Difference (%)**	**(bits/sec/Hz)**	**Difference (%)**
10dB	34.2	26.6	22.2%	27.5	19.6%
20dB	70.8	60.7	14.3%	62.8	11.3%
40dB	161.7	146.7	9.3%	151.4	6.4%

Table 3.5: Outage sum rate comparison for case without de-selection and case with de-selection for the best sum rates.

SNR (dB)	**TCIBF 10%-Outage Sum Rate (Monte Carlo simulations)**		**Percentage Improvement (%)**
	No de-selection (bits/sec/Hz)	**Best number de-selected (bits/sec/Hz)**	
10dB	5	26	420%
20dB	20	58	188%
40dB	115	143	24%

3.1.4 Impact of User Selection on Ergodic TCIBF Sum Rate when $K > M$

The main intention here is to demonstrate the joint impact of (a) user selection, (b) user de-selection (also referred to as user subset selection in the previous section) and (c) different SNR levels, upon the ergodic TCIBF sum rate when the user pool is large ($K > M$). An upper bound will be derived for the purpose of demonstrating the joint impact of these mechanisms. To simplify the task at this stage, the bound will be based on a number of assumptions and the sum rates provided are only good as ballpark estimates.

Let S represent a potential pool of $K > M$ users, that is, there are more single-antenna terminals than the number of base-station-transmit antennas. To employ TCIBF, it is necessary to choose a subset S_r with $K_r = |S_r| \leq M$ users to meet the TCIBF pre-coding constraint. This is commonly referred to in literature as user selection. User selection is potentially beneficial as it has been shown that transmit zero-forcing beamforming approaches the DPC sum capacity asymptotically when the number of users is high [26]. This is due to the presence of multi-user diversity in large user pools and judicious user selection is needed for sum-rate maximization. In other words, a large user pool presents the transmitter with a higher chance of choosing a group of users with high channel magnitudes whose channel directions are matched to the zero-forcing beam directions [26]. This lowers the likelihood of signal attenuation in TCIBF, which occurs when channel inversion is performed on a chosen user subset whose composite channel is poorly conditioned.

When the user pool is large, there are many subsets of S_r that can be considered. The optimal strategy is to perform an exhaustive search of all possible combinations of users to find the optimum subset S_r^{opt}. When the user pool size is very large, the probability that $|S_r^{\text{opt}}| = M$ will be high, where M is the number of base station transmit antennas. As shown in Section 3.1.3 however, $|S_r^{\text{opt}}| < M$ may result when the potential user pool is not much larger than M and/or the channel conditions are poor. Note that the act of selecting any subset with $|S_r^{\text{opt}}| < M$ may be viewed as a *joint* action that comprises the initial selection of S_r with $|S_r| = M$ followed by user/antenna de-selection from within S_r to result in S_r^{opt}, with $|S_r^{\text{opt}}| \leq M$. In practice, this two-step process is true of low-complexity user-selection algorithms that avoid sum-rate evaluations. Note that sum-rate evaluations are computationally intensive since repeated TCIBF pre-coding is involved. These algorithms are normally used to pick $|S_r| = K_r = M$ users since there is no other stopping criterion. The user-selection exercise is then followed by user/antenna de-selection for sum-rate maximization. For consistency with the notation in Section 3.1.3, the resulting subset is labeled as $S_s \subseteq S_r$ where $|S_s| = K_s = \{1,\cdots,K_r\}$. To reflect the two selection mechanisms, S_r^{opt} is used to represent the best subset of $\left|S_r^{\text{opt}}\right| = K_r = M$ users from the potential user pool of K users whereas the optimal subset is denoted as $S_s^{\text{opt}} \subseteq S_r^{\text{opt}}$, where $\left|S_s^{\text{opt}}\right| = K_s = \{1,\cdots,M\}$.

To begin, the effect of user/antenna de-selection is considered in isolation. Suppose that a user set S_r where $|S_r| = K_r = M$ is given without any regard to the effects of user selection. The ergodic TCIBF sum rate after *judicious* user/antenna de-selection to choose $S_s^{\text{opt}} \subseteq S_r$ may be derived from (3.39) as

$$\mathbb{E}\left(R_{Sum}^{\text{CI}}\left(\mathbf{H}_s^{\text{opt}}\right)\right) = \mathbb{E}\left(\max_{\mathbf{H}_S,\mathbf{R}_{SS}} \sum\nolimits_{i=1}^{K_s} \log_2\left(1+\gamma_i/\sigma^2\right)\right), \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \le P$$

$$\mathbb{E}\left(R_{Sum}^{\text{CI}}\left(\mathbf{H}_s^{\text{opt}}\right)\right) = \int_0^\infty r\, f_{R_{sum}^{\text{CI}}}(r)\, dr, \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{ss}) \le P, \tag{3.60}$$

where $f_{R_{sum}^{\text{CI}}}(r)$ is the probability density function of $R_{Sum}^{\text{CI}}\left(\mathbf{H}_s^{\text{opt}}\right)$. Note that $f_{R_{sum}^{\text{CI}}}(r)$ is parameterized by the number of users in the subset S_s^{opt}, which represents the user de-selection mechanism.

Next, the impact of user selection may be jointly reflected using methods from order statistics and a brief outline is given as follows. Suppose that L subsets are drawn from the user pool where each subset has a sum rate of $R_{Sum,l}^{\text{CI}}$ and a set of rates $\{R_{Sum,1}^{\text{CI}}, R_{Sum,2}^{\text{CI}}, \cdots, R_{Sum,L}^{\text{CI}}\}$ is obtained. This set of rates may then be ordered and designated as $R_{Sum,(1):L}^{\text{CI}} < R_{Sum,(2):L}^{\text{CI}} < \cdots < R_{Sum,(L):L}^{\text{CI}}$. It is well known from order statistics theory that $f_{R_{sum,(l):L}^{\text{CI}}}(r)$ the probability distribution function of $R_{Sum,(l):L}^{\text{CI}}$ is given by [62]

$$f_{R_{sum,(l):L}^{\text{CI}}}(r) = \frac{L!}{(l-1)!(L-l)!} F_{R_{sum}^{\text{CI}}}(r)^{l-1}\left(1 - F_{R_{sum}^{\text{CI}}}(r)\right)^{L-l} f_{R_{sum}^{\text{CI}}}(r), \tag{3.61}$$

where $f_{R_{sum}^{\mathrm{CI}}}(r)$ and $F_{R_{sum}^{\mathrm{CI}}}(r)$ are the probability density function and cumulative distribution function of $R_{Sum}^{\mathrm{CI}}(\mathbf{H}_{s,l})$, respectively. The best subset $S_{s,l}^{\max}$ from among the L subsets is the one associated with the highest sum rate $R_{Sum,(L):L}^{\mathrm{CI}}$ and its PDF $f_{R_{sum,(L):L}^{\mathrm{CI}}}(r)$ is simply derived from (3.61) as

$$f_{R_{sum,(L):L}^{\mathrm{CI}}}(r) = LF_{R_{sum}^{\mathrm{CI}}}(r)^{L-1} f_{R_{sum}^{\mathrm{CI}}}(r). \qquad (3.62)$$

Using (3.62) with (3.60), the ergodic TCIBF sum rate expression that captures the joint effect of judicious user selection and subset selection may be expressed as

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\left(\mathbf{H}_{s,l}\right)_{\max}\right) = \int_0^\infty r\, L\, F_{R_{sum}^{\mathrm{CI}}}(r)^{L-1} f_{R_{sum}^{\mathrm{CI}}}(r)\, dr, \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{ss,l}) \leq P. \qquad (3.63)$$

To make use of the order-statistics method described above, some assumptions relating to subset selection must be made. To re-iterate, the optimal strategy involves an exhaustive search of all possible combinations of users to find the optimum subset S_s^{opt}. To reduce complexity, practical selection algorithms usually reduce the number of combinations considered by proceeding in an incremental or decremental manner. Note that in both cases (optimal selection or via algorithms), each user may be evaluated more than once. To suit the order-statistics model however, the aforementioned user selection process is approximated by drawing L subsets $\{S_{s,l},\ l = 1,\cdots,L\}$ simultaneously for consideration. This means that each user is considered only once because each user appears in only one combination. All subset sizes are assumed to be equal in each draw of L subsets and contain $|S_{s,l}| = K_s$ users, where $K_s = 1,\cdots,M \quad \forall l = 1,\cdots,L$ and $\max(K_s) = M$. Next, assuming the existence of very large user pools where $K \to \infty$, each subset $S_{s,l}$ can then be considered independent from all other subsets $S_{s,k}$, where $l \neq k$.

The solution to (3.63) depends on the availability of $f_{R_{sum}^{\mathrm{CI}}}(r)$ and $F_{R_{sum}^{\mathrm{CI}}}(r)$ and efforts are still on-going at the time of writing to solve them. For the purpose of demonstrating the joint effects of user selection and user/antenna de-selection, coupled with parameters like SNR and M, an approximation to (3.63) via bounding techniques is taken.

To begin, the approach in Section 3.1.3.2 is adopted and random user/antenna de-selection is first assumed within each of the L subsets. Each user j within subset l will have a projected channel matrix $\mathbf{H}_{pj,l}$ and Wishart matrices $\mathbf{W}_{Pj,l}(n_l, m_l) = \mathbf{H}_{Pj,l}\mathbf{H}_{Pj,l}^H$ that have unordered eigenvalues $\lambda_{l:L}$ with PDF $f_{\lambda_{l:L}}(\lambda) = n_l^{-1}\sum_{i=1}^{n_l}\varphi_i(\lambda)^2\lambda^{m_l-n_l}e^{-\lambda}$ and associated CDF $F_{\lambda_{l:L}}(\lambda) = \int_{-\infty}^{\lambda} f_{\lambda_{l:L}}(\lambda)\,d\lambda$ where

$$\left.\begin{aligned} n_l &= 1 \\ m_l &= \left(M + q_l - (K_{s,l} - 1)\right) \end{aligned}\right\} \quad \forall l = 1, \cdots, L. \tag{3.64}$$

In this case, the parameter q_l in (3.64) follows the approach in Section 3.1.3.2 and represents the number of users/antennas that are de-selected to form $S_{s,l}$. Since we assume equal size for all $S_{s,l}$ for ease of analysis even when RAS is done, that is, the same for all L sets, then $q_l = q$ and $|S_{s,l}| = K_{s,l} = M$ before de-selection so that (3.64) becomes

$$\left.\begin{aligned} n_l &= 1 \\ m_l &= q + 1 \end{aligned}\right\} \quad \forall l = 1, \cdots, L. \tag{3.65}$$

The next approximation comes from the assumption that all eigenvalues within each of the L subsets take on the same instantaneous value $\lambda_{l:L}$ during each draw. The representative eigenvalues from each of the L chosen subsets may then be sorted as $\lambda_{(l):L}$ where $\lambda_{(1):L} < \lambda_{(2):L} < \cdots < \lambda_{(L):L}$. The best subset from among them is the one associated with the maximum eigenvalue $\lambda_{\max} = \lambda_{(L):L}$. This assumption gives rise to an upper bound since all eigenvalues within the chosen subset has the same instantaneous value of $\lambda_{\max} = \lambda_{(L):L}$. It is well known from order statistics theory that $f_{\lambda_{(l):L}}(\lambda)$ the probability distribution function of $\lambda_{(l):L}$ is [62]

$$f_{\lambda_{(l):L}}(\lambda) = \frac{L!}{(l-1)!(L-l)!} F_{\lambda_{l:L}}(\lambda)^{l-1}\left(1 - F_{\lambda_{l:L}}(\lambda)\right)^{L-l} f_{\lambda_{l:L}}(\lambda). \tag{3.66}$$

Equation (3.66) may then be applied to (3.58) that results in an upper bound expression for the ergodic TCIBF sum rate

$$\mathbb{E}\left(R^{\mathrm{CI}}_{Sum}\left(K_s\right)\right) \le \int_{-\infty}^{\infty} K_s \log_2\left(\frac{P\lambda}{\sigma^2 K_s}\right) f_{\lambda_{(l):L}}(\lambda)\ d\lambda \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \le P. \tag{3.67}$$

The sum rate expression in (3.67) is parameterized by:

(a) M , the number of base station transmit antennas,

(b) L , the number of user subsets drawn for consideration from the user pool

(c) $q_l = q$, the number of users/antennas de-selected from each subset and

(d) SNR, the signal-to-noise ratio..

The maximum $\mathbb{E}\left(R^{\mathrm{CI}}_{Sum}\left(K_s\right)\right)$ given L and M involves finding the best value of $K_s = M - q$. Hence

$$\mathbb{E}\left(R^{\mathrm{CI}}_{Sum}\left(K_s\right)\right)_{\max}\Big|_{L,M} \ge \max_{K_s} \int_{-\infty}^{\infty} K_s \log_2\left(\frac{P\lambda}{\sigma^2 K_s}\right) L\ F_{\lambda_{l:L}}(\lambda)^{L-1} f_{\lambda_{l:L}}(\lambda)\ d\lambda \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \le P. \tag{3.68}$$

Figure 3.11 shows the behavior of $\mathbb{E}\left(R^{\mathrm{CI}}_{Sum}\left(K_s\right)\right)$ using (3.67) under different L, SNR and K_s. The behavior without user subset selection, i.e., when $L = 1$, is identical to those produced by the second lower bound in Figure 3.8. When $L > 1$, the results show better ergodic sum rates, which reflect the benefits of multi-user diversity (MUD) when coupled with user/antenna de-selection (done randomly in this case). The effect of varying K_s within the chosen subset is still significant even when $L > 1$. This means user/antenna de-selection must still be done when searching for the best user subset, especially when the user pool is small. The results also show that the benefits of MUD harnessed via user selection help in the scheduling of more users (i.e., less users de-selected) from within the chosen subset, especially at low SNR. To highlight the case when SNR = 10dB, the

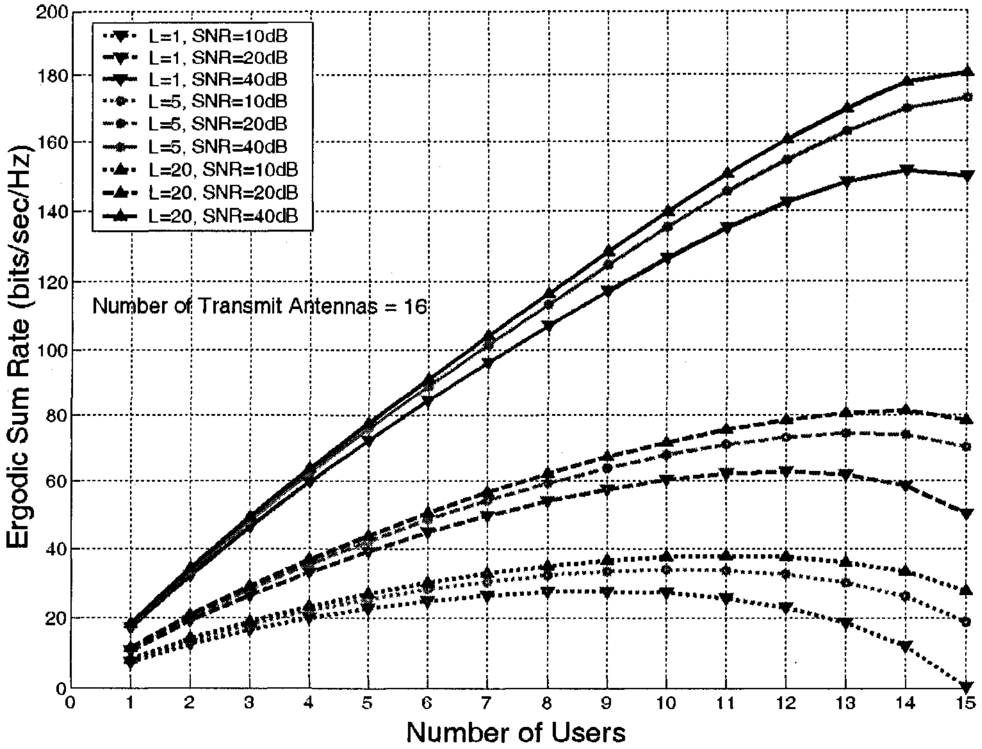


Figure 3.11. Ergodic TCIBF Sum Rate based on (3.67)

number of users scheduled within a chosen group went from 9 to 11 when *L*, the number of subsets drawn, went from 1 to 20, respectively. This is to be expected on average since the chances of finding a subset with users that are close to being orthogonal is higher when the potential user pool is large. This translates to higher eigenvalues on average and higher ergodic sum rates. Conversely, more users/antennas must be de-selected to give adequate channel gains when the potential user pool is small, i.e., *L* is small.

In Figures 3.12 and 3.13, the upper bound of (3.67) is compared against numerical results where user/antenna de-selection is done in a random manner and in a judicious manner, respectively. As shown, (3.67) does upper bound even for the judicious user/antenna de-selection case.

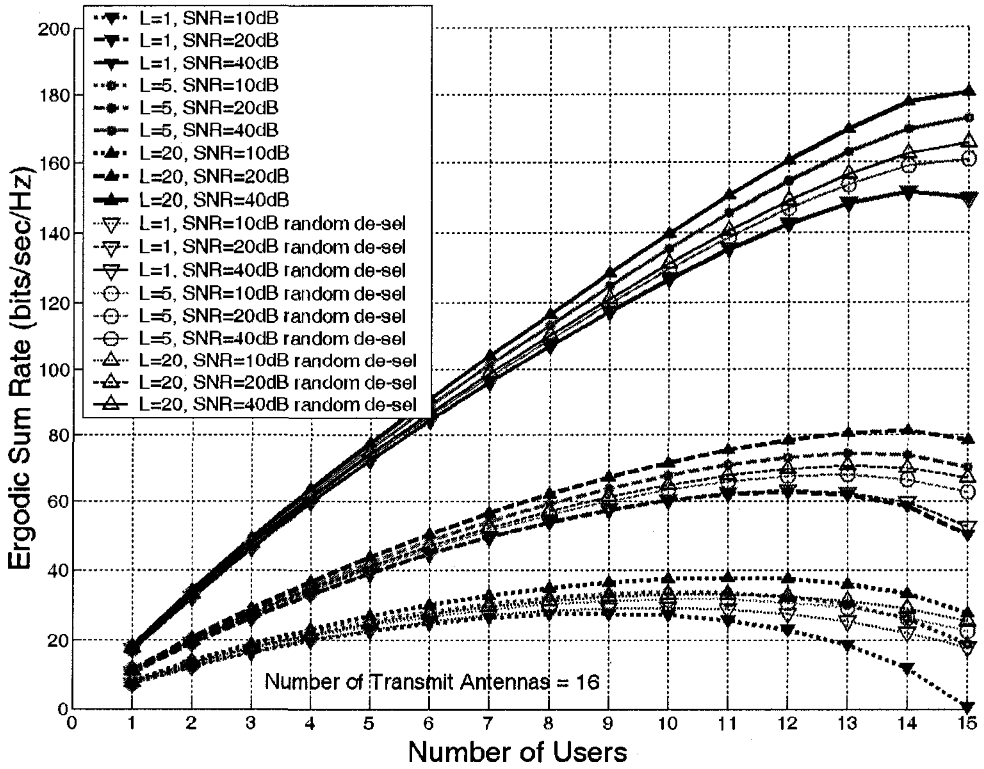


Figure 3.12. Comparing the upper bound in (3.67) with simulated results using random user/antenna de-selection

3.1.5 Sum Rate Scaling Behavior of TCIBF in Large User Pools

It has been shown in [18] that $\mathbb{E}\left(R_{sum}^{\text{DPC}}\right)$, the expected dirty paper coding (DPC) sum rate, scales as

$$\lim_{K\to\infty}\frac{\mathbb{E}\left(R_{sum}^{\text{DPC}}\right)}{M\log\log KN}=1, \tag{3.69}$$

where M is the number of transmit antennas, K is the user pool size and N is the number of antennas per user terminal. It is also shown in [18] that $\mathbb{E}\left(R_{sum}^{\text{BF}}\right)$ the expected beamforming sum rate scales as that for DPC, that is

$$\lim_{K\to\infty}\frac{\mathbb{E}\left(R_{sum}^{\text{BF}}\right)}{M\log\log KN}=1, \tag{3.70}$$

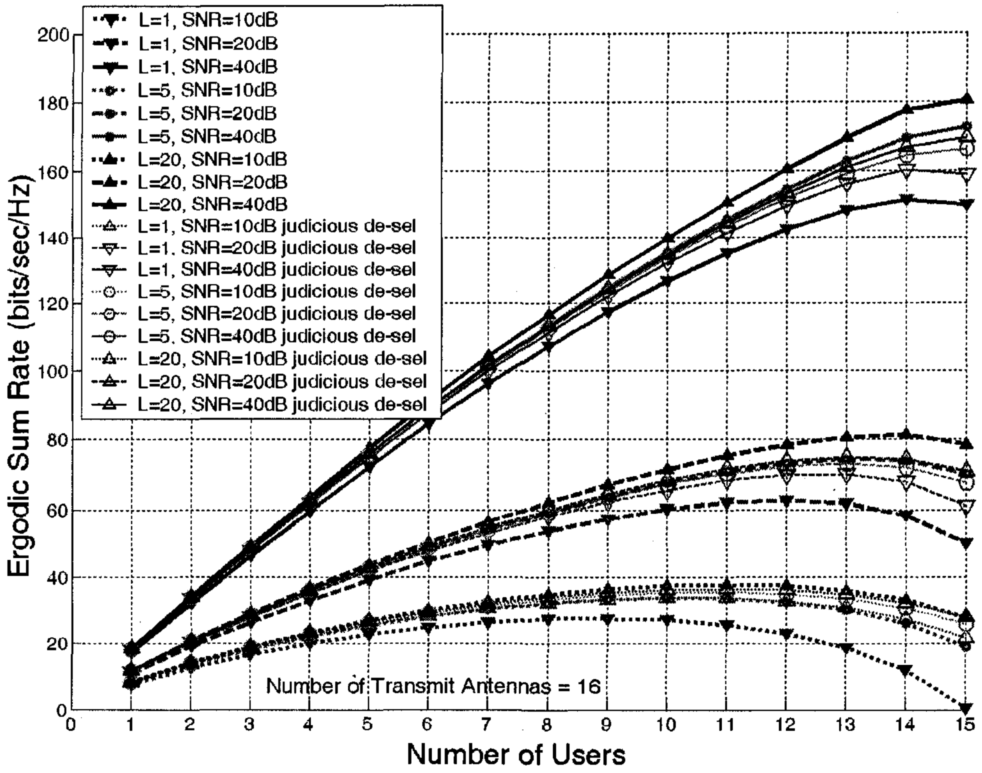


Figure 3.13. Comparing the upper bound in (3.67) with simulated results using judicious user/antenna de-selection

Next, it has been shown in [26] that the zero-forcing beamforming (ZFBF) strategy, while generally sub-optimal, can achieve the same asymptotic sum capacity as that of DPC, when K the number of users goes to infinity. The approach in [26] was based on a selection algorithm known as semi-orthogonal user selection. In this section, it will be shown that transmit channel-inversion beamforming (TCIBF) can also scale as $M \log\log K$ as $K \rightarrow \infty$ (note that $N = 1$ for TCIBF). The simplified method here is based mainly on extreme-value theory and on [71], which dealt with asymptotic throughput analysis for channel-aware scheduling in time-division multiplexing (TDM) systems. An outline of the approach is given next.

A lower bound is first established by excluding the user/antenna de-selection process after selecting the best user subset from the potential user pool of size K. Recall that excluding the user/antenna de-selection process will result in a lower ergodic sum rate than otherwise possible. The associated Wishart matrix is then of full rank, that is, **W**(m,m). The sum rate is further lower bounded by applying the lowest channel gain, as represented by $\lambda_{\min}$, the minimum eigenvalue of **W**(m,m), to all users. Using extreme-order statistics, the ergodic TCIBF sum rate is evaluated when $K \to \infty$, where $f_{\lambda_{\min}}$, the probability density function of $\lambda_{\min}$, is used. Note that if scaling with M is true for the ergodic TCIBF sum rate ($\mathbb{E}\left(R_{sum}^{\mathrm{CI}}\right)$), then, the ergodic sum rate for systems utilizing block diagonalization (BD) will also scale with M because it is lower bounded by $\mathbb{E}\left(R_{sum}^{\mathrm{CI}}\right)$.

From [69], the PDF and CDF of the minimum eigenvalue $\lambda_{\min}$ for **W**(m,m) are

$$f_{\lambda_{\min}}(\lambda) = Me^{-\lambda M}. \tag{3.71}$$

$$F_{\lambda_{\min}}(\lambda) = 1 - e^{-\lambda M}. \tag{3.72}$$

The ergodic TCIBF sum rate expression in (3.58) can be used as a starting point for applying $\lambda_{\min}$

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right) \geq \mathbb{E}\left(M \log\left(\frac{P}{\sigma^2 M}\lambda_{\min}\right)\right) \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{dd}) \leq P, \tag{3.73}$$

where $K_s = M$ is set to reflect the exclusion of the user/antenna de-selection process after selecting the best user subset from the potential user pool. Noting that R_{Sum}^{CI} is a function of $\lambda_{\min}$, we write the following for ease of reference

$$R_{Sum}^{\mathrm{CI}} = T(\lambda_{\min}) = M\log_2\left(\rho\lambda_{\min}\right), \tag{3.74}$$

where $\rho = \dfrac{P}{\sigma^2 M}$, which is the SNR per transmit antenna. For the purpose of scaling with M, the value of ρ will be fixed as M is scaled, i.e., the total power P will be adjusted so that ρ is set constant at the desired value.

The evaluation of $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$ as $K \to \infty$ using extreme-value theory requires $F_{R_{Sum}^{\mathrm{CI}}}(r)$, the CDF of R_{Sum}^{CI}, which is complicated to evaluate. To simplify the evaluation, we note that (3.73) is an approximation to Equation (1) in [71] and a theorem from [71] known as the "limiting throughput distribution (LTD)" theorem can be applied to solve for (3.73). The LTD theorem allows for the evaluation of $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$ without the need to check $F_{R_{Sum}^{\mathrm{CI}}}(r)$ directly. Instead, if it can be shown that

$$\lim_{\lambda \to \infty} \frac{d}{d\lambda}\left[\frac{1-F_{\lambda_{\min}}(\lambda)}{f_{\lambda_{\min}}(\lambda)}\right]=0, \tag{3.75}$$

then, $F_{R_{Sum}^{\mathrm{CI}}}(r)$ the cumulative distribution distribution of the TCIBF rate is

$$F_{R_{Sum}^{\mathrm{CI}}}(r)=F_{\lambda_{\min}}\left(T^{-1}(r)\right), \tag{3.76}$$

and $F_{R_{Sum}^{\mathrm{CI}}}(r)$ belongs to the domain of attraction of the Gumbel distribution.

Using the LTD theorem with $K \to \infty$

$$\frac{\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)-a_K}{b_K} \to E_0, \tag{3.77}$$

where $E_0 = 0.5772\cdots$ is the Euler constant,

$$a_K = \log\left(\rho F_{\lambda_{\min}}^{-1}\left(1-\frac{1}{K}\right)\right) \text{ and} \tag{3.78}$$

$$b_K = \log\left(\frac{F_{\lambda_{\min}}^{-1}\left(1-\frac{1}{Ke}\right)}{F_{\lambda_{\min}}^{-1}\left(1-\frac{1}{K}\right)}\right). \tag{3.79}$$

For $K \to \infty$, $\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right)$ may be evaluated as

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right) \approx M\left(a_K + E_0 b_K\right). \tag{3.80}$$

To begin, the condition in (3.75) is evaluated

$$\lim_{\lambda\to\infty} \frac{d}{d\lambda}\left[\frac{1-F_{\lambda_{\min}}(\lambda)}{f_{\lambda_{\min}}(\lambda)}\right] = \lim_{\lambda\to\infty} \frac{d}{d\lambda}\left[\frac{1-\left(1-e^{-\lambda M}\right)}{Me^{-\lambda M}}\right] = 0. \tag{3.81}$$

This shows that $F_{R_{Sum}^{\mathrm{CI}}}(r)$ belongs to the domain of attraction of the Gumbel distribution and the LTD theorem is indeed applicable. Next, to evaluate $F_{\lambda_{\min}}^{-1}\left(1-K^{-1}\right)$ in (3.78), we note that

$$F_{\lambda_{\min}}\left(\lambda = \frac{\log K}{M}\right) = 1-\exp\left(-\frac{\log K}{M}M\right) = 1-\frac{1}{K} \text{ and} \tag{3.82}$$

$$F_{\lambda_{\min}}\left(\lambda = \frac{\log Ke}{M}\right) = 1-\exp\left(-\frac{\log Ke}{M}M\right) = 1-\frac{1}{Ke}. \tag{3.83}$$

Using these, the values of a_K and b_K are then

$$a_K = \log\left(\rho\frac{\log K}{M}\right) \text{ and} \tag{3.84}$$

$$b_K = \log_2\left(1+\frac{1}{\log K}\right). \tag{3.85}$$

Plugging the values of a_K and b_K in (3.80), the ergodic TCIBF sum rate when $K \to \infty$ is then

$$\mathbb{E}\left(R_{Sum}^{\mathrm{CI}}\right) \approx M\left(\log\left(\rho\frac{\log K}{M}\right) + \underbrace{E_0 \log\left(1+\frac{1}{\log K}\right)}_{\approx 0 \text{ when } K\to\infty}\right)$$

$$\approx M \log \frac{\rho}{M} \log K, \quad \text{where } K \to \infty. \tag{3.86}$$

This means that scaling with M is possible when K and/or SNR are large. In particular, the ergodic TCIBF sum rate $\mathbb{E}\left(R_{Sum}^{\text{CI}}\right)$ scales as for DPC when $K \to \infty$, that is, as $M \log \log K$ [18] because

$$\lim_{K \to \infty} \frac{M \log \frac{\rho}{M} \log K}{M \log \log K} = 1 \tag{3.87}$$

using the l'Hôpital's rule with $t = \log K$ and letting $t \to \infty$.

3.2 Block Diagonalized (BD) Zero-forcing Beamforming (ZFBF) for Multi-Antenna Terminals

3.2.1 Factors Affecting Sum Rates of Block Diagonalized ZFBF Systems

This section begins with systems using direct block diagonalization (direct-BD [27], see Section 2 for system model) to achieve zero-forcing beamforming to users with multi-antenna terminals. Given *S*, a pool of *K* potential users, a subset $S_r \subseteq S$ of $|S_r| = K_r$ users has to be chosen whenever *M*, the number of transmit antennas is less than the total number of receive antennas from the users to be served, that is, $M < \sum_{j=1}^{K} N_j$, where N_j is the number of antennas at user *j*. From Section 2, the sum rate expression for direct-BD is

$$R_{Sum}^{\mathrm{DBD}}\left(\mathbf{H}_j\right) = \max_{\substack{\mathbf{H}_j,\ \mathbf{R}_j, \mathbf{T}_j, \mathbf{R}_{d_jd_j};\\ \text{s.t. tr}(\mathbf{R}_{d_jd_j})=P,\ \mathbf{H}_i\mathbf{T}_j=0, i\neq j}} \sum_{j=1}^{K} \log\det\left(\mathbf{I}_{N_j} + \mathbf{H}_j\mathbf{T}_j\mathbf{R}_{d_jd_j}\mathbf{T}_j^H\mathbf{H}_j^H / \sigma^2\right). \quad (3.88)$$

Since direct-BD achieves zero CCI by projecting each $\mathbf{H}_j$ onto $\tilde{\mathbf{H}}_j^{\perp}$, the null space of $\tilde{\mathbf{H}}_j = [\mathbf{H}_1^T \cdots \mathbf{H}_{j-1}^T\ \mathbf{H}_{j+1}^T\ \mathbf{H}_K^T]^T$, it is preferable that the user channels within S_r are close to being orthogonal. Hence, S_r must be properly chosen to give high projected channel gains in $\mathbf{H}_{p_j} = \mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$ and yield high sum rates. However, sole reliance on user selection does not attain the maximum sum rate because the judicious implementation of receive-antenna selection (RAS) in direct-BD improves the spatial mode gains of $\mathbf{H}_{P_j}$ in two ways.

First, the removal of antennas with high *inter*-terminal correlation from the members of S_r helps the user-channel matrices $\mathbf{H}_j$ to get closer to being mutually orthogonal. This improves the projected channel gains in $\mathbf{H}_{p_j}$ because it *decreases* the angle between the sub-spaces spanned by $\mathbf{H}_j$ and $\tilde{\mathbf{H}}_j^{\perp}$. This can be readily seen by expressing $\mathbf{H}_{p_j} = \mathbf{H}_j\tilde{\mathbf{P}}_j^{\perp}$, where $\tilde{\mathbf{P}}_j^{\perp} = \mathbf{I}_M - \tilde{\mathbf{H}}_j^H\left(\tilde{\mathbf{H}}_j\tilde{\mathbf{H}}_j^H\right)^{-1}\tilde{\mathbf{H}}_j$ is the orthogonal complement projection matrix of $\tilde{\mathbf{H}}_j$. This leads to $\mathbf{H}_{p_j} = \mathbf{H}_j - \mathbf{L}_j$, where $\mathbf{L}_j = \mathbf{H}_j \tilde{\mathbf{H}}_j^H \left(\tilde{\mathbf{H}}_j \tilde{\mathbf{H}}_j^H \right)^{-1} \tilde{\mathbf{H}}_j$ represents the projection loss. Note that in the extreme cases, $\mathbf{H}_{p_j} = \mathbf{H}_j$ and $\mathbf{H}_{p_j} = \mathbf{0}$ result when the row vectors of $\mathbf{H}_j$ are drawn from the row spaces of $\tilde{\mathbf{H}}_j^{\perp}$ and $\tilde{\mathbf{H}}_j$, respectively. This means $0 \le ||\mathbf{H}_{P_j}||_F^2 \le ||\mathbf{H}_j||_F^2$ and approaching orthogonality between all $\mathbf{H}_j$ will help lower the projection losses. The sum-rate impact of $\mathbf{L}_j$ can be seen by approximating (3.88) as

$$R_{Sum}^{\mathrm{DBD}}\left(\mathbf{H}_j\right) \approx \max_{\substack{\mathbf{H}_j,\ \mathbf{R}_j, \mathbf{T}_j, \mathbf{R}_{d_j d_j};\\ \text{s.t. } \mathrm{tr}(\mathbf{R}_{d_j d_j})=P,\ \mathbf{H}_i \mathbf{T}_j = 0,\, i \neq j}} \log_2 \prod_{j=1}^{K_r} \left(\left| \det\left(\mathbf{R}_j \left(\mathbf{H}_j - \mathbf{L}_j \right) \mathbf{V}_j^{(1)} \right) \right|^2 \det\left(\mathbf{R}_{d_j d_j} / \sigma^2 \right) \right), \tag{3.89}$$

where a high SNR regime is assumed and equal power allocation can be applied on all data streams.

Let $S_r' \subseteq S_r$ where $\left(|S_r'| = K_r' \right) \le |S_r|$ be the set of users remaining after RAS. Note that RAS may result in the dropping of one or more users. Let $R_{Sum}^{\prime\mathrm{DBD}}\left(\mathbf{H}_j'\right)$, $\mathbf{R}_j'$, $\mathbf{H}_j'$, $\mathbf{L}_j'$ and $\mathbf{V}_j'^{(1)}$ represent the corresponding entities after RAS. Let $\{U_j \in S_g\} \subseteq S_r'$ be a subset of $K_g \le K_r'$ users with $\left| \det\left(\mathbf{R}_j' \left(\mathbf{H}_j' - \mathbf{L}_j' \right) \mathbf{V}_j'^{(1)} \right) \right|^2 > \left| \det\left(\mathbf{R}_j \left(\mathbf{H}_j - \mathbf{L}_j \right) \mathbf{V}_j^{(1)} \right) \right|^2$. It is possible for $R_{Sum}^{\prime\mathrm{DBD}}\left(\mathbf{H}_j'\right) > R_{Sum}^{\mathrm{DBD}}\left(\mathbf{H}_j\right)$ if

$$\left(\prod_{\{j:\ U_j \in S_g\}} \left| \det\left(\mathbf{B}_j' \right) \right|^2 \right) \left(\prod_{\{m:\ U_m \in S_r' \setminus S_g\}} \left| \det\left(\mathbf{B}_m' \right) \right|^2 \right) > \left(\prod_{\{n:\ U_n \in S_r\}} \left| \det\left(\mathbf{B}_n \right) \right|^2 \right), \tag{3.90}$$

where $\mathbf{B}_l \triangleq \mathbf{R}_l \left(\mathbf{H}_l - \mathbf{L}_l \right) \mathbf{V}_l^{(1)}$, $\{U_j \in S_g;\ j = 1, \cdots K_g\}$ and $U_m \in S_r' \setminus S_g$ means members from S_r' excluding those in S_g. It is clear from (3.90) that higher sum rates may result from RAS when members of S_g have rate gains that outweigh the rate losses in the other users affected by RAS. In addition, RAS has a *mutual* effect among all members of S_r' since the members of S_g may also have undergone RAS and have reduced array sizes.

Second, each antenna removal at a particular terminal provides an additional degree of freedom to *all* other terminals in S_r'. For example, if one antenna is removed from user *k*, all other users *j* have projected channels $\mathbf{H}_{P_j}$ with dimensions

$$N_j \text{ x } (M + 1 - \textstyle\sum_{i=1, i \neq j}^{K_r} N_i), \tag{3.91}$$

where N_i represents the original number of receive antennas at each terminal. The number of columns in $\{\mathbf{H}_{P_j} : \forall j, j \neq k\}$ is increased by one and has the effect of adding more transmission resources to all users other than *k*. This raises the channel gains in $\mathbf{H}_{P_j}$, which can be explained in terms of their singular values. The resulting channel matrix $\mathbf{H}_k'$ has dimensions $(N_k' \text{ x } M)$, where $N_k' = N_k - 1$. User *k*'s single-user capacity and the multi-user sum capacity are correspondingly reduced.

Since RAS is not performed on any other user, we have $\tilde{\mathbf{H}}_k' = \tilde{\mathbf{H}}_k$, i.e., the row null space remains unchanged, which leads to $\tilde{\mathbf{V}}_k'^{(0)} = \tilde{\mathbf{V}}_k^{(0)}$. The projected channel $\mathbf{H}_{p_k}' = \mathbf{H}_k' \tilde{\mathbf{V}}_k'^{(0)}$ has dimensions

$$(N_k - 1) \text{ x } (M - \textstyle\sum_{i=1, i \neq k}^{K_r} N_i), \tag{3.92}$$

that is, a one-row reduction with no change in the column dimension. Let the singular values of $\mathbf{H}_{p_k}'$ be $\sigma_{\max}(\mathbf{H}_{p_k}') \geq \sigma_2(\mathbf{H}_{p_k}') \geq \cdots \geq \sigma_{\min}(\mathbf{H}_{p_k}')$, then $\sigma_{\max}(\mathbf{H}_{p_k}) \geq \sigma_{\max}(\mathbf{H}_{p_k}') \geq \cdots \geq \sigma_{\min}(\mathbf{H}_{p_k}') \geq \sigma_{\min}(\mathbf{H}_{p_k})$ since $\text{rows}(\mathbf{H}_{p_k}) \leq \text{columns}(\mathbf{H}_{p_k})$ [67]. The singular values of $\mathbf{H}_{p_k}'$ lie between those of the original $\mathbf{H}_{p_k}$ and hence, the total channel power gain $\|\mathbf{H}_{p_k}'\|_F^2 < \|\mathbf{H}_{p_k}\|_F^2$, where $\|\mathbf{H}_{p_k}'\|_F^2 \triangleq \text{tr}(\mathbf{H}_{p_k}' \mathbf{H}_{P_k}'^H) = \sum_{i=1}^{N_k'} \lambda_i'$ and λ_i' are the eigen-values of $\mathbf{H}_{p_k}' \mathbf{H}_{P_k}'^H$. For any *other* user *j*, the row dimension of $\tilde{\mathbf{H}}_j$ is reduced by one and let it be represented as $\tilde{\mathbf{H}}_j'$. Let $\tilde{\mathbf{V}}_j'^{(0)}$ be the new orthonormal basis associated with the row null space of $\tilde{\mathbf{H}}_j'$ and the new projected channel be $\mathbf{H}_{P_j}' = \mathbf{H}_j \tilde{\mathbf{V}}_j'^{(0)}$. By virtue of a one-antenna reduction in user k, the column dimension of the projected channels of *all* other users j is increased by one, that is,

$$N_j \text{ x } (M + 1 - \sum_{i=1, i \neq j}^{K_r} N_i). \tag{3.93}$$

Since $\text{rows}(\mathbf{H}_{P_j}) \leq \text{columns}(\mathbf{H}_{P_j})$, then [67]

$$\sigma_{\max}(\mathbf{H}'_{P_j}) \geq \sigma_{\max}(\mathbf{H}_{P_j}) \geq \cdots \geq \sigma_{\min}(\mathbf{H}'_{P_j}) \geq \sigma_{\min}(\mathbf{H}_{P_j}) \tag{3.94}$$

Equation (3.94) shows that the singular values in $\mathbf{H}'_{P_j}$ may be greater than the original singular values of $\mathbf{H}_{P_j}$. Given that $\text{tr}(\mathbf{H}'_{P_j}\mathbf{H}'^{H}_{P_j}) > \text{tr}(\mathbf{H}_{P_j}\mathbf{H}^{H}_{P_j})$ because of the additional column in $\mathbf{H}'_{P_j}$ and that $\text{rank}(\mathbf{H}'_{P_j}) = \text{rank}(\mathbf{H}_{P_j})$, this ensures that $\sigma_i(\mathbf{H}'_{P_j}) > \sigma_i(\mathbf{H}_{P_j})$ will be true for some values of i in (3.94). In turn, this creates the potential for higher *total* channel power gain, i.e., $\|\mathbf{H}'_{P_j}\|_F^2 > \|\mathbf{H}_{P_j}\|_F^2$ and the potential for higher sum rates despite sum capacity loss due to user k. In fact, given that $\|\mathbf{H}_{P_j}\|_F^2 \sim \chi^2_{2N_j^2(M - \Sigma_{i=1, i \neq j}^{Kr} N_i)}$ and $\|\mathbf{H}'_{P_j}\|_F^2 \sim \chi^2_{2N_j^2(M + 1 - \Sigma_{i=1, i \neq j}^{Kr} N_i)}$ where χ^2_{2i} is a chi-square random variable with $2i$ degrees of freedom, the probability of $\|\mathbf{H}'_{P_j}\|_F^2 > \|\mathbf{H}_{P_j}\|_F^2$ is raised due to one more column in $\mathbf{H}'_{P_j}$. When RAS is done on more than one user, it is important to note that (3.94) also applies to those users with *reduced* antenna array sizes due to RAS.

Let ε_k and ε_j represent the total number of receive antennas eliminated from user k and user j, respectively. Let $\{N_j : j = 1, \cdots, K_r\}$ be the original number of antennas at each user j prior to RAS and hence $\varepsilon_j \in \{0, 1, \cdots, N_j\}$. The dimensions of the projected channel $\mathbf{H}'_{p_k}$ for user k can be more generally expressed as

$$\left(N_k - \varepsilon_k\right) \text{ x } \left(M - \sum_{j=1, j \neq k}^{K_r} \left(N_j - \varepsilon_j\right)\right). \tag{3.95}$$

Hence, user k's projected channel will have its row dimension reduced by ε_k after RAS is applied on it, while its column dimension may be *increased* by the amount $\beta_k = \sum_{j=1, j \neq k}^{K_r} (\varepsilon_j)$ when RAS is performed on other users as well. In general, this means that the singular values of those users with reduced array size may still be increased due to additional columns in their projected channels, although the RAS process might have reduced their channel ranks. In this way there is *mutual* benefit to be obtained when RAS is done at all users to remove antennas with high correlation.

Note that the above analysis extends readily to block diagonalization schemes that use $\mathbf{R}_j$, receive-weight matrices for spatial-mode allocation. These schemes have better performance because no antennas are dropped at the user terminals during spatial-mode allocation. Examples of such schemes are the coordinated transmit-receive (CTR) scheme [27] and the null space directed SVD (Nu-SVD) scheme [29]. By replacing each $\mathbf{H}_j$ with a virtual channel $\mathbf{H}_{Vj} = \mathbf{R}_j^H \mathbf{H}_j$, the equivalent of RAS is then spatial mode selection (SMS). In this way, the above arguments on the benefits of RAS for direct-BD are applicable to schemes that operate on the virtual channels, such as CTR and Nu-SVD. The main drawback for such schemes is the need for the base station to send $\mathbf{R}_j$ to each of the chosen users.

Table 3.6 illustrates the benefits of RAS/SMS using direct-BD [27] and Nu-SVD [29] with a fixed channel realization. A system with $K_r = 8$ users, $N_j = 4 \;\; \forall j$, $M = \sum_{j=1}^{K_r} N_j = 32$ and SNR = 20dB is used. This large system is deliberately chosen to better illustrate the impact of RAS/SMS. To avoid exhaustive search, a RAS algorithm from Chapter 5 is used to perform RAS and SMS. As shown, RAS/SMS has a substantial impact on the system sum rates with improvements of ~46% and ~56% for direct-BD and Nu-SVD, respectively. Interestingly, many users with reduced antenna-array sizes or reduced spatial-mode sets enjoy rate increases, e.g., users #1 and #3. The rate loss for users #2 and #5 in the direct-BD scheme is not large despite having 2 antennas removed. The same is also true in Nu-SVD where user #2 has 2 modes removed. As expected, Nu-SVD with SMS performs better than direct-BD with RAS since all receive antennas are utilized. The example demonstrates the mutual benefit among users when judicious RAS/SMS is performed across the chosen group.

Table 3.6: Sum rate improvement due to RAS/SMS – A snapshot

User	#1	#2	#3	#4	#5	#6	#7	#8	Total
Direct-BD without RAS									
#Ants	4	4	4	4	4	4	4	4	
Rate	12.2	12.5	11.7	10.5	12.4	10.8	12.7	11.0	93.8
Direct-BD with RAS									
#Ants	3	2	3	4	2	3	4	3	
Rate	16.8	11.4	16.9	22.7	12.0	16.4	23.9	17.2	**137.4**
Nu-SVD without SMS									
#Modes	4	4	4	4	4	4	4	4	
Rate	12.2	12.5	11.7	10.5	12.4	10.8	12.7	11.0	93.8
Nu-SVD with SMS									
#Modes	3	2	3	3	3	3	4	4	
Rate	17.6	11.9	18.2	17.4	17.4	16.4	23.9	23.1	**145.9**

When there are potential users beyond the chosen group, *additional* sum rate gains due to multi-user diversity may be realized by scheduling more users if the RAS/SMS process has released transmission resources. This is justified by the scaling expression $M \log\log KN$ [18], which shows that M channels must be served to reap the full benefits of multi-user diversity arising from a large user pool and the available degrees of freedom. A greedy procedure for realizing this is: (a) add the next best user to the current group (using a user selection algorithm), (b) perform RAS/SMS to remove antennas/modes that cause low sum rates, (c) iterate until the next chosen user causes a sum rate drop even after RAS/SMS or when the scheduling of M channels is reached. Note that this procedure is also applicable to channel inversion beamforming (TCIBF) systems where the users have single-antenna terminals. However, the improvement is not expected to be as significant when a good user-selection algorithm is already in place.

3.2.2 Ergodic BD Sum Rate Analysis with RAS/SMS and User Selection under Rayleigh Fading Channels

As for the channel-inversion beamforming (TCIBF) case, two cases are considered here, namely (a) the total number of receive antennas matches the total number of transmit antennas, i.e., $\sum_{j=1}^{K} N_j = M$, and (b) a large user pool exists and a subset S_r of $|S_r| = K_r$ users is chosen such that $\sum_{j=1}^{K_r} N_j = M$. Each user terminal has N_j receive antennas and we assume negligible inter-terminal and intra-terminal antenna correlation. A novel approach to establish a lower bound for case (a) under the influence of receive antenna selection (RAS) or spatial mode selection (SMS) is developed here. Importantly, it is shown for case (a) that the maximum ergodic BD sum rates usually occur at $\sum_{j=1}^{K} N_j < M$ for practical SNR ranges and that random RAS/SMS provides fairly good performance compared to judicious RAS/SMS. It will then be extended to cover case (b) where the joint effects of user selection and RAS/SMS are captured.

3.2.2.1 Impact of Receive Antenna Selection (RAS)

This is examined as case (a), where the user pool of *K* users is such that the total number of receive antennas matches the total number of transmit antennas, i.e., $\sum_{j=1}^{K} N_j = M$. This condition meets the BD pre-coding constraint and user selection is therefore not strictly required. Instead, the impact of antenna de-selection from the multi-antenna terminals is examined. Antenna de-selection is more commonly referred to as receive-antenna selection or RAS. Note that the situation may arise where RAS causes a user with poor channel gain to be completely de-selected. It can be said therefore that user selection is subsumed under receive-antenna selection. For the first case (a), the ergodic sum rate for block diagonalized systems is from Chapter 2,

$$\mathbb{E}\left(R_{Sum}^{\mathrm{BD}}\right) = \mathbb{E}\left(\sum_{j=1}^{K} \log_2 \det\left(\mathbf{I}_{N_j} + \mathbf{R}_{d_j d_j} \Sigma_j^2 / \sigma^2\right)\right), \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{dd}) \le P, \tag{3.96}$$

where $\mathbf{\Sigma}_j^2 = diag\left(\lambda_{j,1},\cdots\lambda_{j,N_j}\right)$ where $\mathbf{\Sigma}_j^2$ contains the eigenvalues of $\mathbf{H}_{Pj}\mathbf{H}_{Pj}^H$ and $\mathbf{H}_{Pj} \triangleq \mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$ are projected single user channel with dimensions N_j x $(M-\sum_{i=1,i\neq j}^{K}N_i)$. Next, we assume that all user terminals have an equal number of receive antennas, i.e., $N_j = \eta$ $\forall j$ and $\sum_{j=1}^{K}N_j = M$. The lower bounding approach for TCIBF in Section 3.1.3.1 will be adapted for application here. Specifically, antenna/mode selection will be done on a *random* basis and on the use of unordered eigenvalues of Wishart matrices. We begin by introducing the unordered eigenvalue λ_j of $\mathbf{W}_{Pj}(n_j,m_j) = \mathbf{H}_{Pj}\mathbf{H}_{Pj}^H$ into (3.96) where

$$\left.\begin{aligned} n_j &= N_j \\ m_j &= \left(M-\sum_{i=1,i\neq j}^{K}N_i\right) \end{aligned}\right\} \tag{3.97}$$

To simplify the analysis, we assume a high SNR regime and allow for equal power allocation among the users, i.e., $\mathbf{R}_{d_jd_j} = \left(P/\sum_{j=1}^{K}N_j\right)\mathbf{I}_{N_j}$ where $\mathbf{I}_{N_j}$ is an N_j x N_j identity matrix. Then (3.96) becomes

$$\mathbb{E}\left(R_{Sum}^{\mathrm{BD}}\right) \geq \mathbb{E}\left(\sum_{j=1}^{K}\log_2\det\left(\frac{P}{\sigma^2\sum_{i=1}^{K}N_i}\lambda_j\mathbf{I}_{N_j}\right)\right), \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P,$$

$$\geq \mathbb{E}\left(\sum_{j=1}^{K}\log_2\left(\frac{P}{\sigma^2\sum_{i=1}^{K}N_i}\lambda_j\right)^{N_j}\right), \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P,$$

$$\mathbb{E}\left(R_{Sum}^{\mathrm{BD}}\right) \geq \sum_{j=1}^{K}\mathbb{E}\left(N_j\log_2\left(\frac{P}{\sigma^2\sum_{i=1}^{K}N_i}\lambda_j\right)\right), \quad \text{s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P. \tag{3.98}$$

We want to examine the impact of antenna de-selection, which is more commonly phrased as receive antenna selection or RAS, across the user terminals. If one antenna is removed from user k, then all other users j will have projected channels $\mathbf{H}_{P_j}$ with dimensions $N_j \text{ x } (M+1-\sum_{i=1,i\neq j}^{K}\widehat{N}_i)$, where $\{\widehat{N}_i : i=1,\cdots,K\}$ represents the original number of receive antennas at each terminal. In general, the dimensions of the projected channel $\mathbf{H}'_{p_k}$ for user k can be more generally expressed as

$$\left(\widehat{N}_k-\varepsilon_k\right) \text{x} \left(M-\sum_{j=1,\, j\neq k}^{K}\left(\widehat{N}_j-\varepsilon_j\right)\right), \tag{3.99}$$

where $\varepsilon_j \in \{0,1,\cdots,\widehat{N}_j\}$. Hence, user k's projected channel will have its row dimension reduced by ε_k after RAS while its column dimension may be *increased* by the amount

$$\beta_k=\sum_{j=1,j\neq k}^{K}(\varepsilon_j) \tag{3.100}$$

when RAS is performed on other users as well. When $\widehat{N}_j=\eta \;\; \forall j$, (3.99) becomes

$$\left(\eta-\varepsilon_k\right) \text{x} \left(M+\beta_k-(K-1)\eta\right). \tag{3.101}$$

The parameters of $\mathbf{W}_{Pj}(n_j,m_j)=\mathbf{H}_{Pj}\mathbf{H}_{Pj}^{H}$ will therefore change accordingly when RAS is done across the user terminals. To help visualize the impact of RAS on the BD ergodic sum rate in (3.98), we use a case with 8 users, each equipped with 4 antennas. The base station is equipped with 8 x 4 = 32 transmit antennas. For a start, we would like to observe the impact of a progressive one-antenna reduction from each user, done in a round-robin style. Let λ represent the unordered eigenvalue of a user without RAS and λ' represent the unordered eigenvalue of a user that has a one-antenna reduction. The ergodic sum rate is then

$$\mathbb{E}\left(R_{Sum}^{\text{BD-RAS}}\right) \geq (K-q)\eta\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda\right)\right)+q(\eta-1)\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda'\right)\right),$$
$$\text{s.t. } \operatorname{tr}(\mathbf{R}_{dd})\leq P, \tag{3.102}$$

where q is progressively stepped through $q=0,\cdots,K$. The first expectation in (3.102) is

$$\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda\right)\right)=\int_{-\infty}^{\infty}\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda\right)f_\lambda(\lambda)\ d\lambda,$$

where

$$f_\lambda(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\varphi_i(\lambda)^2\lambda^{m-n}e^{-\lambda}$$

with

$$\left.\begin{array}{c} n=\eta \\ m=\big(M+q-(K-1)\eta\big) \end{array}\right\} \qquad (3.103)$$

The second expectation in (3.102) is

$$\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda'\right)\right)=\int_{-\infty}^{\infty}\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda'\right)f_\lambda(\lambda')\ d\lambda',$$

where

$$f_{\lambda'}(\lambda')=\frac{1}{n'}\sum_{i=1}^{n'}\varphi_i(\lambda')^2\lambda^{\prime m'-n'}e^{-\lambda'}$$

with

$$\left.\begin{array}{c} n'=\eta-1 \\ m'=\big(M+(q-1)-(K-1)\eta\big) \end{array}\right\} \qquad (3.104)$$

Equation (3.102) may be iterated to progressively remove two or more antennas from each user. This is done by the formation of two user groups G_1 and G_2. The first group initially contains all users and the second group is progressively incremented to reflect an antenna removal from each user until all users from G_1 ends up in G_2. The following expression may be used

$$\mathbb{E}\left(R_{Sum}^{\text{BD-RAS}}\right)\geq\underbrace{\left(K-U_2\right)}_{U_1}(\eta-r_1)\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-t)}\lambda\right)\right)+U_2(\eta-r_2)\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-t)}\lambda'\right)\right),$$
$$\text{s.t. } \operatorname{tr}(\mathbf{R}_{dd})\leq P, \qquad (3.105)$$

where U_1 is the number of users in the first group, r_1 is the number of antennas that is already removed from each user in the first group; $U_2 = 0, \cdots K$ is the number of users to be progressively included in the second group and $r_2 = r_1 + 1$ reflects the total number of antennas that would be removed from each user when the next antenna removal is done; and t is the total number of antennas that will be removed after the removal of the next antenna. Note that $(K\eta - t) = U_1(\eta - r_1) + U_2(\eta - r_2)$. To use (3.105), we state the current value of r_1 then $r_2 = r_1 + 1$, i.e., G_2 contains users with one more antenna removed. The number of users in G_2 is sequentially incremented from $U_2 = 0, \cdots K$ and $(K\eta - t) = U_1(\eta - r_1) + U_2(\eta - r_2)$ is updated at each step. The parameters of λ in G_1 are

$$\left.\begin{aligned} n_1 &= \eta - r_1 \\ m_1 &= \left(M + (t - r_1) - (K-1)\eta\right) \end{aligned}\right\} \qquad (3.106)$$

and the parameters of λ' in G_2 are

$$\left.\begin{aligned} n_2 &= \eta - r_2 \\ m_2 &= \left(M + (t - r_2) - (K-1)\eta\right) \end{aligned}\right\} \qquad (3.107)$$

where $t = K\eta - U_1(\eta - r_1) - U_2(\eta - r_2)$ represents the total number of antennas de-selected.

Continuing the example with 8 users and 4 antennas per user, we plot the ergodic BD sum rate given by (3.98) using the method in (3.105), versus the number of antennas de-selected in Figure 3.14. As shown, the ergodic BD sum rate improves even when random RAS is applied and the improvements are substantial. The optimum number of antennas de-selected is high when the SNR is low and vice-versa. The lower-bound ergodic BD sum rate given by (3.98) compares well with the Monte Carlo results where random RAS is done (also shown in Figure 3.14). For example, numerical results yield ≈ 99 b/s/Hz without RAS and ≈ 139 b/s/Hz with RAS at 20dB SNR. This compares well with 82 b/s/Hz and 127 b/s/Hz, respectively obtained via (3.98). We see that $\max\left(\mathbb{E}\left(R_{Sum}^{\text{BD-RAS}}\right)\right)$ occurs at different values of t at different SNR and using (3.105)

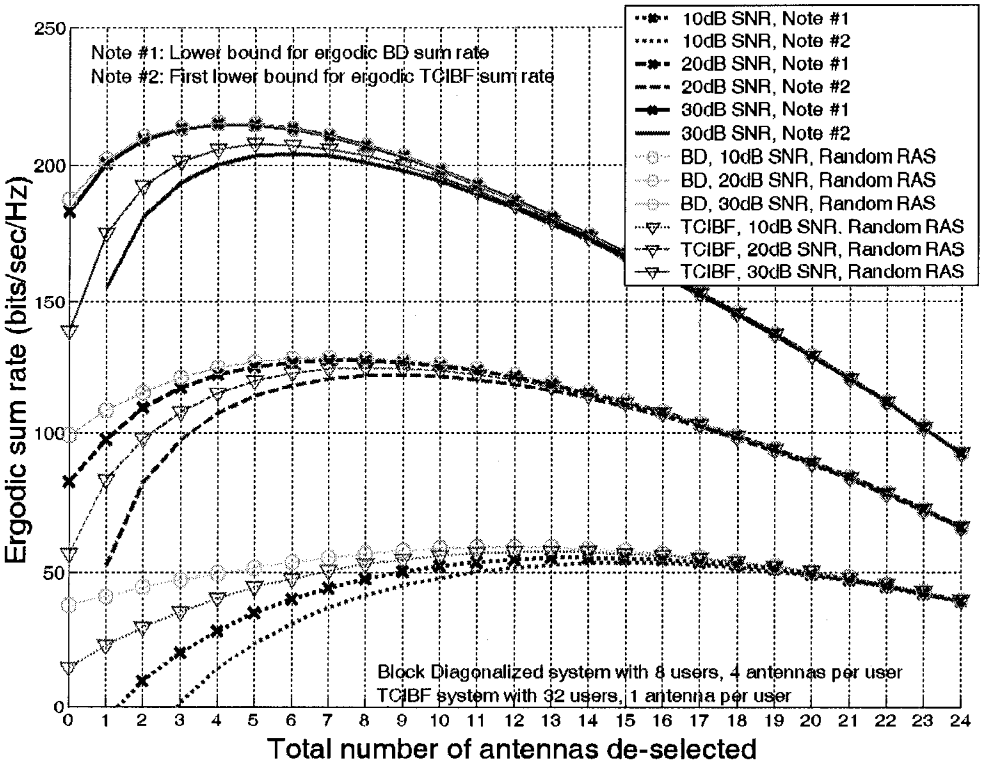


Figure 3.14: Impact of RAS on a block diagonalized ZFBF system

$$\max\left(\mathbb{E}\left(R_{Sum}^{\text{BD-RAS}}\right)\right) \geq \max_{t}\left(\begin{array}{l} U_1(\eta-r_1)\mathbb{E}\left(\log_2\left(\dfrac{P}{\sigma^2(K\eta-t)}\lambda\right)\right) \\ \quad + U_2(\eta-r_2)\mathbb{E}\left(\log_2\left(\dfrac{P}{\sigma^2(K\eta-t)}\lambda'\right)\right) \end{array}\right), \text{ s.t. } \mathrm{tr}(\mathbf{R}_{dd}) \leq P. \qquad (3.108)$$

It will become clear later during the discussion on selection algorithms that user subset selection is actually subsumed within the RAS process.

Next, block diagonalized systems may be lower-bounded by performing channel inversion across all antennas of all user terminals. The lower bound for ergodic TCIBF sum rate in (3.51) may then be modified for use in this context:

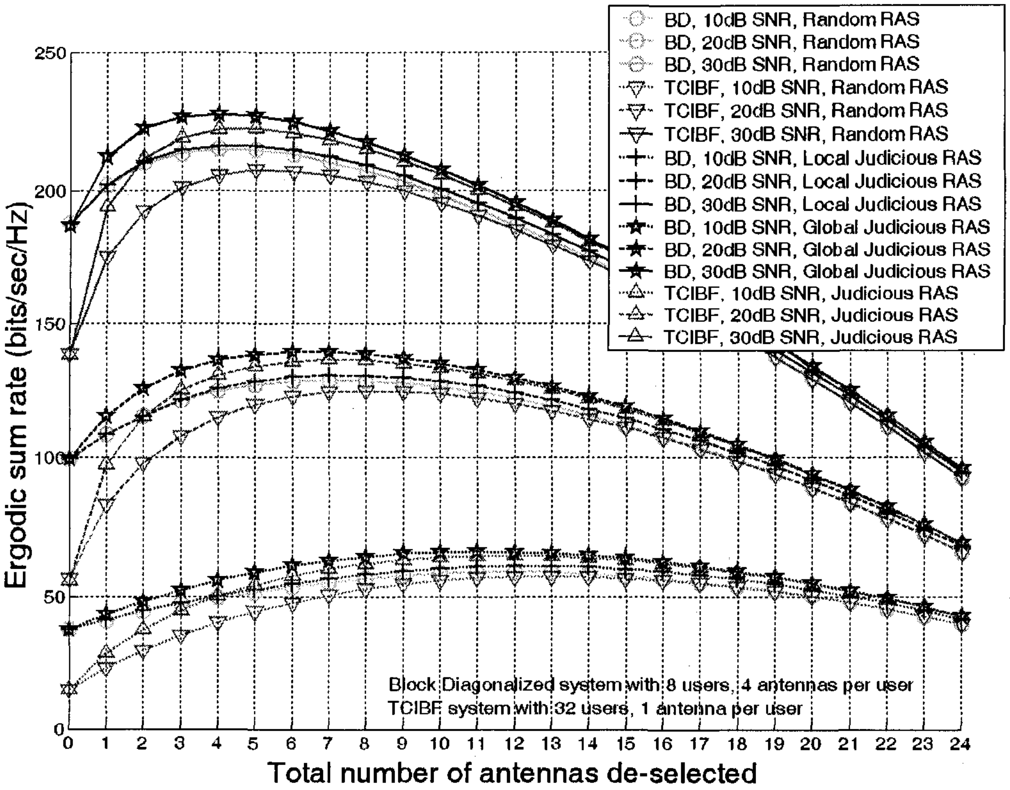


Figure 3.15: Performance of BD ergodic sum rate lower bound compared with simulated results with random RAS and an equivalent TCIBF system

$$\mathbb{E}\left(R_{Sum}^{\text{BD-RAS}}\right) \ge \mathbb{E}\left(R_{CI}\right) \ge (M-t)\log_2\left(\frac{P}{\sigma^2}\left(\frac{M}{(M-t)}-1\right)\right) \qquad \text{s.t. } \operatorname{tr}(\mathbf{R}_{dd}) \le P, \qquad (3.109)$$

where t is the total number of antennas removed from the equivalent TCIBF system with $K\eta$ receive antennas and M transmit antennas. The results using (3.109) for an equivalent TCIBF system with 32 users are also shown in Figure 3.14. It is clear that the lower bound of (3.51) requires the removal of more antennas to result in the maximum sum rate compared to the equivalent BD system. For comparison, numerical results using random RAS on the 32-user TCIBF system is also shown in Figure 3.14.

For comparison against random RAS, numerical results with judicious RAS using the maximum determinant ranking (MDR) algorithm from Chapter 5 are presented in Figure 3.15. For the 8-user BD system, the judicious RAS is done at the localized- and global levels. For localized judicious RAS, the MDR algorithm is employed locally at each user terminal to remove antennas that contribute to poor performance, that is, the base station is not involved in the RAS process. For global judicious RAS, the base station performs RAS on the composite channel matrix that is the concatenation of the individual user channel matrices. As shown in Figure 3.15, the ergodic sum rates arising from localized judicious RAS are only slightly better than random RAS performed at each terminal. However, better performance is obtained when judicious RAS done on a global basis at the base station. For comparison, results for an equivalent 32-user TCIBF system with random and judicious RAS are also shown.

Results for random RAS or localized judicious RAS show the possibility for reducing the channel matrix size to be fed back from each user back to the base station. This is true especially when the SNR levels are low where sum rate maximization involves the de-selection of more antennas. The lower bound in (3.98) may be used to guide the number of users that employ localized RAS for a given SNR level.

At this stage, we have shown that receive-antenna de-selection from some user terminals is needed on average when maximizing the ergodic sum rate for block diagonalized systems. This is commonly referred to as receive antenna selection or RAS. The sum rate is increased even when RAS is randomly done on a round-robin basis. For the case where each terminal has only one antenna, receive-antenna de-selection becomes identical to user de-selection.

3.2.2.2 Impact of User Selection

Using the same approach as Section 3.1.4, the main intention here is to demonstrate the joint impact of (a) user selection, (b) RAS/SMS and (c) different SNR levels, upon the ergodic BD sum rate when the user pool is large ($K > M$). An upper bound will be derived for the purpose of demonstrating the joint impact of these mechanisms. To simplify the task at this stage, the bound will be based on a number of assumptions and the sum rates provided are only good as ballpark estimates. The first expectation in (3.102) is

$$\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda\right)\right)=\int_{\infty}^{\infty}\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda\right)f_{\lambda_{(L):L}}(\lambda)\ d\lambda, \qquad (3.110)$$

where $f_{\lambda_{(L):L}}(\lambda)=L\,F_{\lambda_{l:L}}(\lambda)^{L-1}f_{\lambda_{l:L}}(\lambda)$ and $f_{\lambda_{l:L}}(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\varphi_i(\lambda)^2\lambda^{m-n}e^{-\lambda}$, with

$$\left.\begin{aligned} n&=\eta \\ m&=\left(M+q-(K-1)\eta\right)\end{aligned}\right\}$$

Similarly, the second expectation in (3.102) is

$$\mathbb{E}\left(\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda'\right)\right)=\int_{\infty}^{\infty}\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda'\right)f_{\lambda'_{(L):L}}(\lambda')\ d\lambda', \qquad (3.111)$$

where $f_{\lambda'_{(L):L}}(\lambda')=L\,F_{\lambda'_{l:L}}(\lambda')^{L-1}f_{\lambda'_{l:L}}(\lambda')$ and $f_{\lambda'_{l:L}}(\lambda')=\frac{1}{n'}\sum_{i=1}^{n'}\varphi_i(\lambda')^2\lambda^{m'-n'}e^{-\lambda'}$, with

$$\left.\begin{aligned} n'&=\eta-1 \\ m'&=\left(M+(q-1)-(K-1)\eta\right)\end{aligned}\right\}$$

Putting (3.110) and (3.111) into a form similar to (3.105), we can find the ergodic BD sum rate $\mathbb{E}\left(R_{Sum}^{\text{BD-USEL-RAS}}\right)$ that takes both user subset selection and RAS into account

$$\begin{aligned}\mathbb{E}\left(R_{Sum}^{\text{BD-USEL-RAS}}\right)=U_1(\eta-r_1)&\int_{\infty}^{\infty}\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda\right)f_{\lambda_{(L):L}}(\lambda)\ d\lambda \\ +U_2(\eta-r_2)&\int_{\infty}^{\infty}\log_2\left(\frac{P}{\sigma^2(K\eta-q)}\lambda'\right)f_{\lambda'_{(L):L}}(\lambda')\ d\lambda', \quad \text{s.t. } \operatorname{tr}(\mathbf{R}_{dd})\le P.\end{aligned} \qquad (3.112)$$

Continuing the example with 8 users and 4 antennas per user, we plot the ergodic BD sum rate $\mathbb{E}\left(R_{Sum}^{\text{BD-USEL-RAS}}\right)$ versus the number of antenna de-selected in Figure 3.16. The results show that both user selection and RAS are useful towards ergodic sum rate maximization. The effect of RAS is more pronounced when the potential user pool for subset selection is smaller and vice versa. This is to be expected since the chances of finding eigenvalues in

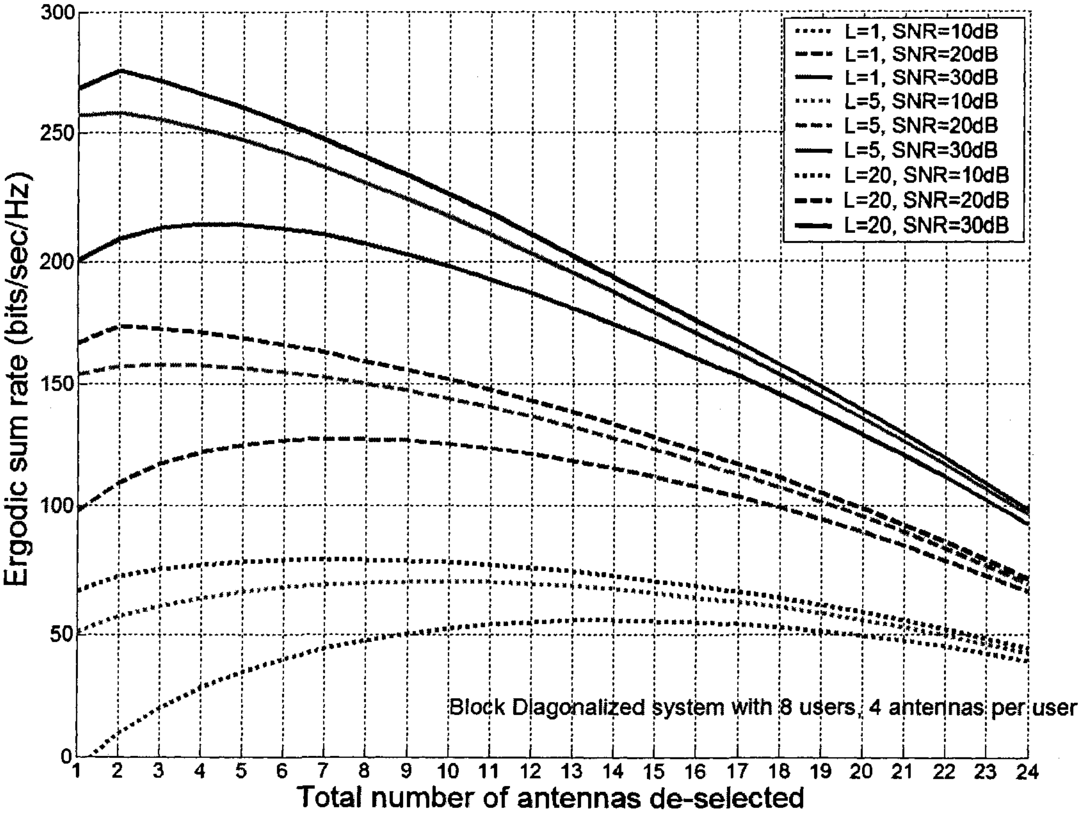


Figure 3.16: Impact of both user selection and RAS on a block diagonalized ZFBF system

the higher range is raised in large user pools, which results in higher ergodic sum rates. In practice, it means that the chances of finding user subsets with users of good channel gains that are close to being orthogonal with each other are higher on average for larger user pools.

3.3 Considerations for Algorithms

Arising from the discussions above, the following is a list of considerations when designing algorithms for user/antenna/mode de-selection:

a. It is shown in Section 3.1.3 that random user/antenna de-selection is useful for increasing the ergodic transmit channel-inversion beamforming (TCIBF) sum rate. This implies that judicious de-selection will give better performance especially under heterogeneous channels.

b. It is shown in Section 3.2.2 that random antenna/mode de-selection is useful for increasing the ergodic block diagonalized (BD) beamforming sum rate. This also implies that judicious de-selection will give better performance especially under heterogeneous channels.

c. For BD systems, a round-robin style of random de-selection produces results that are close to localized judicious de-selection for homogeneous channels. Since the channel matrix size to be fed back to the base station is reduced after a localized RAS (be it random or judicious), a method for reducing the feedback overhead may be developed on this basis. For example, it is shown in Figure 3.15 that around 8 antennas must be removed from the 8-user BD system to achieve the best sum rate when SNR = 20dB. This means that one antenna must be removed from each user and this reduces the size of the channel matrix to be fed back from each user to the base station.

d. For BD systems with large user pools, antenna/mode de-selection should not be done too early during an incremental user selection process. This can be seen from Figure 3.15 where the sum rates go on a downward trend when the number of antennas/modes is too few. This implies that a group of *M* antennas/modes should be chosen first before attempting antenna/mode de-selection.

e. For TCIBF systems with large user pools, user/antenna de-selection should not be done too early during an incremental user selection process. This can be seen from Figure 3.9 where the sum rates go on a downward trend when the number of users/antennas is too few. This implies that a group of *M* users/antennas should be chosen first before attempting user/antenna de-selection.

3.4 Summary

For a given set of single-antenna terminals that is served by Transmit Channel Inversion Beamforming (TCIBF), we have shown the underlying mechanism behind the impact of user de-selection on the TCIBF sum rate. The process of random user de-selection was shown to be changing the parameters of Wishart matrices that may be derived from the overall concatenated system channel matrix or from the individual projected channel vector. The ergodic sum rate expressions for TCIBF demonstrate concavity with variations in K_s, which is the number of users left after de-selection. Optimum values of K_s for maximum ergodic sum rates can be determined for different SNR levels.

When the potential user pool is larger than the TCIBF pre-coding constraint, multi-user diversity (MUD) gain is available when judicious user subset selection is done. The impact of MUD via user subset selection was jointly analyzed with user de-selection using order statistics methods. It was shown that both user de-selection from within a chosen user subset and user subset selection can be combined to yield higher ergodic TCIBF sum rates. It was also shown that user de-selection has less impact when more user subsets are available for selection. This is to be expected since the chances of finding eigenvalues in the higher range is raised and result in higher ergodic sum rates. In practice, it means that the chances of finding user subsets with users of good channel gains that are close to being orthogonal with each other are higher on average for larger user pools.

For block diagonalized (BD) systems, we have also carried out a joint analysis on the impact of user subset selection and antenna de-selection. It was shown that the ergodic sum rates of BD systems benefit from receive antenna selection (RAS) or spatial mode selection (SMS), even when RAS/SMS is randomly done. The basis behind the benefits of both is the same as that experienced by TCIBF. Antenna de-selection is commonly referred to in literature as receive antenna selection or RAS. We note that user de-selection in TCIBF is synonymous with antenna de-selection for the BD case when all terminals are equipped with only one antenna. In addition, the user subset selection process in BD is subsumed under the RAS process, as will be shown in the section on algorithms.

Chapter 4

SELECTION AND ALLOCATION ALGORITHMS FOR TCIBF WITH SINGLE-ANTENNA TERMINALS

We will develop user selection algorithms for the purpose of sum rate maximization under two cases, (i) when $K \leq M$ and (ii) when $K > M$, where K is the user pool size. Proposals for fair scheduling and CSI feedback reduction will be given. Preliminary methods for resource allocation against a given set of QoS requirements will also be given. The challenge is to perform resource allocation while minimizing the sum rate loss. An analysis on the impact of transmit antenna selection (TAS) will also be given.

4.1 User Selection Algorithm for $K \leq M$

Given a pool of $K \leq M$ users, there is no possibility of replacing any user. Under poor channel conditions, the waterfilling process may assign zero power to users associated with low b_i values (see Chapters 2 and 3). Although this helps mitigate the signal attenuation problem in terms of meeting the power constraint, it does not achieve the best sum rate since the poor channel condition still exerts its influence on the remaining b_i values. To analyze the conditions under which TCIBF sum rate improvements would occur with subset reduction, we begin by representing the $1/b_i$ values as

$$\{1/b_i : i = 1, \cdots, K_r\} = \{(A_{11})_r \Delta_r^{-1}, \ldots\ldots, (A_{K_r K_r})_r \Delta_r^{-1}\}, \tag{4.1}$$

where $(A_{ii})_r$ are cofactors associated with the diagonal elements h_{ii} in $\mathbf{H}_r\mathbf{H}_r^H$ and $\Delta_r = \det(\mathbf{H}_r\mathbf{H}_r^H)$. Each $(A_{ii})_r$ is found after eliminating row i and column i in $\mathbf{H}_r\mathbf{H}_r^H$, which corresponds to eliminating row i in $\mathbf{H}_r \in \mathbb{C}^{K_r \mathrm{x} M}$ to give sub-matrix $\mathbf{H}_s \in \mathbb{C}^{K_s \mathrm{x} M}$

where $K_s = K_r - 1$, i.e., eliminating user i. Starting with $K = M$, let $\gamma_i = (\mu_T b_i - \sigma^2)_{k+}$ represent the computation of γ_i after having removed k rows and having waterfilling re-applied. Substituting (4.1) into the expressions for γ_i and the water-level μ_T,

$$\gamma_i = \left(\frac{\Delta_s P}{(A_{ii})_s K_a} \left(1 + \frac{\sigma^2}{\Delta_s P} \sum_{j=1}^{K_a} (A_{jj})_s \right) - \sigma^2 \right)_{k+}, \tag{4.2}$$

where $(A_{ii})_s$ and Δ_s are associated with $\mathbf{H}_s \mathbf{H}_s^H$ and $K_a = K_s - p + 1$. Using (4.2) in the TCIBF sum rate expression,

$$R(\mathbf{H}_s)_{\text{TZFBF}} = \max_{\mathbf{H}_s,\, \mathbf{R}_{ss},\, \text{tr}(\mathbf{R}_{ss}) = P} \left(\sum_{i=1}^{K_s} \log_2 \left(1 + \left(\frac{P \Delta_s}{\sigma^2 K_a (A_{ii})_s} \left(1 + \frac{\sigma^2}{\Delta_s P} \sum_{j=1}^{K_a} (A_{jj})_s \right) - 1 \right)_{k+} \right) \right). \tag{4.3}$$

Considering only those row eliminations that result in non-zero power allocations, i.e., $\gamma_i > 0$, $\forall i$, and $K_a = K_s$, (4.3) is re-written as

$$R(\mathbf{H}_s)_{\text{TZFBF}} = \max_{\mathbf{H}_s,\, \mathbf{R}_{ss},\, \text{tr}(\mathbf{R}_{ss}) = P} \left(\underbrace{K_s \log_2 \left(\frac{\Delta_s P}{\sigma^2} + \sum_{i=1}^{K_s} (A_{ii})_s \right)}_{Term\ I} - \underbrace{\log_2 \prod_{i=1}^{K_s} (A_{ii})_s}_{Term\ II} - \underbrace{K_s \log_2 K_s}_{Term\ III} \right). \tag{4.4}$$

It appears from (4.4) that $\Delta_s P / \sigma^2$ dominates under high SNR where higher Δ_s values will help increase $R(\mathbf{H}_s)_{\text{TZFBF}}$. This leads to a 2-part question, (i) does $S_s \subset S_r$ with $\Delta_s > \Delta_r$ exist, and (ii) if it does, would it result in a higher sum rate?

To answer the first part, note that $\mathbf{H}_r\mathbf{H}_r^H$ is positive definite Hermitian and the inclusion principle [67] implies

$$\lambda_{\min}(\mathbf{H}_r \mathbf{H}_r^H) \le \lambda_{\min}(\mathbf{H}_s \mathbf{H}_s^H) \le \lambda_{\max}(\mathbf{H}_s \mathbf{H}_s^H) \le \lambda_{\max}(\mathbf{H}_r \mathbf{H}_r^H), \tag{4.5}$$

where $\lambda_i(\mathbf{B})$ represents the eigenvalue λ_i of matrix $\mathbf{B}$. The justification for being able to apply the inclusion principle to *any* $\mathbf{H}_s \subset \mathbf{H}_r$ is given in Appendix A. Note that $\Delta_s > \Delta_r$ can happen when the lower eigenvalues of $\mathbf{H}_r\mathbf{H}_r^H$ transit from $\lambda_i(\mathbf{H}_r\mathbf{H}_r^H) < 1$ to $\lambda_i(\mathbf{H}_s\mathbf{H}_s^H) \ge 1$ after a row elimination in $\mathbf{H}_r$. In relation to (4.1), this situation occurs whenever a cofactor $(A_{ii})_r$ is *larger* than its associated determinant Δ_r, i.e., there is one or more $1/b_i > 1$. From (4.1), $\max(1/b_i)$ is associated with $\max((A_{ii})_r) = (A_{mm})_r$, the largest cofactor. This means that eliminating row *m* and column *m* column of $\mathbf{H}_r\mathbf{H}_r^H$, which corresponds to removing row *m* in $\mathbf{H}_r$ to give $\mathbf{H}_s$, will result in a $\mathbf{H}_s\mathbf{H}_s^H$ that possesses the largest determinant $(\Delta_s)_{\max} = (A_{mm})_r$ and more importantly, $(\Delta_s)_{\max} > \Delta_r$.

To answer if $(\Delta_s)_{\max} > \Delta_r$ would lead to a higher sum rate, (4.4) can be approximated at high SNR as

$$R(\mathbf{H}_s)_{\text{TZFBF}} \approx K_s \log_2\left(\Delta_s \frac{E_s}{N_o K_s}\right). \tag{4.6}$$

Equation (4.6) assumes $(\Delta_s P / \sigma^2) > \Sigma_{i=1}^{K_s}(A_{ii})_s$ and that dropping *Term II* is reasonable via the inequality of arithmetic- and geometric- means (AM-GM inequality), where:

$$(K_s)^{-1}\,\Sigma_{i=1}^{K_s}(A_{ii})_s \geq \left(\Pi_{i=1}^{K_s}(A_{ii})_s\right)^{1/K_s} \Rightarrow$$

$$\left(K_s \log_2 \Sigma_{i=1}^{K_s}(A_{ii})_s - K_s \log_2 K_s\right) \;\geq\; \log_2 \Pi_{i=1}^{K_s}(A_{ii})_s, \tag{4.7}$$

with equality only when $\{A_{ii}\}_s = \{c\}$, where *c* is a constant. Given a random matrix $\mathbf{H}_s$, $\{A_{ii}\}_s \neq \{c\}$, and applying (4.7) to (4.4), we see that (*Term I*)>(*Term II*) for all SNR levels because $K_s \log_2 \Sigma_{i=1}^{K_s}(A_{ii})_s > \log_2 \Pi_{i=1}^{K_s}(A_{ii})_s$ even when $\text{SNR} = 0$. Given this, (4.8) must hold for sum rate increases to occur with $\mathbf{H}_s \subset \mathbf{H}_r$, i.e., $R(\mathbf{H}_s)_{\text{TZFBF}} - R(\mathbf{H}_r)_{\text{TZFBF}} > 0$,

$$\Delta_s > K_s \left(\frac{\Delta_r}{K_r}\right)^{\frac{K_r}{K_s}} \left(\frac{E_s}{N_o}\right)^{\frac{K_r}{K_s}-1}. \tag{4.8}$$

The condition on Δ_s in (4.8) is not difficult to meet. Firstly, the existence of highly correlated pairs in $\mathbf{H}_r$ would render $\Delta_r \ll 1$. The removal of one such rows will result in $\Delta_s > \Delta_r$. Next, the removal of a user with very low channel gain will also result in higher Δ_s since

$$\left(\prod_{i=1}^{K_r} \lambda_i(\mathbf{Z}_r) = \Delta_r\right) \leq \prod_{i=1}^{K_r}(z_{ii})_r, \tag{4.9}$$

where $\mathbf{Z}_r = \mathbf{H}_r\mathbf{H}_r^H$ and $(z_{ii})_r = \text{diag}(\mathbf{Z}_r)$. Lastly we note that the power exponents in (4.8) pose no problems since if K_r and K_s are large, then $K_r / K_s \to 1$. However, this also implies that achieving the maximum sum rate with a small user subset is less likely, unless the user channel vectors are all highly correlated.

Next, with respect to the waterfilling process, we see from (4.2) that the influence of $\Sigma_{i=1}^{K_a}(A_{ii})_s$ is less when SNR and Δ_s are high. In addition, the "bucket bottom" is inversely proportional to Δ_s while $b_i \propto \Delta_s$. Hence, choosing the next subset that has the highest Δ_s among all subsets of the same cardinality is also reasonable from the waterfilling viewpoint, i.e., in avoiding zero power assignments so that $K_a \to K_s$.

At this point, we see that choosing each reduced user subset such that Δ_s is maximized is a reasonable approach in terms of the waterfilling process and the sum rate maximization process, under all channel and SNR conditions. This leads to a user selection algorithm that removes the user associated with the current $\max(1/b_i)$ value to result in the next highest determinant Δ_s. At high SNR, this approach would also give the best cooperative MIMO capacity among all such one-row reduced channel matrices. The algorithm's pseudo-code is listed in Table 4.1 and is referred to as Joint Rate Evaluation and User Selection (JREUS) because user selection is made possible during TCIBF rate evaluation since the $1/b_i$ values are available. It avoids the typical arrangement that entails *separate* user selection and TCIBF rate evaluation processes. Rate evaluation and user selection are performed until the sum rate drops, hence incurring a maximum of M steps. This is lower than the $\boldsymbol{O}(2^M)$ steps needed for exhaustive search. It is also lower than many existing algorithms because most operate on the principle of considering each remaining user against a subset of chosen users and incur $\boldsymbol{O}(M^2)$ steps. The extension "-MAX" refers to a JREUS version where user selection is guided by the $\max(1/b_i)$ value.

Note that stopping when $(\Delta_s)_{m+1} < (\Delta_s)_m$ is not optimal, where $(\Delta_s)_m$ represent the determinant of $(\mathbf{H}_s\mathbf{H}_s^H)_m$ at the m^{th} step of user removal. The primary reason is that

Table 4.1 JREUS-MAX Algorithm

```
Start with $K_s=M$ and initial $\mathbf{H}_s=\mathbf{H}_r\subset\mathbf{H}$;
Initialize $R\_max$ = 0, done = 0;

while done == 0
  compute $1/b_i$ values;
  find water level $\mu_T$;
  find $\gamma_i$ values;
  compute channel rate $R\_temp$;
     if $R\_max < R\_temp$
          $R\_max = R\_temp$;
          $\mathbf{H}_s\_found = \mathbf{H}_s$;
     else
          done = 1;
     end_if

  $k$ = argmax($1/b_i$);
  eliminate row $k$ in $\mathbf{H}_s$;
  form new $\mathbf{H}_s$;
end_loop

Output: $R\_max$ and set of active users in
$\mathbf{H}_s\_found$
```

$b_i = \Delta_s/(A_{ii})_s$ are also dependent on the cofactors and they may still increase even when $(\Delta_s)_{m+1} < (\Delta_s)_m$. In addition, (4.2) and (4.4) are also dependent on the subset size K_s.

To reduce complexity of JREUS-MAX, a recursive inversion procedure is given in Appendix B. It makes use of $(\mathbf{H}_r\mathbf{H}_r^H)^{-1}$ to derive the subsequent $(\mathbf{H}_s\mathbf{H}_s^H)^{-1}$ matrices. Numerical results show near-optimal performance using JREUS-MAX for both homogeneous and heterogeneous channels (see below).

4.1.1 Sub-optimality of JREUS-MAX and Alternative Strategies

The high SNR approximation in (4.6) allowed us to isolate the impact of Δ_s, which led to the development of JREUS-MAX. However, a search path that depends solely on $\max(1/b_i)$ is sub-optimal because the cofactors are ignored. We will discuss the possible alternative decision strategies to mitigate this sub-optimality and their effectiveness. The cofactor dependency in (4.2) and (4.4) may be viewed in terms of b_i as

$$R(\mathbf{H}_s)_{\text{TZFBF}} = \max_{\mathbf{H}_s,\, \mathbf{R}_{ss},\, \text{tr}(\mathbf{R}_{ss})=P}\left(K_s \log_2\left(\frac{P}{\sigma^2} + \sum_{j=1}^{K_s} \frac{1}{b_i}\right) + \log_2 \prod\nolimits_{i=1}^{K_s} b_i - K_s \log_2 K_s \right) \text{ and} \quad (4.10)$$

$$\gamma_i = \left(b_i \frac{P}{K_a}\left(1 + \frac{\sigma^2}{P}\sum_{i=1}^{K_a}\frac{1}{b_i}\right) - \sigma^2 \right)_{k+} . \quad (4.11)$$

Starting with a subset $S_r \subset S$ where $|S_r| = M$, let $(S_s)_{\max(1/b_i)} \subset S_r$, be a one-row reduced subset with $\max(\Delta_s)$ among all subsets of the same cardinality. However, since $b_i = \Delta_s / (A_{ii})_s$, higher projected channel gains may be achieved by another subset $S_t \neq (S_s)_{\max(1/b_i)}$. This leads to the idea of sum rate comparisons among *some* subsets of the same cardinality at each iteration stage. Given $b_i = \Delta_s / (A_{ii})_s$, one would examine those subsets whose determinants are close to $\max(\Delta_s)$. A simple algorithm is JREUS-ALT1, which comprises JREUS-MAX plus rate evaluation for another subset associated with the second highest $1/b_i$ value. The subset with the higher sum rate is selected at each stage and the greedy search proceeds until the sum rate drops. The recursive inverse in Appendix B can also be used. Numerical results show that JREUS-ALT1 is better / worse than JREUS-MAX for about 24% / 7% of the time. Despite this, the ergodic sum rate improvement of JREUS-ALT1 over JREUS-MAX is only 0.24%. This is because most improvements are not high while some negative results are high.

For comparison, we will also consider JREUS-ALT2 where the subsets up to the third highest $1/b_i$ value are examined. Note that Dimić's algorithm [50] is actually for the case when *all* subsets of the same cardinality are considered at each stage. Numerical results show that Dimić's algorithm is mostly worse off despite its higher search complexity, except when SNR < 10 dB. It also gets worse when correlation among users is high. This shows that a strategy that considers all subsets raises the probability of wrong search paths that lock onto local maxima. This leads to the notion of thresholding, i.e., consider an alternative subset only if its $1/b_i$ value is within a pre-defined threshold of $\max(1/b_i)$. Numerical results show that depending on the level of user correlation, thresholds of the range 0.58 to 0.90 improve the sum rate performances of JREUS-ALT1 and JREUS-ALT2. However, the improvements are not large and the improvements generally occur when the SNR levels are $< 22\,\text{dB}$ for low to midlevel of correlation. A method for optimum threshold determination is beyond the scope of this thesis. When complexity is of concern, detailed numerical results in Section 4.4 show that JREUS-MAX is adequate in achieving most of the optimal sum rate in the ergodic sense. In addition, performing JREUS-MAX for M steps instead of stopping whenever the sum rate drops is better for about 0.9% of the time and improves the ergodic sum rate by only about 0.03%. Hence the proposed stopping criterion for JREUS-MAX gives good performance with reduced complexity. An alternative search strategy via the equivalent of transmit antenna selection is given in [75]. Again, the ergodic sum rate improvement is not high and it has higher complexity than JREUS-ALT1.

4.2 User Selection Algorithm for $K > M$

When $K > M$, optimal sum rate maximization requires exhaustive search involving $\sum_{i=1}^{K_r=M} \binom{K}{i}$ TCIBF sum rate evaluations. Exhaustive search becomes impractical for large user pools and this has attracted much effort to develop efficient user selection algorithms using various approaches, e.g., [26], [44] – [50].

4.2.1 Brief Survey of Existing Algorithms

An overview of the algorithms in [26], [44] – [50] can be drawn from the optimal beamforming sum rate scaling expression $M \log\log KN$, where $K \gg 1$ is assumed [18]. The intuitive explanation as provided in [24] is that one can always find a roughly orthogonal set of M channels to transmit over when K is very large. Next, the quality of these channels grows roughly as $\log K$ for single antenna terminals because the maximum of independent exponential random variables describing received power distributions grows logarithmically. Given this, a simple user selection algorithm based on the users' channel gains, i.e., the channel vector norm lengths should perform better than random user selection. A total of K norm-length metrics must be examined from which a subset of M users may be chosen. Next, a better selection basis must consider the orthogonality among users or its equivalent. In general, this will incur the examination of $\sum_{j=0}^{M-1}(K-j)=M(2K-M+1)/2$, or $\boldsymbol{O}(MK)$ decision metrics for the algorithms in [26], [44] – [50].

In [44] – [46], the orthogonal complement projection approach is utilized. Selection is done incrementally where the next user is chosen based on the largest projection norm in the null space of the composite channel for the existing group of selected users. In [47] and [48], selection is based on pair-wise metrics derived from the correlation $|\langle \mathbf{h}_k, \mathbf{h}_l \rangle|$ and the cosine $|\langle \mathbf{h}_k, \mathbf{h}_l \rangle| / (\|\mathbf{h}_k\| \|\mathbf{h}_l\|)$ between channel vectors respectively, where $\mathbf{h}_k$ is the k^{th} row of the channel matrix. As expected, the correlation metric does not perform well under both homogeneous and heterogeneous channels since a low gain channel will be confused as having low correlation. The cosine metric does better but still cannot discriminate a high gain channel from a low one. To cater for heterogeneous channels, a related scheme in [49] based on the squared-normalized inner products (SNIP) $|\langle \mathbf{h}_k, \mathbf{h}_l \rangle| / (\|\mathbf{h}_k\|^2 \|\mathbf{h}_l\|^2)$ provides better performance by introducing an inverse proportionality to the channel gain.

To reduce search complexity, a hybrid scheme comprising both orthogonal complement projection and pair-wise cosine-based metrics has been proposed in [26]. By setting a threshold, pair-wise cosine metric comparison with the last chosen candidate is done to reduce the number of potential candidates for consideration. Only candidates with metrics below the threshold are entered for the next round of evaluation via orthogonal complement projection, which caters for heterogeneous channels. In [50], selection is based on repeated use of TCIBF pre-coding with rate evaluation to search for the next best user. In [44], a simple selection scheme based on the user channel gains was also proposed. This is commonly referred to as norm-based selection (NBS) and is simply done by ranking all users according to their channel gains. Simulation results in Section 4.4 show that algorithms using NBS and pair-wise metrics have poorer sum rate maximization performance. In addition, they fail to schedule M users for a large fraction of time.

All schemes in [26], [44] – [50] implement incremental selection, i.e., identifying the next best user at each step. All except [47] and [48] are suited for handling both homogeneous and heterogeneous channels. All operate on the entire pool of K users and hence require the channel state information of all users at the base station, which incurs significant overhead when K is large. Noting that the final active user subset cardinality is $K_r \leq M$, the stopping criterion of the schemes in [26], [44] – [50] comes in one of two ways. In [44], [46], [48] and [50] TCIBF rate evaluation is performed at each step and the selection process stops whenever inclusion of any remaining users causes the sum rate to drop, or when $K_r = M$ is reached. Others like [26] postpone the TCIBF rate evaluation until $K_r = M$ users are first chosen. TCIBF rate evaluation is then performed and the final active user subset that achieves the best sum rate for a particular channel realization is identified.

Next, we consider another class of user selection algorithms by drawing from single-user MIMO links where a major limiting factor is the cost of multiple analog RF chains. Since antenna hardware costs much less than the analog RF chains, a viable way of achieving a large fraction of the MIMO channel capacity is to adaptively select the best-antenna subset from a pool of antennas for coupling to a smaller pool of RF chains. It has been shown that the achievable diversity order through antenna subset selection is the same as that of the full system [34]. Receive antenna selection (RAS) algorithms that strives to maximize the subset channel capacity have been proposed, e.g., in [34] and [55]. The antenna selection process runs parallel to that of user selection encountered in multi-user MIMO systems. In fact, we note from [54] that the multi-user sum capacity is upper-bounded by its equivalent single-user MIMO capacity, i.e., where all receiver antennas can cooperate. This motivates the consideration of receive antenna selection (RAS) algorithms that were developed for single-user MIMO systems, for the purpose of user selection in a multi-user setting. We will evaluate the performance of incremental selection algorithms (IRAS) in [34] and [55] and decremental selection algorithms (DRAS) in [34] and [56]. The algorithms in [34] and [55] are based on the single-user MIMO capacity expression with equal power allocation (high SNR regime implied). The incremental selection algorithms identifies the next antenna whose inclusion maximizes the capacity gain while the decremental selection algorithms identifies the next antenna whose removal minimizes the capacity loss. Two pair-wise approaches are proposed in [56]. The first is the

Table 4.2: User Selection Algorithm Complexity Order Comparison

ALGORITHM	COMPUTATIONAL COMPLEXITY ORDER
Receive Antenna Selection (RAS) Based	
Decremental RAS [33]	$\in \boldsymbol{O}(K^2M^3)$
Incremental RAS [33]	$\in \boldsymbol{O}(KM^3)$
Incremental RAS [34]	$\in \boldsymbol{O}(KM^2)$
Mutual information based method (Decremental RAS based on pair-wise metrics) [35]	$\in \boldsymbol{O}(K^2M)$
Correlation based method (Decremental RAS based on pair-wise metrics) [35]	$\in \boldsymbol{O}(K^2M)$
Orthogonal Complement Projection Based (Incremental selection)	
Tu's algorithm [21]	$\in \boldsymbol{O}(KM^3)$
Berenguer's algorithm [20]	$\in \boldsymbol{O}(KM^3)$
Jiang's algorithm [22] (similar to [21])	$\in \boldsymbol{O}(KM^3)$
Yoo's algorithm [16] – hybridized with cosine-metric	Upper bounded by $\in \boldsymbol{O}(KM^3)$
Based on Repeated Pre-coding	
Dimic's incremental selection algorithm [26] Note that sum rate evaluation is already done.	$\in \boldsymbol{O}(KM^3)$
Pair-wise Metrics	
Squared-normalized inner product [25]	$\in \boldsymbol{O}(K^2M)$
Cosine-based incremental selection [24]	$\in \boldsymbol{O}(K^2M)$
Spatial compatibility metric [23] (similar to CBM in [35])	$\in \boldsymbol{O}(K^2M)$
Frobenius Norm Based	
Norm based selection [20]	$\in \boldsymbol{O}(KM)$

"Correlation Based Method" (CBM), which is based on a pair-wise correlation metric like [47] and as expected, it does not perform well. The second is named "Mutual Information Based Method" (MIBM), which is based on removing an antenna that has maximum pair-wise mutual information with the other antennas.

To compare algorithm complexities, we may compare (i) the number of decision metrics needed and (ii) the computational complexity. For exhaustive search, the decision metric is just the sum rate and the number of decision metrics needed is $\sum_{i=1}^{M}\binom{K}{i}$. For algorithms that depend on orthogonal complement projection, equivalent single-user capacity or "greedy" TCIBF rate evaluations, the number of decision metrics is $\boldsymbol{O}(MK)$ because most examine the remaining users in turn with respect to a currently chosen subset. For algorithms that depend on pair-wise metrics, the number of decision metrics can be reduced from $\sum_{K'=1}^{K}\sum_{i=1}^{K'} i$ to $\boldsymbol{O}(MK)$ because of symmetry and metric re-use for each selection stage. For norm-based user selection, the number of decision metrics is the lowest at *K*. Next, the computational complexities of the algorithms in [26], [44] – [50], [34] – [56] are given in Table 4.2, based on the complexity orders given in those papers or extrapolated from similar papers. Numerical results show that the IRAS algorithm in [55] provides a balance of good performance with lower implementation complexity compared to the best user selection algorithms in [26], [44] – [50].

4.2.2 Incorporating JREUS

Although the JREUS algorithms cannot be used when $K > M$, they can improve the performance of the algorithms in [26], [44] – [49], [34] – [56] when used in tandem with them. The general arrangement is to pre-select a subset S_r with $|S_r| = K_r = M$ users using an algorithm from [26], [44] – [49], [34] – [56] without TCIBF rate evaluation. A JREUS algorithm is then invoked during rate evaluation to select the final active subset $S_s \subset S_r$ that maximizes the sum rate, where $|S_s| = K_s \leq M$. Rate evaluation is not done when pre-selecting the $K_r = M$ users and this approach will not incur significant additional complexity when $|S_s| \to M$ for a large, geographically distributed user pool. Numerical results in Section 4.4 show significant sum rate improvement for the poorer performing algorithms e.g., those based on norm lengths and pair-wise metrics, and marginal improvement for the near-optimal algorithms like [50], [34] and [55]. Additionally, JREUS helps in scheduling more users for a larger fraction of time for algorithms with poorer sum rate performance. Compared to the case without JREUS however, it also reduces the likelihood of scheduling the maximum of *M* users.

4.2.3 Simultaneous Scheduling and Sum Rate Maximization

When the user pool is large, the user selection algorithms in [26], [44] – [49], [34] – [56] will have a higher chance of scheduling the maximum number of *M* users when performing sum rate maximization [18], [24]. However, this is done with different degrees of success and algorithms with poorer sum rate performance tend to schedule fewer users as well. Recalling that at least M users or channels must be served in order to reap the full benefits of multi-user diversity (MUD) arising from a large user pool [18], we propose a method that strives toward simultaneous spatial multiplexing and sum rate maximization. When a JREUS algorithm drops users from the initial list of M users, the original user selection algorithm, e.g., those in [26], [44] – [49], [34] – [56] may be invoked again to choose the next $M-K_s$ user(s) for consideration. Selection is done against the current group already chosen via JREUS. The process stops when inclusion of the next user causes sum rate reduction. Algorithms using pair-wise or channel-gain metrics can re-use the metrics evaluated during the first round of selection. This is not so for projection- or single-user capacity based algorithms where fresh metric computations must be done. The process is iterated in a greedy fashion until $K_s = M$ users are chosen. However, there is still a possibility that a solution with $K_s = M$ may not exist, i.e., when the remaining users cause sum rate reduction. For ease of reference, we will refer to this process as "scheduling and rate maximization" or SRM. In effect, SRM provides an avenue for better MUD exploitation by relying on JREUS to remove poorer performing users from the initial selection, which creates room for the consideration of better users. The SRM process can be iterated greedily and the results show convergence on the scheduling of M users with high probability. Thus SRM enables the algorithms to scale to $M \log\log KN$, given in [18]. As expected, the impact of SRM is more pronounced in algorithms that are based on norm lengths and pair-wise metrics, while improvements for the near-optimal algorithms like [50], [34] and [55] are marginal. To highlight, the norm-based scheme (NBS) from [44] is compared with the near-optimal scheme in [50] for homogeneous channels. The percentage difference in ergodic sum rate is ~28% without JREUS, ~20% with JREUS-MAX and ~13% with JREUS-MAX-SRM. The difference narrows to ~9% using JREUS-MAX-SRM with heterogeneous channels. Simultaneously, the SRM procedure significantly increased the probability of scheduling M users in NBS. To reduce complexity, the recursive inverse algorithm in [50], which operates on an incremental basis, can be used during SRM. Hence, using JREUS-MAX-SRM in tandem with norm-based user selection improves its feasibility for practical deployment while keeping the complexity level close to its original of $\boldsymbol{O}(K)$ decision metric evaluations.

A less complex SRM procedure may be done via an approximation where the original user selection algorithm, e.g., those in [26], [44] – [49], [34] – [56] is used to rank and pre-select $M+\psi$ users. The approximation arises since ranking of the additional ψ users is only done against the first chosen set of M users. The additional ψ users will be considered whenever a JREUS algorithm drops users from the initial list of M users. The magnitude of ψ is dependent on the strength of the algorithm employed and expected to be small for near-optimal algorithms. However, the pre-selection of additional ψ users may present an uncertainty that can be avoided as follows. Algorithms implementing decremental selection are "SRM-ready" since they eliminate the least favourable user at each turn and have therefore ranked the entire user pool except for the chosen subset S_r. The IRAS algorithms in [34] and [55] actually rank all remaining users according to their capacity-gain contribution and the last round of ranking may be used with SRM. This idea can be extended to those using orthogonal complement projections where the last round of projection-gain ranking can be used. Most algorithms with pair-wise metrics would have pre-computed the metrics that could be re-used to rank the remaining users. Finally, the ranking of all users for norm-based algorithms is straightforward. The additional users needed when using SRM with NBS is not expected to be high since the best user subset is likely to be found among users with high channel gains, as pointed out in [24]. The approximate SRM method is used when generating the numerical results in Section 4.4.

4.2.4 Reducing CSI Feedback Requirement during User Selection

Attention is given next to reducing the channel state information (CSI) required at the base station during user selection and the challenge includes avoiding significant sum rate loss. The approach adopted here is guided by [18] where it was shown that the optimal beamforming sum rate scales as $M \log\log KN$ and by the intuitive explanation provided in [24]. They lead to the notion of restricting CSI feedback to those users with the highest channel gains because the likelihood of finding a subset of M users that are roughly orthogonal from among them is high when $K \gg 1$. We refer to this as norm-based CSI feedback reduction (NB-CSIFR) algorithm where a subset of users $S_g \subset S$ with the best channel gains provides CSI feedback to the base station. NB-CSIFR can be accomplished simply via thresholding schemes, i.e., only users with channel gains exceeding a pre-determined level are required to provide CSI feedback. A subset $S_r \subset S_g$ with $|S_r| \leq M$ can then be selected using an algorithm of choice. A similar proposal for reducing CSI feedback requirement was made recently in [49]. Numerical results show insignificant loss in the multi-user sum capacity with NB-CSIFR for $K \gg 1$ even when thresholding has limited feedback to about 38% of the user pool. This is also true for the sum rates achieved by the algorithms in [26], [44] – [50], [34] and [56].

4.2.5 Scheduling Fairness

Scheduling fairness is not an issue for homogeneous channels because all users have equal throughput shares over the long term since they are statistically identical. This is not the case for heterogeneous channels due for example to a near-far situation. In [26], two fair scheduling schemes were proposed, viz., round-robin TCIBF (RR-TCIBF) and proportional fair TCIBF (PF-TCIBF). Theoretical background on the utilization of proportional fair scheduling in wireless systems can be found in [72]. Given space limitations, we briefly highlight that for RR-TCIBF, any near-optimal user selection algorithm will ensure near orthogonality within each group of chosen users and provide performances similar to that in [26]. This is not so for the poorer performing algorithms like those based on NBS or pair-wise metrics and their performance can be improved by the incorporation of JREUS and SRM during the selection of each time-multiplexed group.

Next, since RAS algorithms in [34] and [55] were introduced for user selection, we will give an outline on their adaptation for PF-TCIBF. Beginning with incremental selection, (4.12) shows the single user MIMO capacity expression when one additional row vector $\mathbf{h}_l$ is appended to $\mathbf{H}_{s\{1,\cdots k\}} \in \mathbb{C}^{k \mathrm{x} M}$, which is a channel matrix comprising row vectors $\mathbf{h}_1$ to $\mathbf{h}_k$ that are associated with the current chosen subset S_s.

$$C\left(\mathbf{H}_{s\{1,\cdots k\}};\mathbf{h}_l\right)=\underbrace{\log_2\det\left(I_N+\mathbf{H}^H_{s\{1,\cdots k\}}\mathbf{H}_{s\{1,\cdots k\}}/\sigma^2\right)}_{\text{Capacity of current chosen subset } S_s} \\ +\underbrace{\log_2\left(1+\mathbf{h}_l\left(\sigma^2 I_N+\mathbf{H}^H_{s\{1,\cdots k\}}\mathbf{H}_{s\{1,\cdots k\}}\right)^{-1}\mathbf{h}_l^H\right)}_{\Delta C_l\text{ : Additional capacity due to appended row vector }\mathbf{h}_l}. \tag{4.12}$$

In IRAS, the next chosen antenna is based on $\arg\max_l\left(\Delta C_l\right)$, where ΔC_l is defined in (4.12). For PF-TCIBF, we propose choosing the next user based on $\arg\max_l\left(\mu_l \Delta C_l\right)$, where weight μ_l is the usual inverse of the time-averaged past throughput for user l as defined for example in [26]. This method is easily extended for decremental selection where the next user to be deleted is based on $\arg\min_l\left(\mu_l \Delta C_l\right)$ to minimize the capacity loss. We note also that the weights μ_l can be used for the purpose of CSI feedback reduction, i.e., PF-CSIFR, by setting a threshold value for $\mu_l R_l$, where $R_l \triangleq \log_2(1+\|\mathbf{h}_l\|^2)$.

4.3 Resource Allocation Against QoS Requirements

Up to this point, the emphasis has been on the judicious selection of users for TCIBF sum rate maximization. Since CSI is already available at the base station, waterfilling may also be done to maximize the sum rate of the chosen subset. At this stage however, the individual channel rates are not matched to each user's QoS needs. For some users, the instantaneous channel rates may be in excess of what they need, while others may be in deficit. Preliminary proposals to deal with this issue are given in the following section. These resource allocation proposals are guided by the objective of minimizing the sum rate loss. However, we do not have numerical results at this stage.

4.3.1 Power Allocation

A straightforward idea is to reduce the power allocated to users with excess channel rates and re-distribute the power savings to those in need. This is carried out after the initial power allocation, which is done via waterfilling. To minimize sum rate loss, the re-distribution should be guided by the waterfilling principle, i.e., allocate more power to the better channels. Hence, the excess power will be given to the next best user and any leftover power will be given to the next best user. The process is repeated until all QoS requirements are satisfied or when no excess power is left. When the latter case occurs, there may be one or more users that still do not achieve a channel rate that match their QoS requirements. We may consider the following proposal when more than one user does not have adequate power allocation.

4.3.2 Dropping Users

Essentially, users with very poor channel rates may be dropped in the interest of helping others in the subset. This may be done when power allocation methods fail to yield a solution, especially when more than one user does not have adequate power. In this case, removing the worst user may benefit the remaining users. Rate improvements for the remaining users come not only from the power saved, but also from better null space projections that result in higher projected channel gains.

The initial user subset may have been chosen using an incremental or decremental user selection scheme. For incremental selection, the chosen subset would have already been ranked and de-selection of the next worse user is easily accomplished. However, the de-selection accuracy is dependent on the user selection algorithm previously used. Algorithms that are based on channel gains or pair-wise metrics will give poorer de-selection performance. This may be improved by using JREUS-based algorithms, which have been shown to yield near-optimal results when the user set size $K_r \leq M$. In particular, JREUS-MAX will incur little additional complexity for user de-selection since rate evaluation is needed in any case. For the decremental selection case, the chosen subset is still not ranked. Hence a JREUS-based algorithm should be employed to incur the least additional computational complexity since rate evaluation is needed.

Upon the de-selection of one user, TCIBF pre-coding and waterfilling is done for the new subset. The power allocation scheme in Section 4.3.1 is invoked and the entire process is repeated until a solution is found.

4.3.3 Adding Users

This is possible for the case where $< M$ users are chosen and each user has a channel rate that exceeds its QoS requirement. Selection of the next user is easier in this case because both incremental and decremental user selection algorithms would have ranked all users that are *not* already chosen in the original subset. Using this ranked list, the next user can be included for TCIBF pre-coding and rate evaluation. The power allocation method in Section 4.3.1 may then be invoked and inclusion of the new user will depend on its impact to the original subset. The process may be repeated until the ability to meet the QoS requirements of the most current subset is breached. Note that the last user added may have only attained a fraction of its QoS requirement.

Again, the ranking done by algorithms with poorer performance may be improved by using JREUS algorithms. In this case, the situation is similar to Section 4.2.3 where simultaneous scheduling and sum rate maximization was discussed. To re-iterate, JREUS can be used on the chosen subset to help maximize the sum rate if it was not already done. This process may drop some users and create room for the consideration of more users. The SRM-style procedure may be adopted, where the power allocation method in Section 4.3.1 is carried out for each new user. This process is iterated greedily until the ability to meet the QoS requirements of the most current subset is breached.

4.4 Impact of Transmit Antenna Selection (TAS)

It is known that transmit antenna selection (TAS) methods provide diversity benefits through the provision of more transmit antennas, beyond the required M transmit-chains. This is applicable to both single-user as well as multi-user MIMO systems. It is also clear from [18] that TAS is not useful for *fully* equipped systems where all transmit antennas are accompanied by a RF chain. This is because the multi-user sum rates of optimal DPC and optimal beamforming scale as $M \log\log KN$ and reducing M will reduce the sum rate. We will show that this is true when optimal user selection (USEL) has been done to maximize the TCIBF sum rate but *not* always true when *sub*-optimal USEL algorithms are employed.

4.4.1 Impact of TAS on TCIBF Sum Rate

In line with Section 4.1, we assume that a subset S_r has been chosen from S the pool of K users using an exhaustive search or USEL algorithm. The number of chosen users is $|S_r| = 1 \leq K_r \leq M$ and the associated composite channel matrix is $\mathbf{H}_r \in \mathbb{C}^{K_r \mathrm{x} M}$. Next, we define an arbitrary user subset $S_s \subseteq S_r$ that is the result of applying TAS and perhaps further user de-selection. It has $|S_s| = 1 \leq K_s \leq K_r$ users and the associated composite channel matrix is $\mathbf{H}_s \in \mathbb{C}^{K_s \mathrm{x} M'}$, where $M' = M - \sigma$ and σ represents the number of transmit antennas removed. In addition, $K_s \leq M'$ so that the TCIBF pre-coding constraint is met. Waterfilling is done over S_s so that S_a the final active user subset of K_a users is found. Considering only those row eliminations that result in non-zero power allocations, i.e., $\gamma_i > 0$, $\forall i$ in (4.3), i.e., $K_a = K_s$ and (4.4) applies. Although changing M' may affect K_a in (4.4), it will be shown that the analysis here still applies. The analytical results will be applicable to any search method, viz., joint exhaustive USEL-TAS search, decoupled exhaustive USEL-TAS search and any sub-optimal USEL/RAS algorithm.

To evaluate the effects of removing column vectors from $\mathbf{H}_s$, we assume $K_s < M'$ without loss of generality. Let $\mathbf{H}_k$ represent a channel sub-matrix with one column removed from $\mathbf{H}_s$. Sub-matrix $\mathbf{H}_k$ will be K_s x$(M'-1)$ where $K_s \le (M'-1)$ and $\mathbf{H}_k\mathbf{H}_k^H$ still has dimensions K_s x K_s. Hence (4.4) still applies and the cofactors and determinant are now notated as $(A_{ii})_k$ and Δ_k. Let the singular values of $\mathbf{H}_k$ and $\mathbf{H}_s$ be $\sigma_{\max}(\mathbf{H}_k) \ge \sigma_2(\mathbf{H}_k) \ge .... \ge \sigma_{\min}(\mathbf{H}_k)$ and $\sigma_{\max}(\mathbf{H}_s) \ge \sigma_2(\mathbf{H}_s) \ge .... \ge \sigma_{\min}(\mathbf{H}_s)$ respectively. With $K_s < M'$ in $\mathbf{H}_s$, we note that [67]

$$\sigma_{\max}(\mathbf{H}_s) \ge \sigma_{\max}(\mathbf{H}_k) \ge \sigma_2(\mathbf{H}_s) \ge \sigma_2(\mathbf{H}_k) \ge \cdots\cdots \ge \sigma_{\min}(\mathbf{H}_s) \ge \sigma_{\min}(\mathbf{H}_k) \quad (4.13)$$

Although equalities are present in (4.13), note that

$$\left(\det(\mathbf{H}_k\mathbf{H}_k^H) = \Delta_k\right) < \left(\det(\mathbf{H}_s\mathbf{H}_s^H) = \Delta_s\right) \quad (4.14)$$

$$\text{because} \qquad (z_{ii})_k < (z_{ii})_s \quad \forall\, i = 1, \cdots K_s, \quad (4.15)$$

where $(z_{ii})_k$ and $(z_{ii})_s$ are the principal diagonal elements of $\mathbf{Z}_k = \mathbf{H}_k\mathbf{H}_k^H$ and $\mathbf{Z}_s = \mathbf{H}_s\mathbf{H}_s^H$ respectively. Equation (4.15) is true because $\mathbf{H}_k$ has one column less than $\mathbf{H}_s$ and hence $\text{trace}(\mathbf{Z}_k) < \text{trace}(\mathbf{Z}_s)$. This means that

$$\sum_{i=1}^{K_s} \lambda_i(\mathbf{Z}_k) < \sum_{i=1}^{K_s} \lambda_i(\mathbf{Z}_s), \quad (4.16)$$

so that some inequalities must exist in (4.13) and hence (4.14) holds. At high SNR levels, it is clear from (4.14) that $R(\mathbf{H}_k) < R(\mathbf{H}_s)$ using (4.4). At low SNR levels however, the outcome of (4.4) is not so straightforward and a comparison of $(A_{ii})_s$ and $(A_{ii})_k$ is needed before any conclusions could be made.

Let an arbitrary matrix $\mathbf{H}_x \in \mathbb{C}^{m\mathrm{x}n}$ where $m \le n$ have cofactors $\{A_{jj}; j = 1 \cdots m\}_x$, associated with the diagonal elements of $\mathbf{Z}_x = \mathbf{H}_x\mathbf{H}_x^H \in \mathbb{C}^{m\mathrm{x}m}$. From [67]

$$\prod_{i=1}^{m-1}\lambda_i(\mathbf{Z}_x) \le \det\left((\mathbf{Z}_x)_{m-1}^j\right) \le \prod_{i=1}^{m-1}\lambda_{1+i}(\mathbf{Z}_x) \tag{4.17}$$

$$\Rightarrow \prod_{i=1}^{m-1}\lambda_i(\mathbf{Z}_x) \le \{A_{jj}\}_x \le \prod_{i=1}^{m-1}\lambda_{1+i}(\mathbf{Z}_x) \tag{4.18}$$

where $(\mathbf{Z}_x)_{m-1}^j$ are the $(m-1)\mathrm{x}(m-1)$ principal sub-matrices of $\mathbf{H}_x\mathbf{H}_x^H$ that are associated with the set of cofactors $\{A_{jj}\}_x$. Note that (4.18) is true because of the Inclusion Principle (see Appendix) where the eigenvalues of $\mathbf{Z}_x$ do not change when the corresponding rows and columns are interchanged to find the $(\mathbf{Z}_x)_{m-1}^j$ that is associated with its $\{A_{jj}\}_x$. Applying (4.18) on $\mathbf{Z}_k$ and $\mathbf{Z}_s$, we first see via (4.13) that

$$\prod_{i=1}^{K_s-1}\lambda_i\left(\mathbf{Z}_k\right) < \prod_{i=1}^{K_s-1}\lambda_i\left(\mathbf{Z}_s\right) \quad \text{and} \tag{4.19}$$

$$\prod_{i=1}^{K_s-1}\lambda_{1+i}\left(\mathbf{Z}_k\right) < \prod_{i=1}^{K_s-1}\lambda_{1+i}\left(\mathbf{Z}_s\right). \tag{4.20}$$

With (4.18), (4.19) and (4.20) in mind, we see from (4.1) that

$$\frac{1}{\lambda_{\max}(\mathbf{Z}_k)} \le \left\{\frac{1}{b_i}\right\}_k \le \frac{1}{\lambda_{\min}(\mathbf{Z}_k)} \tag{4.21}$$

and

$$\frac{1}{\lambda_{\max}(\mathbf{Z}_s)} \le \left\{\frac{1}{b_i}\right\}_s \le \frac{1}{\lambda_{\min}(\mathbf{Z}_s)}, \tag{4.22}$$

where $i = 1, \ldots\ldots, K_s$. Since $\lambda_{\max}(\mathbf{Z}_k) \le \lambda_{\max}(\mathbf{Z}_s)$ and $\lambda_{\min}(\mathbf{Z}_k) \le \lambda_{\min}(\mathbf{Z}_s)$, it can be implied that one or more $(1/b_i)_k$ values are greater than $\max(1/b_i)_s$ values. It is also clear that $\min(1/b_i)_k > \min(1/b_i)_s$. It is then clear from the power constraint expression in (2.3) that the set of $\{\gamma_i\}_k$ values contains some elements that are lower than $\min(\gamma_i)_s$. It is also clear that $\min(\gamma_i)_k < \min(\gamma_i)_s$. Hence $R(\mathbf{H}_k) < R(\mathbf{H}_s)$ via (2.2) and this is true for *any* SNR and channel condition as reflected by the sub-matrix determinants and cofactors of $\mathbf{H}_k$

and $\mathbf{H}_s$. Even though the above analysis was done for the one-column de-selection case, it is obvious from (4.19) – (4.22) that the results apply to any number of columns removed from $\mathbf{H}_s$ as long as $K_s < M'$ is true, prior to each step of column de-selection (note that TAS cannot proceed in TCIBF when $K_s = M'$). Since this condition meets the TCIBF pre-coding constraint, we can draw the first conclusion that reducing the number of transmit antennas for any (K_s, M') antenna combination will *not* increase the TCIBF sum rate. The combination may arise from any joint search method (e.g., joint exhaustive TAS-USEL or de-coupled TAS-USEL), any USEL algorithm or during waterfilling.

Next, it can be readily inferred that the converse is true, i.e., increasing the number of transmit antennas for any antenna combination will increase the sum rate. A repeated application of this reasoning shows that any TCIBF system should utilize all available transmit antennas for sum rate maximization. Very importantly, this means that TAS is not useful in realizing the optimal TCIBF sum rate under any SNR level and any channel condition. Hence, an exhaustive joint- or exhaustive decoupled- USEL-TAS search is not needed when finding the optimal solution set for sum rate maximization; instead, only an exhaustive USEL search is needed.

4.4.2 Combining TAS with Sub-optimal User Selection Algorithms

When *sub-optimal* USEL algorithms such as JREUS or those in [44] – [49] are used to avoid exhaustive search, the chosen subset may be the result of following a search path that leads to a local maximum. It will be shown that TAS may help assist the USEL process in getting out of a local maximum when the optimal set is contained *within* the sub-optimal set. Fig. 4.1 is provided for visualization by depicting the singular value transition paths for an example with 6 users. Each transition row represents the removal of a row/column vector. When USEL is done, the transition paths can be understood from (2.9). Assume for the moment that an *optimal* USEL search resulted in a 3-user subset and their singular values are shown as triangles (upper set). Next, assume that a *sub*-optimal USEL algorithm has chosen a 4-user subset instead and their singular values are shown as

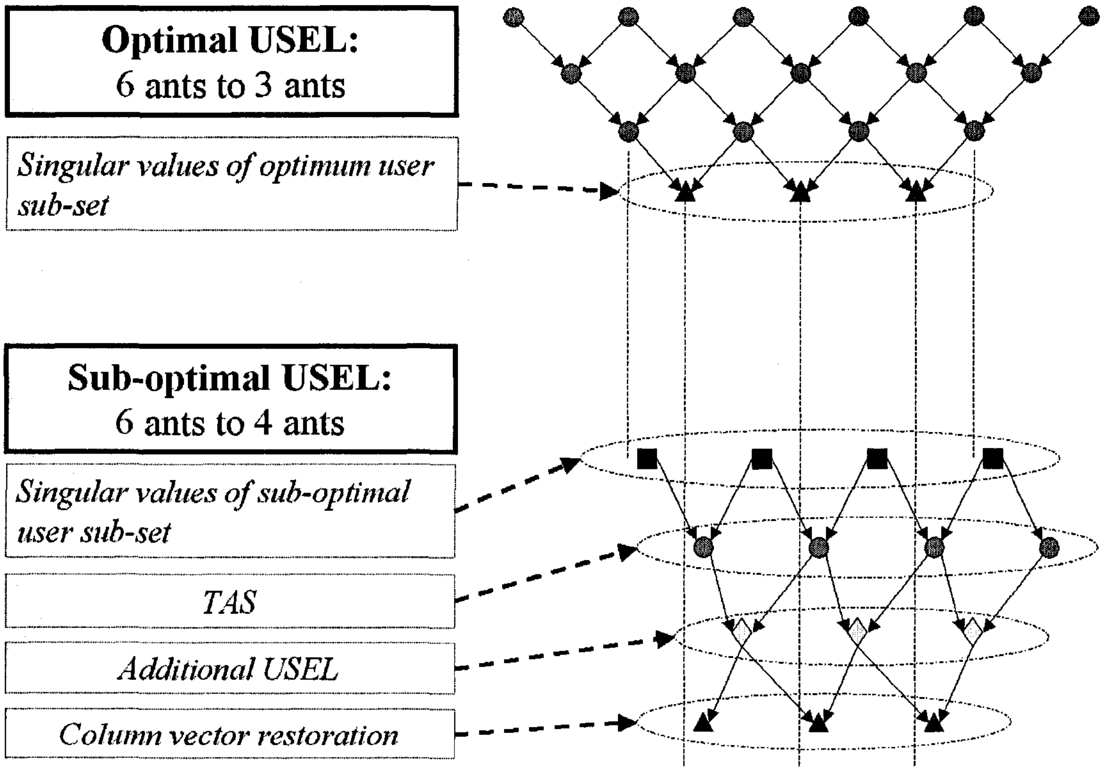

Optimal USEL:
6 ants to 3 ants
Singular values of optimum user sub-set
Sub-optimal USEL:
6 ants to 4 ants
Singular values of sub-optimal user sub-set
TAS
Additional USEL
Column vector restoration

Figure 4.1. Singular Value Transition Diagram

squares. Fig. 4.1 illustrates how the use of TAS followed by USEL could result in a set of singular values (diamond shape) that are *closer* to the optimum values (upper triangular set). The singular value transitions due to TAS can be understood from (4.13) and the subsequent transitions arising from USEL is again due to (2.9). Numerical results have shown that this procedure does result in sum rate gains in some cases, especially if the optimum subset is contained within the sub-optimal subset. Note that the USEL procedure following each TAS may contain more than one row de-selection. For the TAS stage however, it is clear from the previous section (Section 4.4.1) that no more than one column vector should be de-selected at a time. For ease of reference, we will name this procedure as DSEL (decoupled TAS-USEL) and more details are given in Section 4.4.3.

Next, it was shown in Section 4.4.1 that any TCIBF system should utilize all available transmit antennas for sum rate maximization. Drawing on this, we apply the converse by restoring the de-selected column vector back into the sub-matrix arising from a TAS-USEL cycle. For the example in Fig. 4.1, it will result in moving the singular values into a higher region (lower triangular set), which helps approach the optimal subset singular values (upper triangular set). We will refer to this algorithm as MDSEL (modified DSEL). Numerical results have confirmed that whenever DSEL provides a sum rate increase, restoring the removed column vector will *always* increase that sum rate further.

Note however that the expected gain from incorporating TAS is not high, especially when the USEL algorithm is already near optimal. For example, the JREUS algorithm is nearly optimal especially at high SNR and numerical results in the next section will show that the occurrence rate where TAS does make a difference is not high and the sum rate increase is also not high. This is to be expected because TAS does not help achieve the optimal sum rate when optimal USEL is done. This can be easily inferred from Fig. 4.1 where a TAS-USEL cycle done on the optimal selection will depart from the original optimal selection indicated as triangles. Next, it is clear from the above discussion that the scheme can be adapted for use with *any* USEL algorithm because the TAS process is essentially independent of the USEL algorithm utilized.

4.4.3 Guidelines on Sub-optimal Decoupled Search Strategies

Drawing from single-user MIMO systems where transmit antenna selection (TAS) was first proposed, the optimal antenna subset selection requires a joint exhaustive search of all possible transmit- and receive- antenna combinations. Various sub-optimal decoupled search methods have been proposed to avoid the computationally intensive joint exhaustive search. One example of decoupled exhaustive search is given in [79] where TAS is done before RAS. In general, algorithms such as those in [34] and [55] may be employed for further complexity reduction.

In TCIBF, any decoupled search algorithm must first take the pre-coding constraint into account, which will have impact on the search order. From Section 4.4.1, we note that performing TAS alone when using sub-optimal USEL algorithms is always not useful. Rather, TAS must always be followed by USEL. Next, TAS should generally not be used too early in any decoupled search strategy because it moves the singular values to a lower range as shown by (4.13). Hence strategies like one-step alternation between USEL and TAS should be avoided. In line with the TCIBF pre-coding constraint and with the fact that only USEL is needed for optimal search, it is expedient to perform any sub-optimal USEL algorithm first. As guided by Section 4.4.2, TAS is then performed next where a sequential elimination of each column vector coupled with USEL is done. This process may be repeated until the subset with the best sum rate is found. The procedure as outlined is the decoupled TAS-USEL or DSEL algorithm. Next, whenever DSEL results in sum rate gains, the column vector removed by TAS should be restored to give a higher sum rate (i.e., MDSEL is used).

4.5 Numerical Results

4.5.1 Comparing Algorithms for $K \leq M$

We focus on comparing the JREUS family with the near-optimal algorithms in [26], [44] – [31], [50] and [34] – [55]. Those based on pair-wise metrics or channel gains are excluded since they also require rate evaluation but do not provide good returns for the complexity involved. The algorithm in [45] is used to represent all orthogonal complement projection based methods. Fig. 4.2(a) shows the sum rates for $K = M = 8$ users when correlation among the users is zero in a homogeneous channel. The presence of spatial fading correlation among users is captured by modelling the channel as $\mathbf{H}_r = \mathbf{R}_u^{1/2}\mathbf{H}_w$, where $\mathbf{H}_w$ is the i.i.d. spatially white channel and $\mathbf{R}_u$ is positive definite Hermitian matrix that specifies the user correlations. An exponential correlation model is used where each element r_{ij} in $\mathbf{R}_u$ is $r_{ij} = \rho^{|i-j|}$, where ρ is the maximum correlation between two users. To facilitate examination, Fig. 4.2(b) shows the percentage sum rate difference of each algorithm compared to exhaustive search, i.e., $(R_{\text{Exh}} - R_{\text{Algo}}) / R_{\text{Exh}}$. As shown, the JREUS family performs well for $\text{SNR} > 10\,\text{dB}$ and thresholding for JREUS-ALT1 and JREUS-ALT2 helps when $\text{SNR} < 20\,\text{dB}$. Both JREUS-ALT1 and JREUS-ALT2 perform very well without thresholding for $\text{SNR} > 20\,\text{dB}$ whereas Dimić's algorithm [50] is worse off. This means that the consideration of all alternative subsets is not ideal as it raises the likelihood of solutions yielding local maximums. Note that JREUS-MAX has the same performance as Gorokhov's DRAS algorithm [34] at high

SNR. This is expected since each row removal by JREUS-MAX gives the next best MIMO capacity at high SNR when comparing subsets of the same cardinality. This is equivalent to minimizing the DRAS capacity loss. Similarly, the IRAS [34] and [55] performances are close to those of orthogonal complement projection at high SNR levels. Fig. 4.3 shows the percentage sum rate difference for $\rho = 0.50$ and $\rho = 0.95$. In summary, JREUS-MAX provides good performance when $K \leq M$ with low complexity in practical SNR and correlation ranges.

4.5.2 Comparing Algorithms for $K > M$

The sum rate maximization performance of the algorithms in [44] – [50] and [34], [55], and [56] are evaluated with and without CSIFR, JREUS and SRM. Fig. 4.4 shows the sum rate versus K for a system with 8 transmit-antennas, 20dB SNR, zero correlation among users and operating in a homogeneous channel without CSI feedback reduction. As shown, Dimić's algorithm performs best for large K whereas the DRAS algorithm in [34] performs best when $K \leq 32$. IRAS performs better for larger user pools than DRAS and is close to Dimić's. All algorithms benefit from JREUS-MAX and SRM, except for Dimić's algorithm for which they do not apply. The improvement is significant for the poorer performing algorithms. For example, the sum rate difference between norm-based selection (NBS) and Dimić's algorithm is ~28% when $K = 120$. This improves to ~20% with JREUS-MAX and ~13% with SRM. The approximated SRM scheme as outlined in Section 4.2.3 is used in all numerical results shown here. We also observe that the gap narrows between Dimić's sum rate and the multi-user sum capacity (derived using [73]) as K becomes large. Next, Fig. 4.5 shows that SRM is successful in maximizing the number of scheduled users while maximizing the sum rate simultaneously. The improvement is significant for the poorer performing schemes.

Fig. 4.6(a) shows the same system operating under a heterogeneous channel without CSIFR where the relative average-SNR levels among all users is varied uniformly in a 10dB range. As expected, the correlation-based algorithm performs poorly and the cosine-based algorithm performance registers a significant drop. Fig. 4.6(b) shows the same system with CSIFR where the threshold is set at 4dB, i.e., any user whose channel gain is 4dB below the average homogeneous channel SNR will be dropped. At this threshold, CSI feedback occurs for ~38% of the user pool on average. As shown, there is no significant loss of multi-user sum capacity or the algorithms' sum rates for $K \geq 40$ even though the effective user pool has shrunken. This affirms the approach outlined in Section 4.2.4. In addition, the correlation- and cosine-based algorithm, which includes the squared-normalized inner products (SNIP), improved significantly since thresholding performs pre-selection and results in fewer mistakes due to low gain users. The performance difference of NBS with JREUS-MAX-SRM has also narrowed to within ~9% of Dimić's algorithm with or without CSIFR. Given that NBS has a computational complexity reduction of $\boldsymbol{O}(M^2)$ compared to Dimić's algorithm, this result further improves the feasibility of employing NBS in practice. Fig. 4.7 compares the scheduling performance of NBS (with simultaneous sum rate maximization) versus Dimić's algorithm for CSIFR with thresholds of 3dB, 4dB and 5dB. As shown, the number of users scheduled for NBS is higher than Dimić's when JREUS-MAX-SRM is employed.

In summary, the incorporation of JREUS-MAX and SRM provides simultaneous improvement of sum rate and user scheduling performance, especially for the poorer performing algorithms. This improves the feasibility of employing the norm-based selection (NBS) algorithm in practice, which is attractive given its low complexity and natural fit for NB-CSIFR. Next, incremental RAS algorithms in [34] and [55] have demonstrated performance that is close to the best algorithm in [50]. The IRAS implementation in [55] has a lower complexity than [50] by a factor of M. This improves the feasibility of achieving performances above that offered by NBS when the potential user pool is large.

4.5.3 Sample Results for TAS with TCIBF

To verify the findings in Section 4.4.1, simulations are conducted on an 8-user TCIBF system with $M = 8$ transmit antennas. The results arising from joint exhaustive TAS-

USEL are *identical* to the case where *only* exhaustive USEL is done. This serves to confirm the analysis that TAS does *not* help in achieving the optimal TCIBF sum rate.

Next, simulations are done to illustrate the effect of TAS when *sub-optimal* USEL algorithms are used. The near-optimal JREUS algorithm is used for USEL. The maximum sum rate gain for DSEL over USEL alone (using JREUS) is 7.7%, 8.2%, 14.4% and 20.8% for correlations of 0.0, 0.2, 0.5 and 0.9 respectively. Fig. 4.8 shows a sample result when the correlation is 0.0. For most cases, the increase tends to occur around 0dB SNR except when correlation is high. The maximum sum rate gain for MDSEL over DSEL is 7.1%, 8.8%, 8.9% and 9.6% for correlations of 0.0, 0.2, 0.5 and 0.9 respectively. Fig. 4.9 shows the percentage of time that DSEL resulted in better sum rates than using JREUS alone. As expected, the occurrence rate is low, especially when SNR is high where JREUS performs better, except when correlation is high. The results show that the incorporation of TAS with sub-optimal USEL algorithms does have the potential for channel sum rate gains. However, when the USEL algorithm is already near optimal, the contribution from TAS is not high and the occurrence frequency is low. This is to be expected because TAS does not help achieve the optimal sum rate when optimal USEL is already done.

Fig. 4.2(a) TCIBF Sum Rate vs SNR

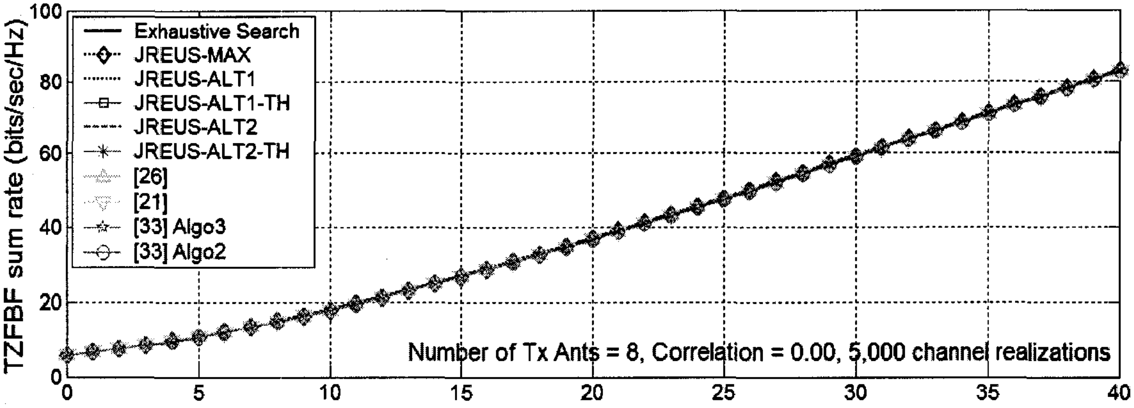


Fig. 4.2(b) TCIBF Sum Rate Difference (%) from Optimal vs SNR; Correlation = 0.00

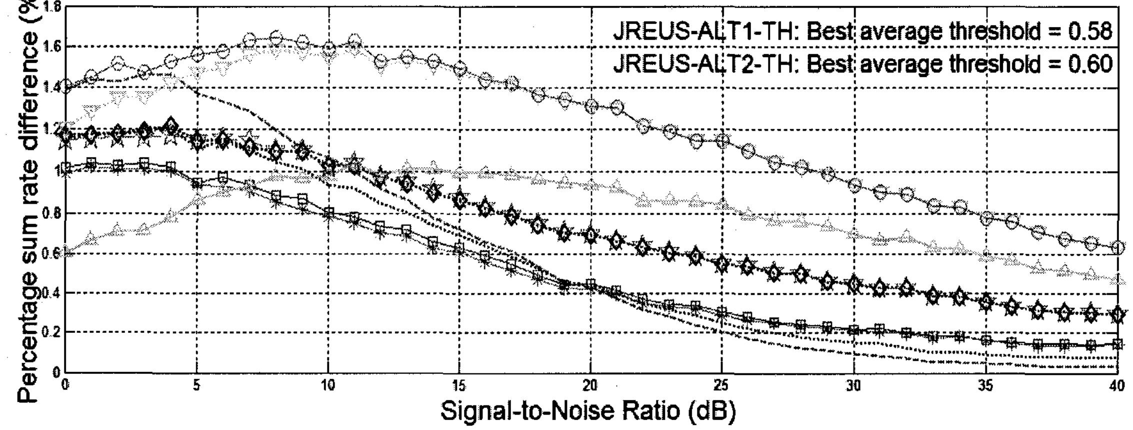


Fig. 4.3(a) TCIBF Sum Rate Difference (%) from Optimal vs SNR; Correlation = 0.50

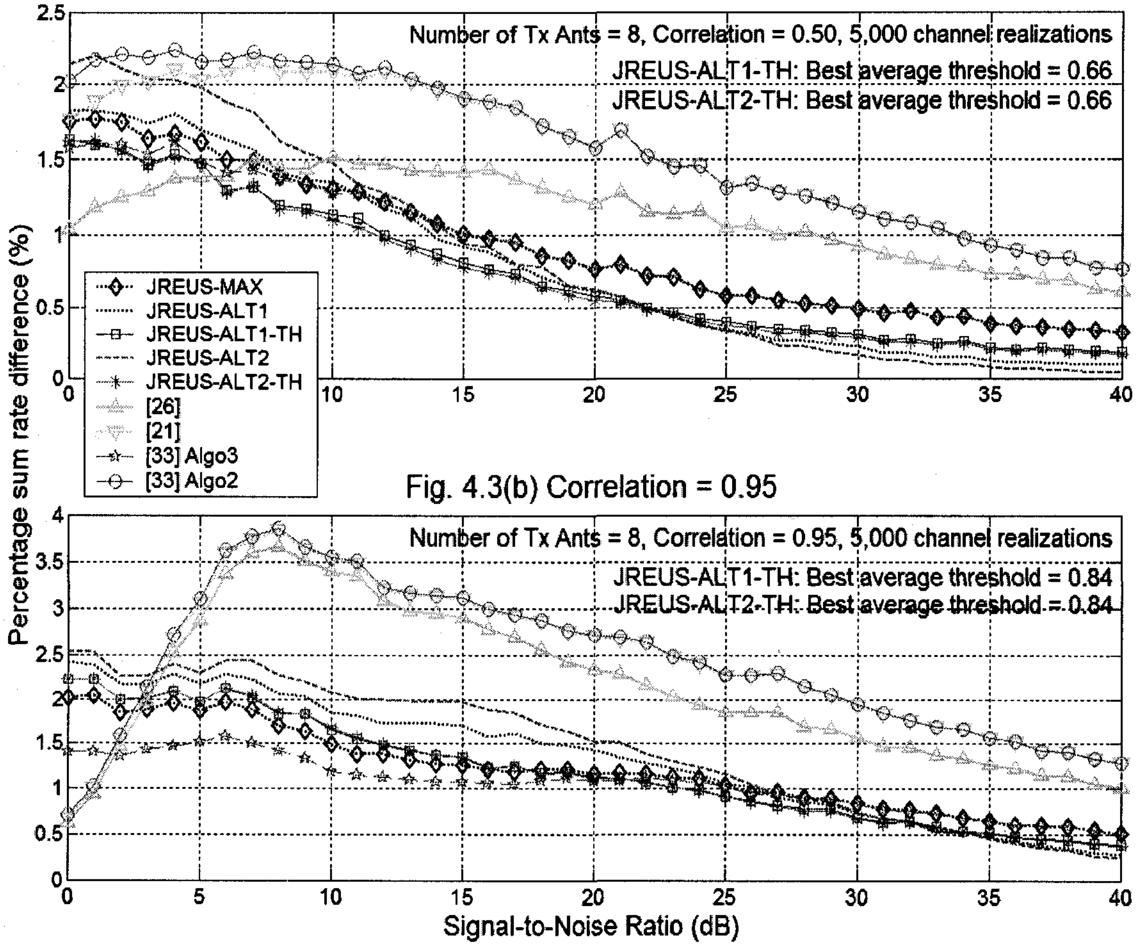


Fig. 4.3(b) Correlation = 0.95

Fig. 4.4 Sum Rate for Algorithms for K > M

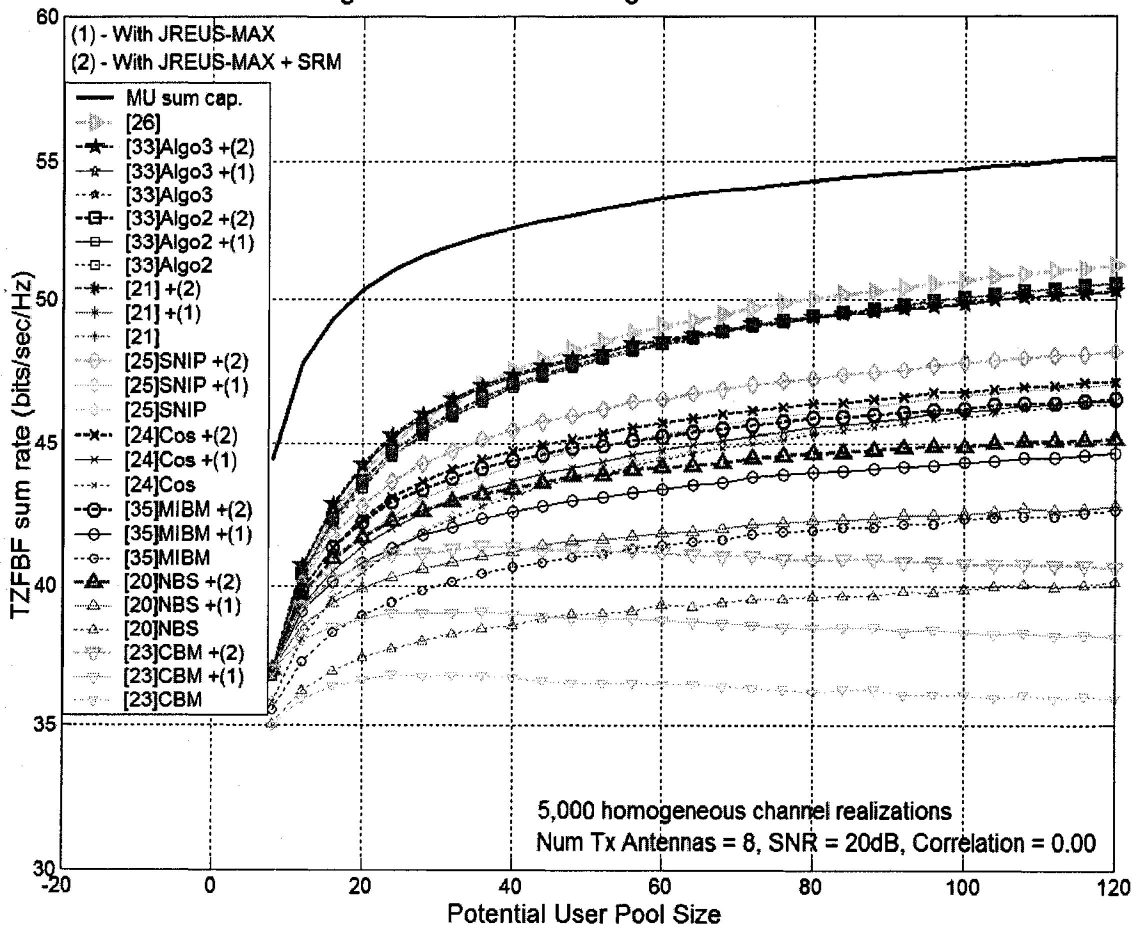
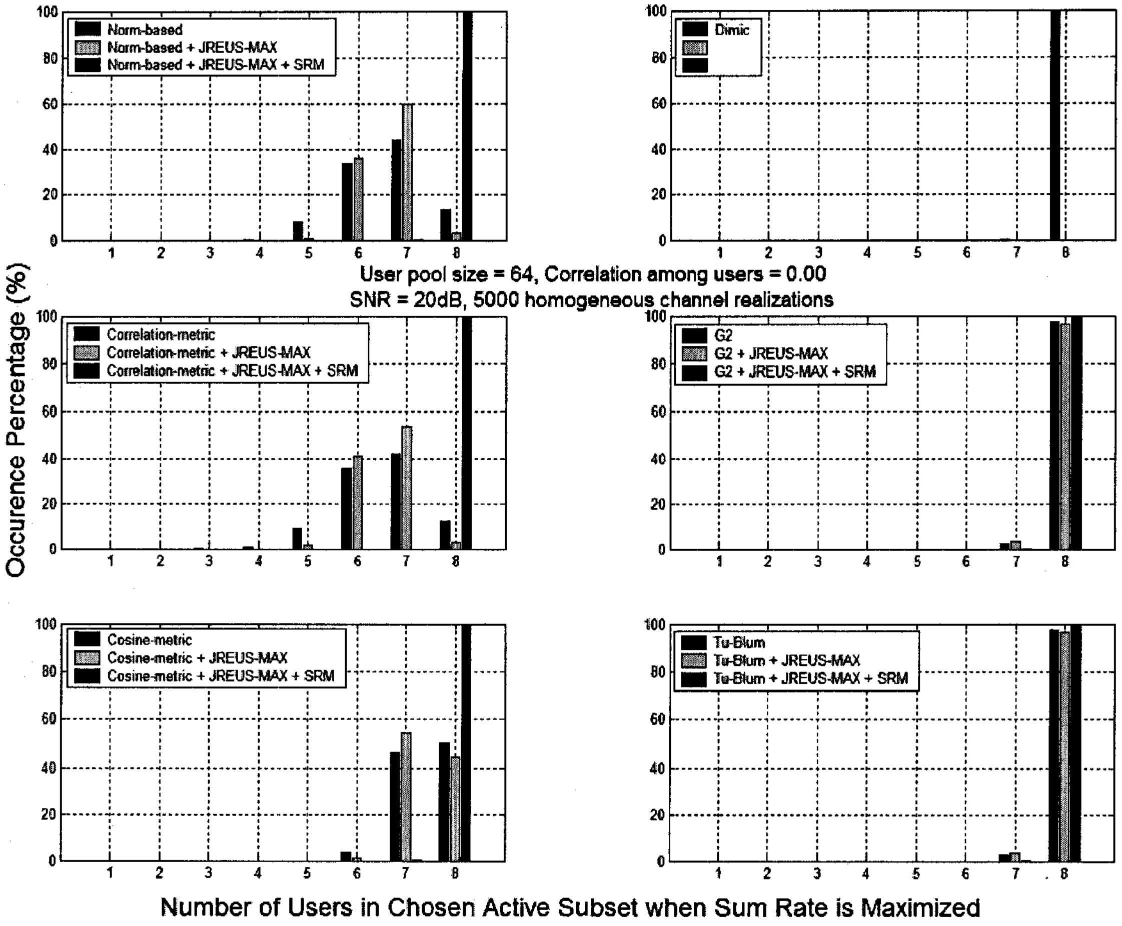


Fig. 4.5 Scheduling Maximization Performance (Homogeneous Channel)



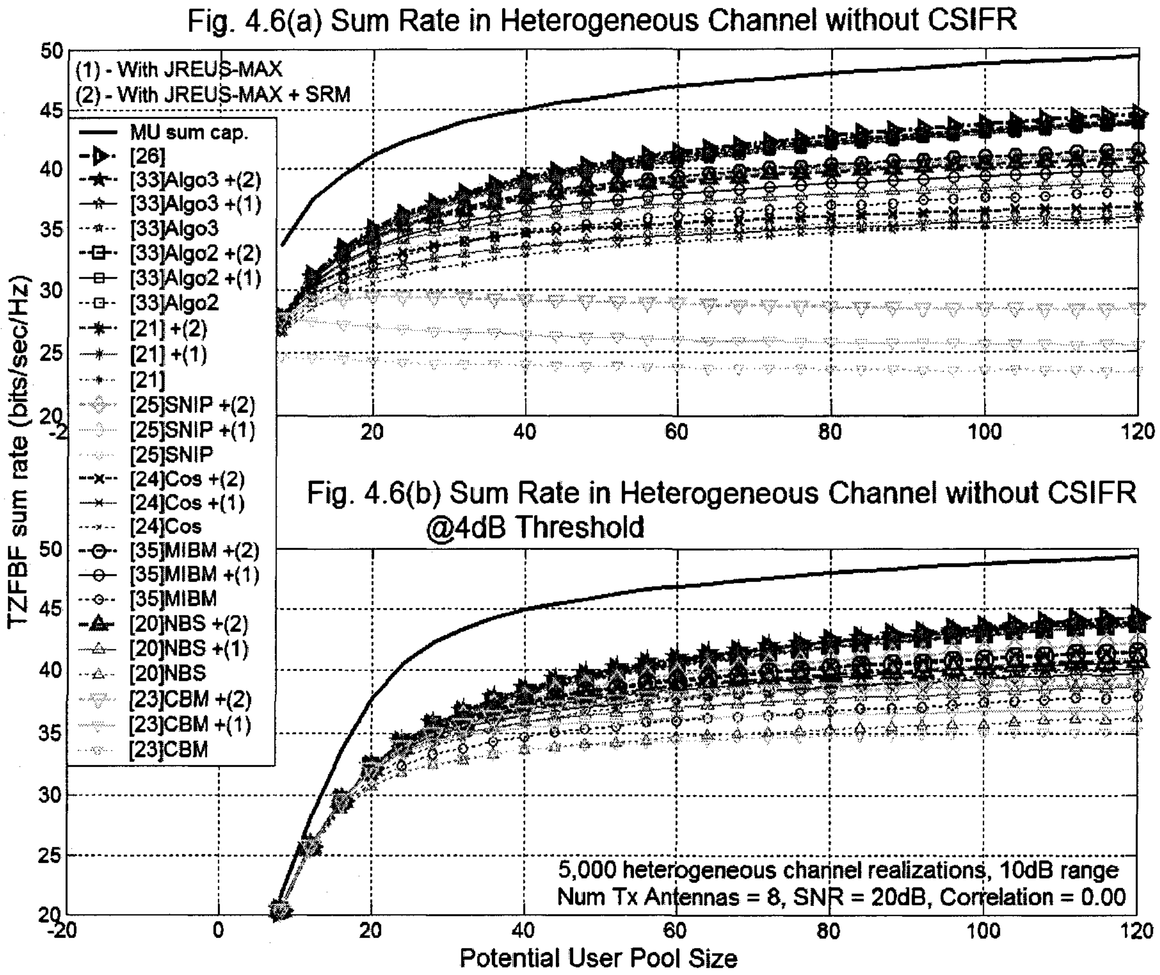


Fig. 4.6(a) Sum Rate in Heterogeneous Channel without CSIFR

Fig. 4.6(b) Sum Rate in Heterogeneous Channel without CSIFR @4dB Threshold

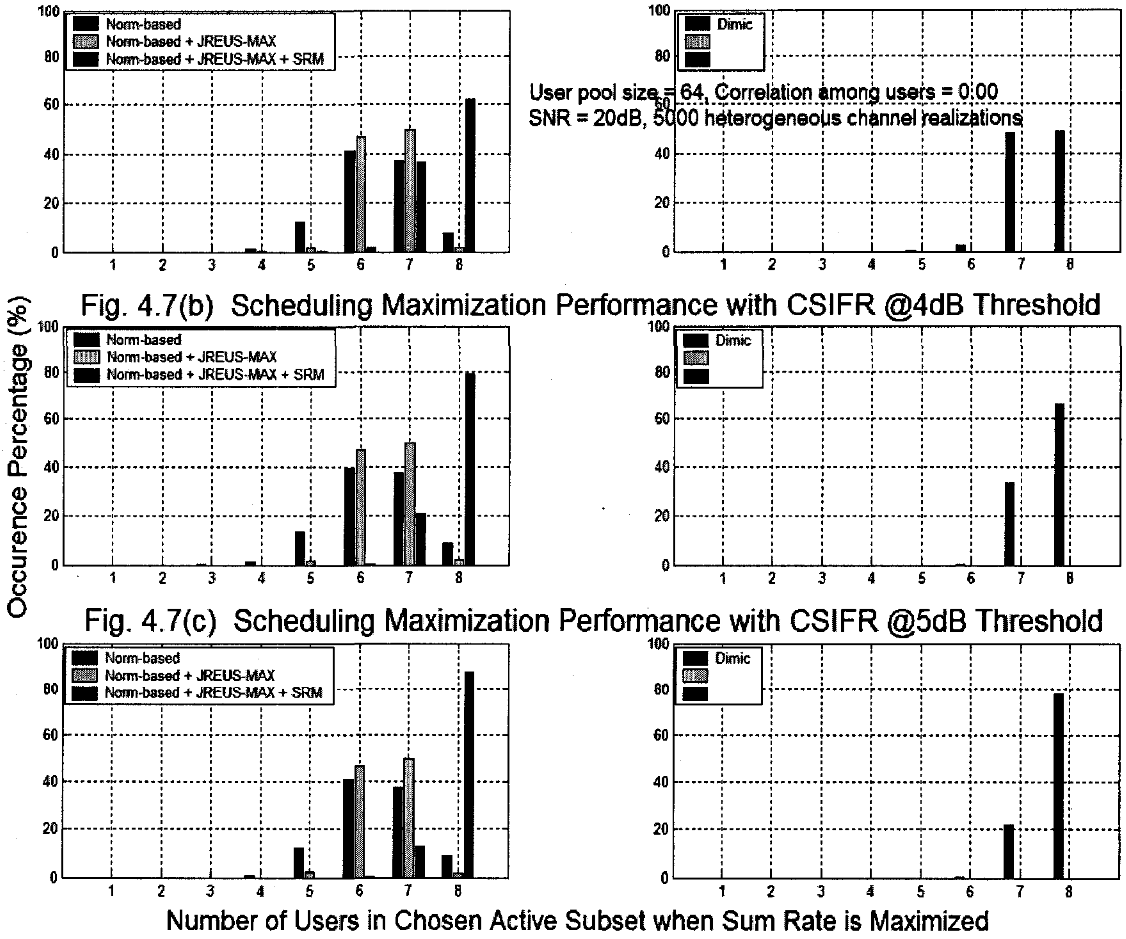


Fig. 4.7(a) Scheduling Maximization Performance with CSIFR @3dB Threshold

Fig. 4.7(b) Scheduling Maximization Performance with CSIFR @4dB Threshold

Fig. 4.7(c) Scheduling Maximization Performance with CSIFR @5dB Threshold

Fig. 4.8. Channel Rate Difference Ratio: DSEL vs JREUS; DSEL vs MDSEL

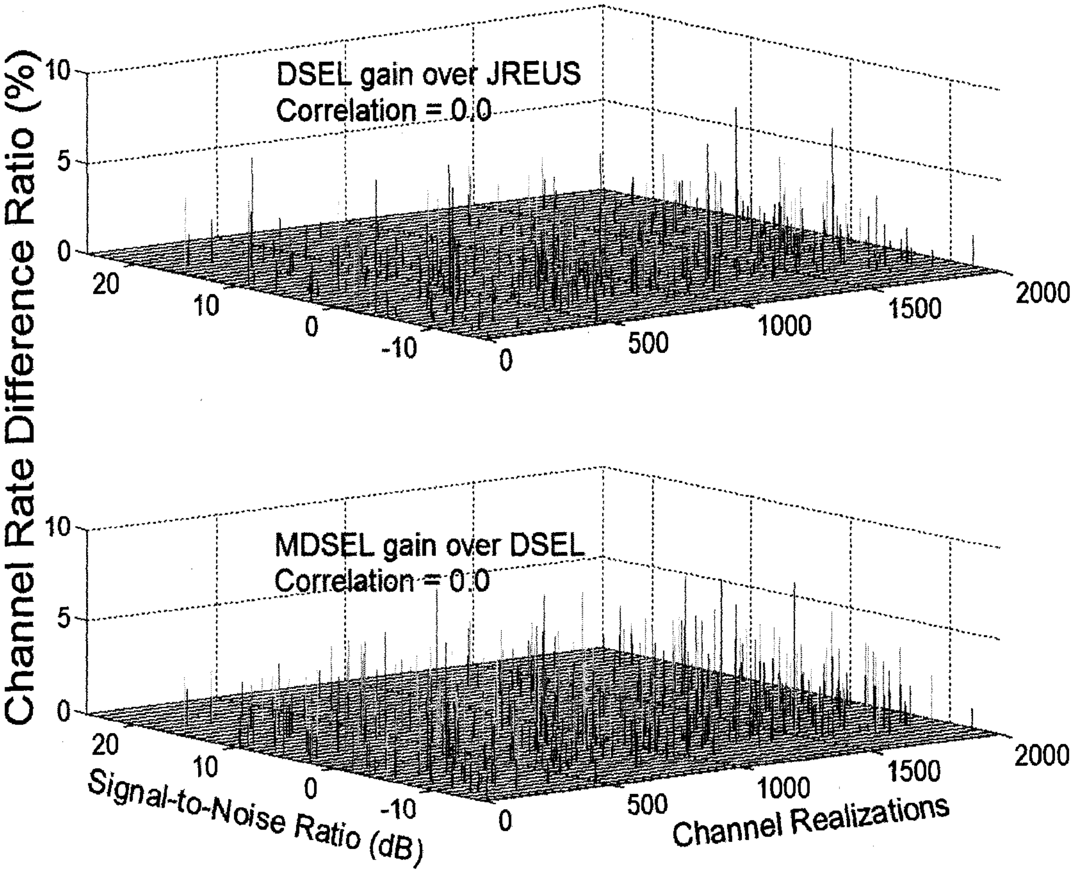
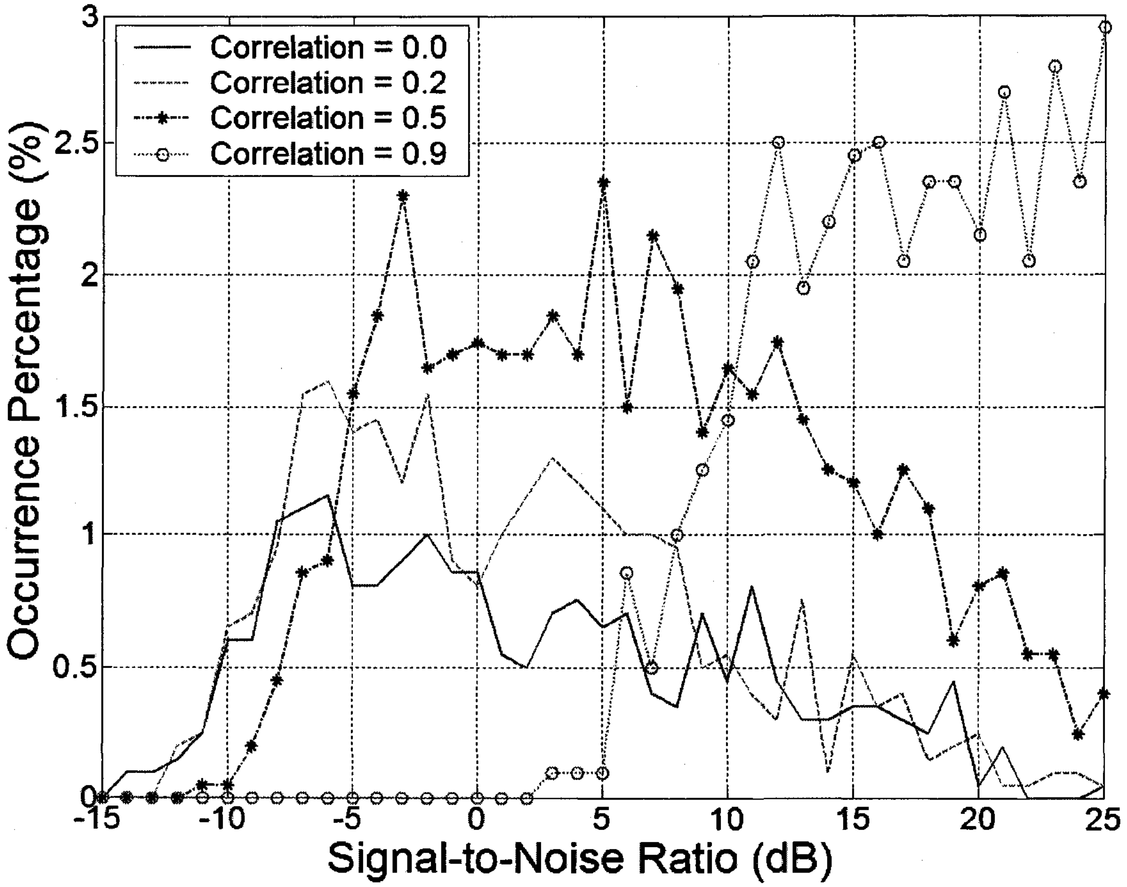


Fig. 4.9. Percentage Time with DSEL Gain Contribution over JREUS



Chapter 5

SELECTION AND ALLOCATION ALGORITHMS FOR BLOCK DIAGONALIZED (BD) SYSTEMS

5.1 Overview

It has been shown in Chapter 3 that receive antenna selection (RAS) or spatial mode selection (SMS) is necessary for sum rate maximization in block diagonalized space-division multiplexing (BD-SDM) systems. It is also mentioned in Chapter 3 that the user selection problem in BD-SDM can be subsumed within the receive antenna selection (RAS) or spatial mode selection (SMS) process. Details on RAS/SMS for BD-SDM are given in this chapter, along with efficient ways of implementing RAS/SMS and user selection jointly.

Differing from RAS for *single*-user MIMO systems, it is necessary to account for the differences between intra-terminal and inter-terminal processing when performing RAS for BD-SDM systems. In general, this necessitates the repeated use of BD pre-coding to check the rate contribution of each antenna. This incurs high computational complexity and ways to avert this are addressed below. Note that the need to account for these differences is also implicitly recognized in existing user selection algorithms like those in [33] and [47] where BD pre-coding is repeatedly done to check the rate contribution of each user. This is also true for TCIBF with single-antenna terminals where for instance, repeated TCIBF pre-coding and rate evaluations are done in [50] during user/antenna selection.

Optimal receive antenna selection (RAS) requires an exhaustive search over all antenna combinations and a BD-SDM rate evaluation is needed for each combination. For a BD-SDM system with M transmit antennas and a pool of K potential users, the number of rate evaluations needed is

$$\sum_{j=1}^{M}\binom{\eta K}{j} \in O(\eta^{M} K^{M}), \tag{5.1.1}$$

where all terminals are assumed to have the same number of antennas (η). The complexity order is $\in O(\eta^{M} K^{M})$ for $M << \eta K$ because $\binom{n}{k} \leq (n^{k}/k!)$. Since an exhaustive search over all possible antenna combinations is needed for sum rate maximization, it is clear that user selection is automatically subsumed under this exhaustive RAS process. Given the need for BD-SDM rate evaluations, only incremental RAS (IRAS) is possible when designing algorithms for complexity reduction. User selection can still be subsumed under the IRAS process that operates on a single-antenna selection (SAS) basis. A decremental (DRAS) approach is not possible since the BD-SDM pre-coding constraint is not met when the user pool is large.

However, implementing incremental-RAS algorithms on a single-antenna selection basis (IRAS-SAS) incurs the following problems: (a) the presence of high intra-terminal correlation often results in the choice of too many users, each with insufficient resources to meet the individual QoS needs, and (b) the need for high computational complexity since selection is done on the SAS basis. Although the complexity order for IRAS-SAS is lower than exhaustive search at $\in O(M\eta K)$ as shown in (5.1.2), the number of rate evaluations is still very high at

$$\sum_{j=0}^{M-1}(\eta K - j) = M\left(2\eta K - M + 1\right)/2. \tag{5.1.2}$$

To help overcome the first shortcoming of single-antenna selection (SAS) where too many users are chosen, the concept of "block antenna selection (BAS)" is introduced. The equivalent "block mode selection (BMS)" is applicable for BD systems that make use of receive-processing matrices for mode allocation. In BAS/BMS, the antennas or modes are chosen on a block basis, that is, antennas/modes are chosen as subsets from each user terminal. In this way, the user selection process is still subsumed under a BAS or BMS process.

Besides providing a means of avoiding the choice of too many users, the BAS/BMS approach also paves the way for algorithms that require fewer rate evaluations. This helps address the second shortcoming of high computational complexity when RAS is implemented on an SAS basis. Additional computational reduction can be realized if BD pre-coding and/or rate evaluations could be avoided during RAS. Seven algorithms are proposed along these lines for the purpose of block antenna/mode selection. They are based on the modification of existing algorithms, for example, a simplified incremental selection method based on mutual null space projections from [58] is proposed to avoid the need for repeated BD pre-coding. Another four are based on existing single-user RAS algorithms and they are computationally more efficient than [33] largely because neither repeated BD-SDM pre-coding nor rate evaluations are required. In this way, decremental BAS/BMS becomes possible since there are no dimensional constraints due to BD pre-coding nor projection requirements. In line with the intent of this thesis, particular attention is paid to the performance enhancement of channel-gain based algorithms because of their low complexity as well as potential for partial CSI feedback. Next, RAS algorithms that derive selection metrics from user- or antenna pairs can also be modified for BAS/BMS. However, attention is not given to them because their levels of performance improvement and complexity reduction are not significant.

BAS/BMS algorithms should be implemented with guidance from the analyses done in Chapter 3. For BD-SDM systems that serve large user pools, results from Chapter 3 show that RAS/SMS should not be done too early during an incremental user/antenna/mode selection process. This can be seen from Figure 3.15 where the sum rates go on a downtrend when the number of antennas/modes is too few. This result lends credence to a decoupled approach where user selection is done before RAS/SMS. In this case, the block size is set at the maximum and users are chosen along with their entire set of antennas. One approach is to choose a group of users with a total of M receive antennas, where M is the number of transmit antennas. RAS is then performed to remove those antennas that are committed to low rate returns. The additional room created may then allow for the scheduling of more users. This process may be iterated and referred to as the "simultaneous scheduling and sum-rate maximization" (SSRM) scheme. The decoupled user-selection and RAS/SMS approach is attractive from the viewpoint of complexity reduction.

It is shown in Section 3.2.2 that *random* antenna/mode de-selection done in a round-robin style among users of a BD-SDM is useful for increasing the ergodic BD-SDM sum rate. For BD systems, a round-robin style of random de-selection produces results that are close to localized judicious de-selection for homogeneous channels. Since the channel matrix size to be fed back to the base station is reduced after a localized RAS (be it random or judicious), a method for reducing the feedback overhead may be developed on this basis. For example, it is shown in Figure 3.15 that around 8 antennas must be removed from the 8-user BD system to achieve the best sum rate when SNR = 20dB. This means that one antenna must be removed from each user and this reduces the size of the channel matrix to be fed back from each user to the base station. This method may be considered if CSI feedback reduction during beamforming is of paramount importance.

Next, it is worthwhile to highlight the impact of small user pools upon the RAS algorithms. To begin, incremental RAS done on the single-antenna selection (SAS) approach may still yield favorable selection when the user pool is small since the selected antennas are less dispersed over different users. In fact, decremental RAS (DRAS-SAS) may also be feasible for small user pools if rate evaluations can be postponed until the pre-coding constraint is met. DRAS-SAS is interesting because it may produce better performance than IRAS-SAS if the total number of receive antennas is close to the number of transmit antennas.

Given the interest for decoupled user-selection and RAS/SMS, attention is paid to the case when the combined pool of receive antennas from all users is $\sum_{j=1}^{K} N_j \leq M$. This situation is akin to a transmit channel-inversion beamforming (TCIBF) system that is serving M single-antenna terminals. In this case, it is possible to employ a single-user RAS algorithm that does not involve repeated TCIBF pre-coding within its user/antenna selection process as in [50] nor projections onto the orthogonal complement subspace. Decremental RAS can be done and better sum rate performance can be obtained as shown in Chapter 4. It is possible to adopt a similar approach to serve BD-SDM systems when $\sum_{j=1}^{K} N_j \leq M$. An efficient, near-optimal algorithm implementing *decremental* RAS on a

SAS basis is shown below. Based on maximum determinant ranking (MDR), it has low complexity and requires a maximum of only M rate evaluations instead of $\sum_{i=1}^{M} i = M(M+1)/2$ rate evaluations needed in typical RAS algorithms, e.g., [34], [55], [58]. Compared to IRAS algorithms, e.g., [58], MDR has better performance and incurs less complexity on average because $(M - N') < N'$ occurs with high probability, where N' is the final number of receive antennas after RAS.

Besides the decoupled user selection and RAS approach, the MDR algorithm can be used in general to improve the sum rate performance of any BAS/BMS algorithm for the usual case where $\sum_{j=1}^{K} N_j > M$, This starts with the use of a BAS/BMS algorithm to choose a group of users and antennas that meets the BD pre-coding constraint. A RAS procedure is then performed using MDR for sum rate maximization. The BAS/BMS-MDR combination for sum rate maximization may free transmission resources that allow the scheduling of additional users, which can be judiciously done for further sum rate maximization. This procedure draws its guidance from the optimal beamforming sum rate that scales as $M \log\log KN$ [18], which means that one should strive for the scheduling of M channels when maximizing the beamforming sum rate. The procedure does so by providing a better means of exploiting the multi-user diversity (MUD). Similar to the single-antenna terminal case, the procedure will raise the feasibility of employing lower complexity BAS/BMS algorithms in practice, for example channel-gain based algorithms.

Since the RAS/SMS process in MDR involves spatial channel ranking, it provides a systematic way for resource allocation to meet individual QoS requirements while minimizing sum rate loss. Essentially, the decremental RAS/SMS process in MDR is useful since it can help identify the next worst antenna/mode to discard or add. The approach is very similar to the case for single-antenna terminals and preliminary resource allocation methods are given.

5.2 A Near-Optimal Decremental RAS/SMS Algorithm when $N = \sum_{j=1}^{K} N_j \leq M$: Maximum Determinant Ranking (MDR)

When the combined pool of receive antennas from all users is $\sum_{j=1}^{K} N_j \leq M$, a simple but near-optimal RAS scheme that is based on the equivalent single-user capacity may be developed. As pointed out earlier, there is a need to discriminate between intra-terminal and inter-terminal correlations when performing RAS/SMS in BD-SDM. In general, this is accomplished by the incorporation of the BD-SDM rate evaluation or its equivalent during the selection process as done in [33] and [58].

To avoid this computationally intensive process, we propose a decremental RAS (DRAS) process that operates on a single-antenna selection (SAS) basis for this case. The theoretical basis for this approach is from the Sato upper bound for multi-user systems [54] where the multi-user sum rate is upper bounded by the capacity of an equivalent single-user, that is, assuming cooperation among all receive antennas. The preference for a decremental approach over an incremental one is the fact that the DRAS process is able to account for the *joint* contributions of all *remaining* antennas and therefore provide better performance than IRAS [34]. The DRAS approach is also more attractive than IRAS in this case because it incurs less complexity on average since $(M - N') < N'$ occurs with high probability, where N' is the final number of receive antennas after RAS.

The algorithm is based on maximum determinant ranking (MDR) and is similar to the DRAS algorithm (Algorithm III) in [34], i.e., it de-selects the next antenna on the basis of minimizing the equivalent single-user capacity loss. Note that the BD-SDM pre-coding constraint is met here and MDR requires a maximum of only M rate evaluations. This is much lower than the methods proposed in [33] and [58], which would incur a maximum of $M(M+1)/2$ rate evaluations for this case. Note that the user selection process is subsumed under this DRAS-SAS process.

Table 5.1 MDR Algorithm

```
Initialize
    $\mathbf{H}_0 = \mathbf{H}$;
    $\mathbf{S}$ = []; %Ranked antenna index set

for $R$ = 0 to $N$-2
    Compute $\chi_j = \mathrm{diag}(\mathbf{H}_R\mathbf{H}_R^H)^{-1}$;
    (%Note: use recursive inverse in Appendix for $R$ > 0)
    $\alpha_R = \underset{j}{\mathrm{argmax}}(\chi_j)$;

    $f(\alpha_R): \alpha_R \rightarrow \mathrm{row}(\mathbf{H}_0)$; %Map $\alpha_R$ to original row index in $\mathbf{H}_0$
    $\mathbf{S} = [\mathbf{S}; f(\alpha_R)]$; %Add new row index into ranking set
    Update $\mathbf{H}_{(R+1),\alpha_R}$; %Remove row $\alpha_R$ from $\mathbf{H}_R$
end

Output: $\mathbf{S}$; %Antennas ranked from worst to best
```

5.2.1 The MDR Algorithm

Let $\mathbf{H} \in \mathbb{C}^{N \times M}$ be the associated composite channel matrix that is derived from the concatenation of all user channel matrices, where $N = \left(\sum_{j=1}^{K} N_j\right) \leq M$. Let $\mathbf{H}_R$ represent the channel matrix *after* the removal of R rows from $\mathbf{H} \in \mathbb{C}^{N \times M}$. The approach here is similar to Algorithm III in [34] but takes on a simplified form that maximizes the determinant of $\mathbf{H}_{(R+1),j}\mathbf{H}_{(R+1),j}^H$, where $\mathbf{H}_{(R+1),j}$ is a one-row reduced composite channel matrix after removing row j from $\mathbf{H}_R$. The row number α_R chosen from $\mathbf{H}_R$ is such that

$$\alpha_R = \underset{j}{\mathrm{argmax}}\left(\det\left(\mathbf{H}_{(R+1),j}\mathbf{H}_{(R+1),j}^H\right)\right). \tag{5.2.1}$$

The basis for this approach is derived from the high SNR approximation for the single-user channel capacity C_{su}, i.e., $C_{su}(\mathbf{H}_R) \approx \log_2(\rho/M)^N \det\left(\mathbf{H}_R\mathbf{H}_R^H\right)$, where ρ is the average SNR. Hence α_R is chosen such that the resulting $\det\left(\mathbf{H}_{(R+1),\alpha_R}\mathbf{H}_{(R+1),\alpha_R}^H\right)$ is the largest among all $\mathbf{H}_{(R+1),j}$ so that the capacity loss $C_{\mathrm{loss}} = C_{su}(\mathbf{H}_R) - C_{su}(\mathbf{H}_{(R+1),\alpha_R})$ is minimized. However, the stopping criterion is not clear and the examination of all possible combinations requires $N(N+1)/2 \in O(N^2)$ steps.

One efficient way of implementing (5.2.1) is via a method similar to JREUS in Chapter 4. Let $\chi_j = \left[(\mathbf{H}_R\mathbf{H}_R^H)^{-1}\right]_{j,j}$, i.e., χ_j are the diagonal elements of $(\mathbf{H}_R\mathbf{H}_R^H)^{-1}$ and they may be re-written as

$$\chi_j = \left\{A_{11}\Delta^{-1}, ..., A_{N'N'}\Delta^{-1}\right\}, \tag{5.2.2}$$

where A_{jj} are cofactors associated with the diagonal elements h_{jj} in $\mathbf{H}_R\mathbf{H}_R^H$, $\Delta = \det(\mathbf{H}_R\mathbf{H}_R^H)$ and $N' = N - R$. Each A_{jj} is found after eliminating row j and column j in $\mathbf{H}_R\mathbf{H}_R^H$, which corresponds to eliminating row j in $\mathbf{H}_R$ to result in $\mathbf{H}_{(R+1),j}$, i.e., $A_{jj} = \det(\mathbf{H}_{(R+1),j}\mathbf{H}_{(R+1),j}^H)$. From (5.2.2) we see that $\max(A_{jj})$ corresponds to $\max(\chi_j)$ and hence the next largest determinant among all $\det\left(\mathbf{H}_{(R+1),j}\mathbf{H}_{(R+1),j}^H\right)$ can be easily found. In this way, the ranking of all antennas from worst to best can be recursively done in $N-1$ steps, i.e., for $R = 0, \cdots, N-2$. For convenience, we will refer to this scheme as "maximum determinant ranking" or MDR and its pseudo-code is given in Table 5.1. BD-SDM rate evaluation is then done from $\mathbf{H}_0$ down to $\mathbf{H}_{N-1}$ by removing antennas one at a time, according to the ranking done by MDR. This incurs N rate evaluations and is much less complex than the $N(N+1)/2 \in O(N^2)$ steps required if all determinant combinations were examined. Further complexity reduction can be achieved by the following stopping criterion: Perform BD-SDM rate evaluation after each antenna de-selection and stop the process whenever the sum rate drops. Simulations have shown only a slight difference in results. This incurs less complexity on average since $(N - N') < N'$ occurs with high probability, where N' is the final number of receive antennas after RAS.

To reduce computational complexity, a recursive inverse method is given in Appendix B so that all subsequent $(\mathbf{H}_R\mathbf{H}_R^H)^{-1}$ where $R = 1, \cdots, N-2$, can be found from the initial $(\mathbf{H}_{R=0}\mathbf{H}_{R=0}^H)^{-1}$. In this way, the computational complexity of MDR is dictated mainly by the evaluation of $(\mathbf{H}_{R=0}\mathbf{H}_{R=0}^H)^{-1}$, which is approximately $\in O(N^3)$. The computational complexity of other RAS and user selection algorithms are given in Table 4.2. MDR has the same complexity order as those algorithms based on pair-wise metrics and the incremental RAS algorithm in [55]. Numerical results show that MDR is near optimal and has performance that exceeds other near-optimal algorithms with higher complexity, e.g., Algorithm III in [34].

5.2.2 Poorest Spatial Mode Elimination (PSME) Algorithm

In Nu-SVD, there is a one-to-one correspondence between the spatial mode gains and the columns of the post-processing matrices defined as $\mathbf{R}_j$ in [29], which are dimensioned according to the desired number of spatial modes for each user j. It is therefore possible to implement a simple SMS algorithm that proceeds by removing the column in $\mathbf{R}_j$ associated with poorest spatial mode to be eliminated. For convenience, we will refer to such a SMS algorithm as "poorest spatial mode elimination (or PSME)". The Nu-SVD process is repeated after each column-elimination and the elimination process is stopped whenever the next iteration results in a lower sum rate. Numerical results show that the PSME and MDR algorithms have identical performance. This is expected since MDR seeks for the next highest determinant while PSME removes the smallest eigenvalue. Hence PSME is the preferred method for SMS in Nu-SVD since it incurs negligible computational load compared to MDR. Note that PSME does not apply to CTR since a one-to-one correspondence between the spatial modes and the columns of $\mathbf{W}_j$ in CTR does not exist.

5.3 Block Antenna/Mode Selection (BAS/BMS) Algorithms

The "block antenna/mode selection (BAS/BMS)" concept is proposed to allow the joint consideration of antenna/mode selection and user selection without the shortcomings of selection on a single-antenna basis, which includes (a) the presence of high intra-terminal correlation often results in the choice of too many users, each with insufficient resources to meet the individual QoS needs, and (b) the need for high computational complexity since selection is done on the SAS basis. The same reasoning applies to spatial mode selection (SMS). In this way, the user selection process can be subsumed under a BAS or BMS process.

We note that optimal BAS/BMS requires an exhaustive search and the number of rate evaluations required is the same as (5.1.1). In this section, we introduce seven algorithms for BAS/BMS that incur less complexity than (5.1.2) for IRAS-SAS. Among them, the more efficient algorithms are based on existing RAS algorithms and are computational more efficient than [33] or [58], largely because neither rate evaluations via repeated BD-SDM pre-coding nor mutual null space projections are required. In this way, decremental BAS/BMS is also possible since dimensional constraints associated with BD-SDM pre-coding or mutual null space projections no longer apply.

5.3.1 Implementing BAS / BMS Using the Approach in [33]

To begin, a straightforward BAS/BMS scheme with less complexity than (5.1.1) is via an adaptation of [33]'s method where BD-SDM pre-coding is systematically used to assess the rate contribution of every antenna subset of every user, including choosing an empty set from a user (i.e., performing user selection). The number of rate evaluations required in the original scheme in [33] when choosing K' users from a pool of K is

$$\begin{aligned}\sum\nolimits_{j=0}^{K'-1}(K-j) &= \sum\nolimits_{j=0}^{K-1}(K-j)-\sum\nolimits_{j=0}^{K-K'-1}(K-K'-j)\\ &= K'\left(2K-K'+1\right)/2.\end{aligned} \tag{5.3.1}$$

Assuming each terminal has the same η antennas, the number of rate evaluations required when incorporating BAS/BMS into [33] is

$$\sum\nolimits_{j=0}^{K'-1}(K-j)2^{\eta} = 2^{\eta}K'(2K-K'+1)/2. \tag{5.3.2}$$

Note that each terminal will have 2^{η} combinations, including the empty set, which means the user is de-selected. The complexity orders of (5.3.1) and (5.3.2) are $\in O(K'K)$ and $\in O(2^{\eta}K'K)$ respectively. Noting that $M=\eta K'$, the complexity order of (5.1.2) is $\in O(\eta^2 K'K)$, which is close to $\in O(2^{\eta}K'K)$ in (5.3.2) for the practical range of $\eta \le 4$. For reference, (5.3.2) has higher complexity when $\eta > 4$. It appears at this stage that using the BAS/BMS approach does not help lower the complexity compared to the IRAS-SAS approach although it helps to address the issue of choosing too many users.

5.3.1.1 Using Pre-selection via Localized BAS / BMS with [33]

Localized BAS is defined here as choosing antenna subsets from a terminal without consideration for other terminals in the BD-SDM context. Localized BAS draws its motivation from the fact that de-selecting antennas with high intra-terminal correlation incurs low percentage rate losses for the affected user while it releases transmit resources for the potential of high rate returns at other terminals. Very importantly, localized BAS also draws its justification from Section 3.2.2 where it was shown that *random* antenna/mode de-selection done in a round-robin style among users of a BD-SDM is useful for increasing the ergodic BD-SDM sum rate. This means that localized BAS has a high probability of providing good ergodic performance when done judiciously.

The same reasoning applies to BD schemes that work with projected channels, e.g., the CTR method in [27] and the Nu-SVD method in [29]. There, the starting point would be to use all columns of $\mathbf{U}_j$ from $\mathbf{H}_j = \mathbf{U}_j \mathbf{\Sigma}_j \mathbf{V}_j^H$ to activate all spatial modes.

There are many ways of identifying the antennas/modes with high correlation. For example, RAS algorithms and user selection algorithms for multi-user systems with single-antenna terminals may be used. An efficient way is similar to the JREUS algorithm in Section 4.1. The explanation for this can be derived from (2.36) where b_i measures the projection magnitude of the i row vector into the null space of all other row vectors. In this way, we may treat the antennas at the same terminal as single-antenna users and the magnitudes of their mutually orthogonal projections may be measured in terms of the maximum TCIBF sum rate obtained via JREUS. This process is repeated across all user terminals to remove antennas/modes with high intra-terminal correlation.

Taking guidance from Section 3.2.2, the following discussion is based on an example where each user is equipped with 4 antennas. Rate evaluations will be done round-robin style where each user is considered in turn for two cases, namely, (a) without antenna de-selection and (b) with one antenna de-selected on a local intra-terminal basis. The best users that yield the highest sum rate while meeting the BD pre-coding constraint are then selected from this pool. Further RAS may then be done on the chosen group using the MDR algorithm and further scheduling may be done if the pre-coding constraint is not exceeded. This process is iterated until the sum rate drops or when the pre-coding constraint is exceeded.

5.3.1.2 Using BAS / BMS with Localized Antenna/Mode Ranking with [33]

In this approach, all antennas/modes at each terminal are ranked locally. The motivation behind this scheme is as stated in the previous section. This may be done using any RAS or user selection algorithm and rate evaluation is not needed. In particular the JREUS approach may be used for high performance and low complexity. In this way, the number of rate evaluations may be reduced to

$$\sum_{j=0}^{K'-1}(K-j)(\eta+1)=(\eta+1)K'(2K-K'+1)/2, \tag{5.3.3}$$

where $(\eta+1)$ is the number of choices at each terminal. The complexity order is $\in O\big((\eta+1)K'K\big)$, which means the complexity of (5.3.3) is lower than that of IRAS-SAS in (5.1.2) or the original BAS without localized antenna de-selection in (5.3.2) when $\eta \geq 2$.

Taking a similar approach as that in Section 5.3.1.1, further reduction in the number of rate evaluations may be obtained when implementing localized de-selection using the ranked list on a round-robin basis for all users.

5.3.2 Implementing BAS / BMS Without Rate Evaluation

The motivation behind this is to avoid repeated BD-SDM rate evaluations during selection, which is a computationally heavy process that involves SVD. One possible approach is by adapting single-user MIMO RAS algorithms e.g., [34], [55], [56], [75] or user selection algorithms for single-antenna terminals e.g., [26], [44] – [49]. These algorithms employ methods like capacity maximization, orthogonal complement projection, channel-gain metrics and pair-wise metrics based on correlation, cosine and squared normalized inner products, and combinations thereof. For a pool of K single-antenna terminals or antennas in general, the simplest channel-gain based selection algorithm requires only K decision metrics while most other schemes require $K'(2K-K'+1)/2 \in O(K'K)$ decision metrics. In general, algorithms using pair-wise metrics do not perform as well as those based on capacity maximization or orthogonal complement projection.

These algorithms may be adapted for BAS/BMS as follows: (a) For capacity maximization methods, blocks of row vectors may be chosen or de-selected for incremental and decremental selection respectively. (b) For orthogonal complement projection methods, projection may be done in blocks instead of single antennas or modes. (c) For channel gain metrics, a "block gain" metric may be defined. (d) For pair-wise metrics, "block-wise" metrics may be obtained from the basic pair-wise metrics. We will highlight the schemes in (a) and (b) given their good performance and (c), given its simplicity. The schemes in (d) do not provide high performance or significant complexity reduction. We want to highlight that all material written for BAS are applicable to BMS as well in the following sections. It is also important to stress that guidance from the analyses done in Chapter 3 should be followed when implementing BAS/BMS algorithms for block diagonalized systems. To give visibility on the intrinsic complexity of each algorithm however, the complexity evaluation for each algorithm in the following sections is done without any consideration for Chapter 3.

5.3.2.1 BAS/BMS using Norm-Based Selection (NBS)

The simplest antenna selection algorithm is based on the power of the received signals [75]. This algorithm selects antennas with the largest channel gains (norms) and performs well only at low signal-to-noise ratios. It can be easily adapted for BAS by ranking α_j, which is the sum of all channel gains associated with each multi-antenna terminal, where

$$\{\alpha_j;\ j=1,\ldots,K\} = \sum\nolimits_{i=1}^{N_j} \left\| \mathbf{h}_{i;j} \right\|^2, \tag{5.3.4}$$

where $\mathbf{h}_{i;j}$ is the i^{th} channel row vector of user *j* and N_j is the number of receive antennas at user *j*. As shown in (5.3.4), there are *K* decision metrics. Norm-based BAS (NBS-BAS) is accomplished by choosing the *K*' users that are associated with the *K*' highest α_j values. Assuming η antennas at each terminal, the number of complex multiplications is $MK\eta$.

Taking a similar approach as that in Section 5.3.1.1, localized de-selection may be done based on the metric α_j for each user *j*. Again, further RAS/SMS can be done using the MDR algorithm and further scheduling may be done if the pre-coding constraint is not exceeded. Numerical results show good sum rate improvement when this iterative scheme is used.

5.3.2.2 BAS/BMS based on Orthogonal Complement Projections (OCP)

This can be based on Algorithm I in [34], which is a RAS algorithm for single-user MIMO systems, or on the algorithm in [44], which is a user selection algorithm for single-antenna terminals. They can be adapted to operate on blocks of antennas for BAS/BMS. Since dimensional constraints have to be met, the orthogonal complement projection approach is only possible on an incremental selection basis.

Consider a multi-user MIMO system with *M* transmit antennas and *K* users, each with N_k receive antennas. The goal is to choose K' users out of the original pool of *K* users using incremental BAS/BMS. Let $\mathbf{H}_j$ be the channel matrix for user *j* and let S_c be the set of n users that are already chosen where $r_1,....,r_n$ are indices of the chosen users. Let $\mathbf{H}_{r_1,\ldots,r_n}$ be the composite channel matrix for S_c obtained by the concatenation of the selected users' channel matrices $\mathbf{H}_{r_1},\cdots,\mathbf{H}_{r_n}$.

The orthogonal complement projection (OCP) method simply selects the next user k that has the largest projection norms in the null space of $\mathbf{H}_{r_1,\ldots,r_n}$, denoted as $\mathbf{H}_{r_1,\ldots,r_n}^{\perp}$. Let $\mathbf{H}_{pk}=\mathbf{H}_k\mathbf{P}_{r_1,\ldots,r_n}^{\perp}$ be the projected matrix of $\mathbf{H}_k$ onto $\mathbf{H}_{r_1,\ldots,r_n}^{\perp}$ where $\mathbf{P}_{r_1,\ldots,r_n}^{\perp}=\mathbf{I}_M-\mathbf{H}_{r_1,\ldots,r_n}^{H}\left(\mathbf{H}_{r_1,\ldots,r_n}\mathbf{H}_{r_1,\ldots,r_n}^{H}\right)^{-1}\mathbf{H}_{r_1,\ldots,r_n}$. The choice among $\mathbf{H}_{pk}$ may be determined by

$$u_s = \arg\max_k \left(\left\| \mathbf{H}_{pk} \right\|_F \right), \tag{5.3.5}$$

where u_s is the index of the chosen user and $\left\| \mathbf{H}_{pk} \right\|_F = \mathrm{tr}\left(\mathbf{H}_{pk} \mathbf{H}_{pk}^H \right)$. Alternatively, u_s may be based on the following

$$u_s = \arg\max_k \left(\det\left(\mathbf{H}_{pk} \mathbf{H}_{pk}^H \right) \right), \tag{5.3.6}$$

or

$$u_s = \arg\max_k \left(\log_2 \det\left(1 + \rho \mathbf{H}_{pk} \mathbf{H}_{pk}^H \right) \right). \tag{5.3.7}$$

Next, we address the number of decision metrics needed. When each terminal is equipped with η antennas and OCP is used for *user* selection and not antenna subset selection, the number of decision metrics needed for choosing K' users out of K users is

$$\sum_{j=0}^{K'-1} (K - j) = K'\left(2K - K'+1\right)/2, \tag{5.3.8}$$

which is the same as (5.3.1). When antenna subset selection is desired, each terminal presents 2^η choices and hence the number of decision is the same as (5.3.2). The localized antenna/mode ranking method in Section 5.3.1.2 can be done at each terminal to reduce the choices to $(\eta + 1)$ at each terminal. This reduces the number of decision metrics and is equal to (5.3.3). In all cases here, the computational cost is lower than those methods involving [33] because no BD-SDM pre-coding and rate evaluations are involved in each decision metric. A detailed assessment on the computational complexity is not done yet at this stage. Taking a similar approach as that in Section 5.3.1.1, a round-robin style of de-selection may be done so that an initial group of users with the highest sum rate can be chosen. Further RAS/SMS can then be done to further improve the sum rate using the MDR algorithm and additional scheduling may be done if the pre-coding constraint is not exceeded.

5.3.2.3 BAS/BMS based on Equivalent Single-User Capacity Maximization

The algorithms proposed here are based on the RAS algorithms in [34]. The incremental algorithm is based on maximizing the capacity gain of a single-user MIMO system for the next chosen antenna. The decremental algorithm is based on minimizing the capacity loss of a single-user MIMO system for the next de-selected antenna. In this multi-user context, the equivalent single user is formed assuming cooperation among all user antennas. The equivalent single-user channel matrix is formed by the concatenation of all user channel matrices. We begin by giving the background to the incremental RAS scheme known as Algorithm II in [34].

5.3.2.3.1 Background on Incremental RAS (IRAS) in [34]

Consider a point-to-point MIMO system with M transmit antennas and L receive antennas where $L > M$. Let l represent the receive antenna index. The objective here is to choose M receive antennas using IRAS. Let S_c be a set of n antennas that has already been chosen and let $r_1,....,r_n$ be the indices of the chosen antennas. Let $\mathbf{H}_{r_1},\cdots,\mathbf{H}_{r_n}$ represent their corresponding channel row vectors and $\mathbf{H}_{r_1,\ldots,r_n} \in \mathbb{C}^{nxM}$ be its associated channel matrix that is a stack of the row vectors $\mathbf{H}_{r_1},\cdots,\mathbf{H}_{r_n}$. Suppose that the next antenna l is to be added to S_c and let $\mathbf{H}_l$ be its associated channel row vector. The resulting single-user MIMO channel capacity when $\mathbf{H}_l$ is appended to $\mathbf{H}_{r_1,\ldots,r_n}$ is

$$\begin{aligned} C_{\text{IRAS}}(\mathbf{H}_{r_1,\ldots,r_n};\mathbf{H}_l) &= \log_2\left\{\det\left(\mathbf{I}_M + \rho(\mathbf{H}^H_{r_1,\ldots,r_n}\mathbf{H}_{r_1,\ldots,r_n} + \mathbf{H}_l^H\mathbf{H}_l)\right)\right\} \\ &= \log_2\left\{\det\left(\underbrace{\mathbf{I}_M + \rho\mathbf{H}^H_{r_1,\ldots,r_n}\mathbf{H}_{r_1,\ldots,r_n}}_{\mathbf{A}} + \underbrace{\rho\mathbf{H}_l^H}_{\mathbf{x}}\underbrace{\mathbf{H}_l}_{\mathbf{y}^H}\right)\right\}, \end{aligned} \qquad (5.3.9)$$

where ρ is the average SNR. Note that when $n < M$, the normal expression for capacity in (5.3.9) would have used $\mathbf{H}_{r_1,\ldots,r_n}\mathbf{H}^H_{r_1,\ldots,r_n}$ instead of $\mathbf{H}^H_{r_1,\ldots,r_n}\mathbf{H}_{r_1,\ldots,r_n}$ as above. Equation (5.3.9) is numerically correct since $\det(\mathbf{I}_m + \underbrace{\mathbf{C}}_{(m\text{x}n)}\underbrace{\mathbf{B}}_{(n\text{x}m)}) = \det(\mathbf{I}_n + \underbrace{\mathbf{B}}_{(n\text{x}m)}\underbrace{\mathbf{C}}_{(m\text{x}n)})$ and this allows for appending of $\mathbf{H}_l$ to be accounted for in an additive way. Given the identity $\det(\mathbf{A})*\det(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})=\det(\mathbf{D})*\det(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})$, one may write

$$\det(\mathbf{A}+\mathbf{x}\mathbf{y}^H)=\det(\mathbf{A})*(1+\mathbf{y}^H\mathbf{A}^{-1}\mathbf{x}), \tag{5.3.10}$$

where $\mathbf{y}^H\mathbf{A}^{-1}\mathbf{x}$ is a scalar. Using (5.3.10), (5.3.9) may be expressed as

$$\begin{aligned} C_{\mathrm{IRAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_l) &= \log_2\left\{\det\left(\mathbf{I}_M+\rho\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)\left(1+\mathbf{H}_l\left(\mathbf{I}_M+\rho\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)^{-1}\rho\mathbf{H}_l^H\right)\right\} \\ &= \underbrace{\log_2\det\left(\mathbf{I}_M+\rho\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)}_{\text{Capacity of original channel matrix}} + \underbrace{\log_2\left(1+\underbrace{\mathbf{H}_l\left(\rho^{-1}I_M+\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)^{-1}\mathbf{H}_l^H}_{\beta_l=\text{Scalar term to be maximized}}\right)}_{\Delta C_{\mathrm{IRAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_l)=\text{Additional capacity due to additional row vector }\mathbf{H}_l}. \end{aligned} \tag{5.3.11}$$

Note the definition of $\Delta C_{\mathrm{IRAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_l)$ in (5.3.11), which is the additional capacity gain due to the additional row vector $\mathbf{H}_l$. To maximize $\Delta C_{\mathrm{IRAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_l)$, we wish to identify a user l so that

$$(\beta_l)_{\max}=\max_{l\notin S_c}\mathbf{H}_l\left(\rho^{-1}\mathbf{I}_M+\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)^{-1}\mathbf{H}_l^H. \tag{5.3.12}$$

The matrix inverse within (5.3.12) is the most computationally heavy portion that may be made less using a recursive inverse procedure that is based on the Woodbury matrix identity

$$\left(\mathbf{A}^{-1}+\mathbf{U}\mathbf{C}\mathbf{V}^H\right)^{-1}=\mathbf{A}-\mathbf{A}\mathbf{U}\left(\mathbf{C}+\mathbf{V}^H\mathbf{A}\mathbf{U}\right)^{-1}\mathbf{V}^H\mathbf{A}. \tag{5.3.13}$$

By letting $\mathbf{C}=\mathbf{I}$, $\mathbf{U}=\mathbf{H}^H_{r_n}$ and $\mathbf{V}^H=\mathbf{H}_{r_n}$, we get the Sherman-Morrison identity. Defining $\mathbf{A}^{-1}_{n+1}\triangleq\left(\rho^{-1}\mathbf{I}_M+\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)^{-1}$, the following recursive inverse procedure is obtained

$$\mathbf{A}_{n+1}^{-1} \triangleq \left(\rho^{-1}\mathbf{I}_M + \left[\mathbf{H}_{r_1,\dots,r_{n-1}};\mathbf{H}_{r_n}\right]^H \left[\mathbf{H}_{r_1,\dots,r_{n-1}};\mathbf{H}_{r_n}\right]\right)^{-1} = \left(\underbrace{\rho^{-1}\mathbf{I}_M + \mathbf{H}^H_{r_1,\dots,r_{n-1}}\mathbf{H}_{r_1,\dots,r_{n-1}}}_{\mathbf{A}_n} + \underbrace{\mathbf{H}^H_{r_n}}_{\mathbf{U}}\underbrace{\mathbf{H}_{r_n}}_{\mathbf{V}^H}\right)^{-1}$$

$$= \mathbf{A}_n^{-1} - \frac{\mathbf{A}_n^{-1}\mathbf{H}^H_{r_n}\mathbf{H}_{r_n}\mathbf{A}_n^{-1}}{\left(1+\mathbf{H}_{r_n}\mathbf{A}_n^{-1}\mathbf{H}^H_{r_n}\right)},$$

(5.3.14)

where $\mathbf{H}_{r_n}$ arises from the previously chosen row vector. IRAS starts by identifying the first antenna, which is the row vector with the maximum norm length, i.e.,

$$\mathbf{H}_1 = \max_l \left\|\mathbf{H}_l\right\|^2, \quad l = 1,\dots.,L, \qquad (5.3.15)$$

where $\mathbf{H}_l$ are the channel row vectors of all receive antennas. The initial value for $\mathbf{A}_n^{-1}$ is $\mathbf{A}_1^{-1} = \rho^{-1}\mathbf{I}_M$ and n is subsequently varied in the range $n \in \{2,\dots..,M\}$. For convenience, we will refer to this scheme as GIS-SAS, i.e., Gorokhov's incremental selection based on single-antenna selection (SAS).

5.3.2.3.2 Background on Decremental RAS (DRAS) in [34]

For DRAS, the algorithm begins with $n = K$ antennas in the set S_c, where K is the total number of antennas. For DRAS, S_c may be defined as the set of remaining antennas. De-selection is done by removing antennas from S_c such that the single-user MIMO capacity loss is minimized and the resulting capacity expression is similar to (5.3.11) where

$$C_{\text{DRAS}}(\mathbf{H}_{r_1,\dots,r_n} \setminus \mathbf{H}_l) = \underbrace{\log_2 \det\left(\mathbf{I}_M + \rho\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)}_{\text{Capacity of original channel matrix}} + \underbrace{\log_2\left(1 - \underbrace{\mathbf{H}_l\left(\rho^{-1}\mathbf{I}_M + \mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)^{-1}\mathbf{H}_l^H}_{\delta_l = \text{Scalar term to be minimized}}\right)}_{\Delta C_{\text{DRAS}}(\mathbf{H}_{r_1,\dots,r_n}\setminus\mathbf{H}_l) = \text{Capacity loss after removing row vector } \mathbf{H}_l},$$

(5.3.16)

where $\mathbf{H}_{r_1,\dots,r_n} \backslash \mathbf{H}_l$ represents the removal of row vector $\mathbf{H}_l$ from $\mathbf{H}_{r_1,\dots,r_n}$. To minimize $\Delta C_{\text{DRAS}}(\mathbf{H}_{r_1,\dots,r_n} \backslash \mathbf{H}_l)$, we wish to identify a user l so that

$$(\delta_l)_{\min} = \min_{l \in S_c} \mathbf{H}_l \left(\rho^{-1}\mathbf{I}_M + \mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}\right)^{-1} \mathbf{H}_l^H \tag{5.3.17}$$

A recursive inverse procedure similar to that in (5.3.14) can also be applied to (5.3.17) to lower computational cost

$$\begin{aligned}\mathbf{A}_{n-1}^{-1} \triangleq \left(\rho^{-1}\mathbf{I}_M + \left[\mathbf{H}_{r_1,\dots,r_n} \backslash \mathbf{H}_l\right]^H \left[\mathbf{H}_{r_1,\dots,r_n} \backslash \mathbf{H}_l\right]\right)^{-1} &= \left(\underbrace{\rho^{-1}\mathbf{I}_M + \mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}}_{\mathbf{A}_n} - \underbrace{\mathbf{H}_l^H}_{\mathbf{U}}\underbrace{\mathbf{H}_l}_{\mathbf{V}^H}\right)^{-1} \\ &= \mathbf{A}_n^{-1} + \frac{\mathbf{A}_n^{-1}\mathbf{H}_l^H\mathbf{H}_l\mathbf{A}_n^{-1}}{\left(1-\mathbf{H}_l\mathbf{A}_n^{-1}\mathbf{H}_l^H\right)},\end{aligned} \tag{5.3.18}$$

where $\mathbf{H}_l$ arises from the previously chosen antenna. For the DRAS case, the process starts with finding the inverse of $\mathbf{A}_{n=K}^{-1}$. This makes GDS-SAS more complex than GIS-SAS, however its performance is better. For convenience, we will refer to this scheme as GDS-SAS, i.e., Gorokhov's decremental selection based on single-antenna selection (SAS).

5.3.2.3.3 Incremental BAS/BMS

This is done by adapting the IRAS algorithm in [34] given in Section 5.3.2.3.1. For convenience, we will refer to this scheme as GIS-BAS, i.e., Gorokhov's incremental selection based on block-antenna selection (BAS). This is referred interchangeably to as GIS-BMS for block spatial mode selection. Consider a multi-user MIMO system with M transmit antennas and K users, each with N_k receive antennas. The goal is to choose K' users out of the original pool of K users using incremental BAS/BMS. Let $\mathbf{H}_j$ be the channel matrix for user j and let S_c be the set of n users that are already chosen where $r_1,\dots\dots,r_n$ are indices of the chosen users. Let $\mathbf{H}_{r_1,\dots,r_n}$ be the composite channel matrix for

S_c obtained by the concatenation of the selected users' channel matrices $\mathbf{H}_{r_1},\cdots,\mathbf{H}_{r_n}$. When a new user k is added to S_c, the new composite channel is denoted as $\left[\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_k\right]$. Assuming cooperation among all receive antennas in S_c, the increase in the equivalent single-user MIMO capacity is

$$C_{\text{IBAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_k)=\log_2\det(\mathbf{I}_M+\rho\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}+\rho\mathbf{H}^H_k\mathbf{H}_k). \tag{5.3.19}$$

This is similar to (5.3.9) and hence

$$C_{\text{IBAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_k)=\underbrace{\log_2\det(\rho\mathbf{G}_n)}_{\text{Capacity of original channel matrix}}+\underbrace{\log_2\det(\mathbf{I}_{N_k}+\mathbf{H}_k\mathbf{G}_n^{-1}\mathbf{H}_k^H)}_{\Delta C_{\text{IBAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_k)\text{=additional capacity to be maximized}}, \tag{5.3.20}$$

where $\mathbf{G}_n\triangleq\rho^{-1}\mathbf{I}_M+\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}$ and N_k is the number of antennas/modes in user k. To maximize $C_{\text{IBAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_k)$, we perform the following

$$\Delta C_{\text{IBAS}}(\mathbf{H}_{r_1,\dots,r_n};\mathbf{H}_k)_{\max}=\max_{k\notin S_c}\log_2\det(\mathbf{I}_{N_k}+\mathbf{H}_k\mathbf{G}_n^{-1}\mathbf{H}_k^H). \tag{5.3.21}$$

To reduce computational load, the current inverse $\mathbf{G}_n^{-1}$ in (5.3.21) may be recursively done using the Woodbury identity in (5.3.13) with $\mathbf{C}=\mathbf{I}$, $\mathbf{U}=\mathbf{H}^H_{r_{n-1}}$ and $\mathbf{V}^H=\mathbf{H}_{r_{n-1}}$, where $\mathbf{H}_{r_{n-1}}$ is the previously chosen user channel matrix

$$\mathbf{G}_{n+1}^{-1}=\mathbf{G}_n^{-1}-\mathbf{G}_n^{-1}\mathbf{H}^H_{r_n}\left(\mathbf{I}_{N_{r_n}}+\mathbf{H}_{r_n}\mathbf{G}_n^{-1}\mathbf{H}^H_{r_n}\right)^{-1}\mathbf{H}_{r_n}\mathbf{G}_n^{-1}, \tag{5.3.22}$$

where $\mathbf{H}_{r_n}$ is the previously chosen user channel matrix, $n\in\{1,\dots.,K'-1\}$, K' is the number of users to be chosen and $\mathbf{G}_1^{-1}=\rho^{-1}\mathbf{I}_{N_{r_1}}$. Selection of the first user r_1 may be found by

$$r_1=\arg\max_{1\le j\le K}\left(\det\left(\mathbf{I}_{N_j}+\rho\mathbf{H}_j\mathbf{H}_j^H\right)\right). \tag{5.3.23}$$

An approximation to (5.3.23) may be done for complexity reduction using the identity

$$\ln\det(\mathbf{A})\le\text{trace}(\mathbf{A})-m, \tag{5.3.24}$$

Table 5.2: Pseudo-Code for GIS-BAS Algorithm

```
Find $r_1 = \arg\max_{1\le k\le K}\left(\det\left(\mathbf{I}_{N_k}+\rho\mathbf{H}_k\mathbf{H}_k^H\right)\right)$ and set $\mathbf{G}_1^{-1}=\rho^{-1}\mathbf{I}_{N_{r_1}}$ ;
Set $S_c=\{\,\}$ and update $S_c$ with $r_1$ ;

For $n$ = 1 to $K'$-1

    Update $\mathbf{G}_{n+1}^{-1}=\mathbf{G}_n^{-1}-\mathbf{G}_n^{-1}\mathbf{H}_{r_n}^H\left(\mathbf{I}_{N_{r_n}}+\mathbf{H}_{r_n}\mathbf{G}_n^{-1}\mathbf{H}_{r_n}^H\right)^{-1}\mathbf{H}_{r_n}\mathbf{G}_n^{-1}$;
    Find $r_{n+1}=\arg\max_{k\notin S_c}\det(\mathbf{I}_{N_k}+\mathbf{H}_k\mathbf{G}_{n+1}^{-1}\mathbf{H}_k^H)$ ;
    Update $S_c$ with $r_{n+1}$ ;

End
```

where **A** is a $m \times m$ matrix. Numerical results show that the approximation in (5.3.24) performs poorly when high intra-terminal antenna correlation exists. The performance for GIS-BAS approach that of NBS-BAS under such conditions. The pseudo-code for GIS-BAS is given in Table 5.2. Note that the list of users in S_c obtained via GIS-BAS is ranked from best to worst.

Next, we address the number of decision metrics needed. When each terminal is equipped with η antennas and GIS-BAS is used for *user* selection and not antenna subset selection, the number of decision metrics needed for choosing K' users out of K users is

$$\sum_{j=0}^{K'-1}(K-j)=K'\left(2K-K'+1\right)/2, \qquad (5.3.25)$$

which is the same as (5.3.1). When antenna subset selection is desired, each terminal presents 2^{η} choices and hence the number of decision is the same as (5.3.2). The localized antenna/mode ranking method in Section 5.3.1.2 can be done at each terminal to reduce the choices to $(\eta+1)$ at each terminal. This reduces the number of decision metrics and is equal to (5.3.3). In all cases here, the computational cost is lower than methods involving [33] because no BD-SDM pre-coding and rate evaluations are involved in each decision metric. A detailed assessment on the computational complexity is not done yet at this stage. Taking a similar approach as that in Section 5.3.1.1, a round-robin style of de-selection may be done so that an initial group of users with the highest sum rate can be chosen. Further RAS/SMS can then be done to further improve the sum rate using the

MDR algorithm and additional scheduling may be done if the pre-coding constraint is not exceeded.

5.3.2.3.4 Decremental BAS/BMS

This is done by adapting the DRAS algorithm in [34] given in Section 5.3.2.3.2. For convenience, we will refer to this scheme as GDS-BAS, i.e., Gorokhov's decremental selection based on block-antenna selection (BAS). This is interchangeably referred to as GDS-BMS for block spatial mode selection. As for GDS-SAS, the development of GDS-BAS is very similar to GIS-BAS. For GDS-BAS, we start with the entire pool of K users in S_c and the equivalent single-user capacity is $C_{\text{DBAS}}(\mathbf{H}_{r_1,\dots,r_K})$. When a user k is removed, the capacity becomes

$$C_{\text{DBAS}}(\mathbf{H}_{r_1,\dots,r_n}\setminus\mathbf{H}_k)=\log_2\det(\mathbf{I}_M+\rho\mathbf{H}^H_{r_1,\dots,r_n}\mathbf{H}_{r_1,\dots,r_n}-\rho\mathbf{H}^H_k\mathbf{H}_k). \tag{5.3.26}$$

The expressions corresponding to (5.3.20), (5.3.21) and (5.3.22) are

$$C_{\text{DBAS}}(\mathbf{H}_{r_1,\dots,r_n}\setminus\mathbf{H}_k)=\underbrace{\log_2\det(\rho\mathbf{G}_n)}_{\text{Capacity of original channel matrix}}+\underbrace{\log_2\det(\mathbf{I}_{N_k}-\mathbf{H}_k\mathbf{G}_n^{-1}\mathbf{H}_k^H)}_{\Delta C_{\text{DBAS}}(\mathbf{H}_{r_1,\dots,r_n}\setminus\mathbf{H}_k)=\text{capacity loss to be minimized}}, \tag{5.3.27}$$

$$\Delta C_{\text{DBAS}}(\mathbf{H}_{r_1,\dots,r_n}\setminus\mathbf{H}_k)_{\min}=\min_{k\in S_c}\log_2\det(\mathbf{I}_{N_k}-\mathbf{H}_k\mathbf{G}_n^{-1}\mathbf{H}_k^H) \tag{5.3.28}$$

and

$$\mathbf{G}_{n-1}^{-1}=\mathbf{G}_n^{-1}+\left(\mathbf{G}_n^{-1}\mathbf{H}_{r_n}^H\left(\mathbf{I}_{N_{r_n}}-\mathbf{H}_{r_n}\mathbf{G}_n^{-1}\mathbf{H}_{r_n}^H\right)^{-1}\right)\mathbf{H}_{r_n}\mathbf{G}_n^{-1}, \tag{5.3.29}$$

where $\mathbf{H}_{r_n}$ is the previously chosen user channel matrix. The pseudo-code for GDS-BAS is given in Table 5.3. Note that the list of users remaining in S_c after GDS-BAS is not ranked.

Next, we address the number of decision metrics needed. When choosing K' users out of K users, the number of users to be de-selected is $K-K'$. When each terminal is equipped with η antennas and GIS-BAS is used for *user* selection and not antenna subset selection, the number of decision metrics needed for choosing K' users out of K users is

Table 5.3: Pseudo-Code for GDS-BAS Algorithm

```
Set $S_c = \{1, ...., K\}$, $\mathbf{G}_K^{-1} = (\rho^{-1}\mathbf{I}_M + \mathbf{H}^H_{r_1,..r_K}\mathbf{H}_{r_1,..r_K})^{-1}$ and find
$r_K = \arg\min_{1\le k\le K,\, k\in S_c} \det\left(\mathbf{I}_{N_k} - \mathbf{H}_k\mathbf{G}_K^{-1}\mathbf{H}_k^H\right)$;
Update $S_c \backslash r_K$, the set of remaining users;
For $n$ = $K$ to $K$-$K'$+1
    Update $\mathbf{G}_{n-1}^{-1} = \mathbf{G}_n^{-1} + \left(\mathbf{G}_n^{-1}\mathbf{H}_{r_n}^H\left(\mathbf{I}_{N_{r_n}} - \mathbf{H}_{r_n}\mathbf{G}_n^{-1}\mathbf{H}_{r_n}^H\right)^{-1}\right)\mathbf{H}_{r_n}\mathbf{G}_n^{-1}$;
    Find $r_{n-1} = \arg\min_{k\in S_c} \det(\mathbf{I}_{N_k} - \mathbf{H}_k\mathbf{G}_{n-1}^{-1}\mathbf{H}_k^H)$ ;
    Update $S_c \backslash r_{n-1}$ ;
End
```

$$\begin{aligned}\sum\nolimits_{j=0}^{K-K'-1}(K-j) &= \sum\nolimits_{j=0}^{K-1}(K-j) - \sum\nolimits_{j=0}^{K'-1}(K'-j)\\ &= (K^2 + K - K'^2 - K')/2.\end{aligned} \tag{5.3.30}$$

The complexity order is $\in O(K^2)$, which is higher than GIS-BAS unless $K' \to K$. When antenna subset selection is desired, each terminal presents 2^η choices and hence the number of decision is

$$\sum\nolimits_{j=0}^{K'+1}(K-j)2^\eta = 2^\eta(K^2 + K - K'^2 - K')/2, \tag{5.3.31}$$

which has complexity order $\in O(2^\eta K^2)$. The localized antenna/mode ranking method in Section 5.3.1.2 can be done at each terminal to reduce the choices to $(\eta+1)$ at each terminal. This reduces the number of decision metrics to

$$\sum\nolimits_{j=0}^{K'+1}(K-j)(\eta+1) = (\eta+1)(K^2 + K - K'^2 - K')/2, \tag{5.3.32}$$

which has complexity order $\in O((\eta+1)K^2)$. In all, GDS-BAS requires more decision metrics than GIS-BAS unless $K' \to K$. A detailed assessment on the computational complexity is not done yet at this stage. Taking a similar approach as that in Section 5.3.1.1, a round-robin style of de-selection may be done so that an initial group of users with the highest sum rate can be chosen. Further RAS/SMS can then be done to further improve the sum rate using the MDR algorithm and additional scheduling may be done if the pre-coding constraint is not exceeded.

5.4 Decoupled User Selection and RAS/SMS

From Section 5.3, it was shown that the required number of decision metrics is lowest (see (5.3.25) and (5.3.30)) when selection is done at the user level and not at the antenna-subset level. One means of keeping close to these numbers is to perform user selection first, followed by RAS/SMS on the chosen users. For example, the GIS-BAS algorithm is used to select K' out of K users, where $\sum_{j=1}^{K'} N_j \leq M$ to meet BD-SDM pre-coding constraints. This is followed by RAS/SMS using any RAS algorithm, e.g., the MDR algorithm from Section 5.2. Using MDR, the maximum number of decision metrics needed is

$$M + \sum_{j=0}^{K'-1} (K - j) = M + K'\left(2K - K'+1\right)/2, \qquad (5.4.1)$$

which is still $\in O(K'K)$ as desired. In fact, the number of rate evaluations in the RAS/SMS phase is expected to be $< M$ because $\left(M - N'\right) < N'$ occurs with high probability, where N' is the final number of antennas/modes after RAS/SMS.

The RAS/SMS process may create room for the scheduling of additional users when $\sum_{j=1}^{K'} N_j < M$ after RAS/SMS. Judicious scheduling may be done by means of the original user selection ranking. For example, all un-selected users would have been ranked during the last round of GIS-BAS user selection. Similarly, all de-selected users would have been ranked in GDS-BAS. Strictly speaking though, the current user and antenna set would have changed after RAS/SMS and a fresh ranking is needed. However, the previous user selection ranking could still be used as an approximation. This is particularly applicable to NBS-BAS. In this way, we strive towards the scheduling of M channels. This helps in approaching the optimal beamforming sum rate, which scales as $M \log\log KN$ [18]. Numerical results show that this iterative scheme provides significant improvement to the sum rate performance of NBS schemes. This is attractive since NBS schemes have low complexity. In addition, it is important to note that NBS schemes have low CSI feedback requirement and hence their adoption will facilitate the deployment of BD schemes in practical systems. More details are given in Section 5.6 below.

5.5 Resource Allocation in BD Systems

Resource allocation involves power and spatial mode allocation. We focus on the latter, which requires (a) determining the number of spatial modes per user and (b) making a choice out of $\binom{N_{\max}}{N_j}$ combinations when a user j is allocated N_j modes, which is less than its maximum $N_{\max}$. The two decisions cannot be made in isolation since a choice at one user has rate impacts on other users. When a group of users are chosen based on the sum rate maximization criterion, some of the users may not have the required channel rate while others may have excess rates. This is usually still true after RAS/SMS has been performed on the chosen group to help maximize the sum rate. Performing resource allocation after sum rate maximization via RAS/SMS is the correct order of events since the poorer antennas/modes have been weeded out. Although power scaling provides a means of resource allocation, it is not efficient when users with QoS deficit have low channel rates.

The same RAS/SMS algorithms used for sum rate maximization in BD-SDM can also be used to provide a systematic mechanism for resource allocation. This is due to their ability to rank the antennas or spatial modes in an order that represent their contribution to the overall sum rate. Removing an antenna or mode with low contribution will result in a low rate loss to the affected user and a low loss to the overall sum rate. This mechanism is useful when reducing the rates of those users with excess rate in order to aid those that are lacking. One may proceed by dividing the selected user pool into 2 groups, viz., those with excess rates (Group #1) and those who are in deficit (Group #2). The rate allocation process may then proceed by eliminating the worst antenna or mode within Group #1. This may be done using any RAS/SMS algorithm, e.g., MDR. If the elimination causes a user to go from Group #1 to Group #2, undo the elimination and go for the next worse antenna or mode in Group #1. Repeat this process until all individual user rates are satisfied.

Note that a solution may not be found and other allocation policies may then be invoked, e.g., serving the higher priority users. In this case the RAS/SMS algorithms are of help again as it can identify the worst antennas and modes to be eliminated so that the overall rate loss impact is minimized. In fact, the proposed method may be used in tandem with any other scheduling methods, for example, priority ranking according to the buffer lengths.

In this way, the proposed method does away with the need to make *a priori* decisions on the number of antennas/modes at each terminal. It also solves the combinatorial problem that presents itself when subsets of antennas/modes are to be chosen at some users. Further adjustments to the final transmission rate and powers may be done via power scaling. In the case where there are still excess rates after all users are satisfied, the RAS/SMS process may be used to free resources to allow the scheduling of more users.

5.6 Reducing CSI Feedback Requirement

A major hindrance to the adoption of spatial multiplexing is the need for channel state information (CSI) at the base station, which can incur an enormous amount of overhead when the user pool is large. Zero-forcing beamforming systems require timely and accurate channel estimates for good performance and the problem is compounded when exploitation of multi-user diversity via judicious user selection is desired. In general, better sum-rate maximization is achieved when user selection schemes at the base station have access to the full channel matrix of each user under consideration. Arising from this, a major concern is that the delays associated with the CSI feedback of many users may go beyond the channel coherence time and cause processing errors. This has motivated much research effort to find ways of mitigating this problem. There are two broad approaches to mitigate the problem associated with making CSI available at the base station, namely, (a) limited-bandwidth CSI feedback and (b) partial CSI feedback.

This dissertation focuses on the latter case of partial CSI feedback during the user selection process and during the beamforming process. A straightforward method for reducing CSI feedback during user selection in ZFBF systems is to base the selection metric on the channel gain of each user. The reduction is due to fact that each channel gain value may be transmitted as a scalar to the base station, which is much less than transmitting the full channel matrix of each user. Further feedback reduction may be achieved by thresholding methods, that is, only those users who are above the pre-defined gain level will report their values back to the base station. The base station ranks the gain values and obtain the channel matrices only from the best chosen users. In this way, the amount of CSI fed back is drastically reduced compared to the case where a user selection scheme requires feedback from all potential users.

However, channel-gain based user selection results in poor ZFBF performance because the chosen channel directions may not line up well with the zero-forcing directions. To address this, the base station may employ RAS/SMS to weed out those antennas/modes that give low rate returns. This may free up transmit resources for the consideration of additional users. The original ranking list may be used and channel matrices of the next best users are obtained. The cycle repeats itself until no additional room is left or when the channel rate of the originally chosen group becomes insufficient. In this way, good sum rate performance along with CSI feedback reduction can be achieved using channel-gain based user selection schemes. This iterative process is referred to as "simultaneous scheduling and sum-rate maximization" (SSRM). Numerical results below show significant improvements to channel-gain based user selection when RAS plus SSRM are employed. In fact, channel-gain based user selection schemes become on par with the more complex schemes that required the full channel matrix of all users under consideration.

Another interesting scheme that is a variant of channel-gain based selection is proposed in [40] where a form of polling is used. The next chosen user is one with the highest projection magnitudes in the null space of a currently chosen user group. Decisions at the BS are made on a single scalar feedback from each user while full CSI feedback is required only from the chosen users, thus attaining the goal of feedback reduction. The process begins by broadcasting to all users with a pre-coding matrix $\mathbf{T}_1 = \mathbf{I}$. Each user j will compute its single-user capacity $C_j(\mathbf{H}_j\mathbf{T}_1)$ and report it to the BS. The BS will choose the first user u_1 where $u_1 = \arg\max_j \left(C_j(\mathbf{H}_j\mathbf{T}_1)\right)$ and require its channel matrix $\mathbf{H}_1$ to be fed back. Next, the base station will broadcast to all remaining users after pre-coding with $\mathbf{T}_2$ where $\mathbf{T}_2$ is the null space of $\mathbf{H}_1$. The second user u_2 where $u_2 = \arg\max_j \left(C_j(\mathbf{H}_j\mathbf{T}_2)\right)$ is then chosen. The process is repeated with pre-coding matrix $\mathbf{T}_3$ where $\mathbf{T}_3$ is the null space of $\tilde{\mathbf{H}}_2 = [\mathbf{H}_1^T \quad \mathbf{H}_2^T]^T$. In this way, a subset $S_r \subset S$ of K_r users may be chosen. The user selection approach in [40] is similar to schemes that make use of orthogonal complement projection. The adoption of the scheme in [40] requires the simplest form of block antenna selection, which is to decouple the user selection and RAS processes.

In relation to CSI feedback reduction during the ZFBF beamforming process, the analysis and results in Chapter 3 will show a possible method that is based on localized antenna/mode selection done at each user terminal, without the involvement of the base station. It is shown in Section 3.2.2 that *random* antenna/mode de-selection done in a round-robin style among users of a BD-SDM is useful for increasing the ergodic BD-SDM sum rate. For BD systems, a round-robin style of random de-selection produces results that are close to localized judicious de-selection for homogeneous channels. Since the channel matrix size to be fed back to the base station is reduced after a localized RAS (be it random or judicious), a method for reducing the feedback overhead may be developed on this basis. For example, it is shown in Figure 3.15 that around 8 antennas must be removed from the 8-user BD system to achieve the best sum rate when SNR = 20dB. This means that one antenna must be removed from each user and this reduces the size of the channel matrix to be fed back from each user to the base station. This method may be considered if CSI feedback reduction during beamforming is of paramount importance.

5.7 Proposed Integrated Process

The user selection and RAS processes may be jointly done using BAS algorithms. To accommodate CSI feedback requirement reduction schemes such as channel-gain based selection like NBS or the scheme in [40], a decoupled user selection (USEL) and RAS approach must be taken. The user selection scheme of choice may be invoked first to choose K_r users. The availability of channel sub-matrices $\mathbf{H}_1 \cdots \mathbf{H}_{K_r}$ at the BS enables RAS to be performed. Let $\mathbf{H}'_1 \cdots \mathbf{H}'_{K_r}$ be the new channel sub-matrices after RAS. If transmission resources are made available after RAS, the user selection process may be repeated. For the scheme in [40], selection may be repeated with $\mathbf{T}_{K_r}$ to choose the next user, where $\mathbf{T}_{K_r}$ is the null space of $\tilde{\mathbf{H}}'_{K_r} = [\mathbf{H}'^T_1 \quad \cdots \quad \mathbf{H}'^T_{K_r}]^T$. This USEL-RAS cycle may be repeated until M channels are scheduled.

Resource allocation may take place next, making use of the antenna/mode ranking provided by the RAS/SMS algorithms to minimize rate loss. The resource allocation method proposed in Section 5.5 may be used. In the case where there are still excess rates after all users are satisfied, the RAS/SMS process may be used to free resources to allow the scheduling of more users. The USEL-RAS process in the previous paragraph may be repeated until M channels are scheduled. In this way, CSI feedback reduction is achieved along with BD-SDM sum rate maximization and resource allocation.

5.8 Numerical Results

The presence of spatial fading correlation in $\mathbf{H}_j$ is captured by modeling the channel as $\mathbf{H}_j = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$, where $\mathbf{H}_w$ is the i.i.d. spatially white channel and $\mathbf{R}_r$ and $\mathbf{R}_t$ are positive definite Hermitian matrices that specify the receive and transmit correlations respectively. We assume that the base station antennas are well spaced enough to allow $\mathbf{R}_t = \mathbf{I}_M$ and the users are well separated enough to consider only the intra-terminal antenna correlation. An exponential correlation model is used where each element r_{ij} in $\mathbf{R}_r$ is $r_{ij} = \rho^{|i-j|}$, where ρ is the maximum correlation between two antennas at each user terminal.

5.8.1 Impact of RAS/SMS on BD Systems

Fig. 5.1 shows the ergodic sum rates of direct-BD and Nu-SVD with and without RAS/SMS for a 4-user system each with 2 antennas. The RAS/SMS is done via exhaustive search, MDR, G3 and NB, where G3 is the decremental RAS "Algorithm III" from [34] and NB is norm-based selection. As shown, RAS/SMS provides substantial sum rate gain and helps BD-SDM schemes to approach the multi-user sum capacity (derived using [73]). The MDR algorithm is near optimal for both direct-BD and Nu-SVD. The MDR performance is slightly better than the decremental RAS scheme named Algorithm III from [34] and named as G3 in Fig. 5.1. This is pleasing as MDR has computational complexity of $\in O(M^3)$ while G3 is upper bounded by $\in O(M^5)$. The NB-RAS algorithm does not perform as well as MDR but is attractive from the complexity viewpoint.

Fig. 5.2 compares the performance of direct-BD, CTR and Nu-SVD using the MDR and PSME algorithms for RAS/SMS. A larger system comprising 8 users, each with 4 antennas is used. As shown, all three BD-SDM schemes benefited from RAS/SMS. As expected, Nu-SVD provides the best performance among the three schemes. MDR and PSME provide the same results for Nu-SVD. However, PSME incurs negligible computational load compared to MDR and is therefore the preferred means of SMS for Nu-SVD. Both MDR and PSME are used for CTR and as expected, MDR provides better performance since a one-to-one correspondence between the spatial modes and the columns of $\mathbf{W}_j$ does not exist in CTR. In all, Nu-SVD provides the best performance but it is computationally more expensive as it is an iterative algorithm. CTR is attractive in that its performance is only slightly worse than Nu-SVD and it provides a means of mode selection at a computational cost that is practically the same as direct-BD.

5.8.2 Example of Additional User Scheduling

As described before, the RAS/SMS process may free resources that allow the scheduling of additional users. The BD-SDM sum rate may increase in the process since it strives towards the scheduling of M channels, which is needed when trying to approach the optimal beamforming sum rate. Fig. 5.3 shows a direct-BD system with 32 transmit antennas and users that are equipped with 4-antenna terminals. A group of 8 users is initially selected and Fig. 5.3(a) shows the ergodic sum rate of this 8-user group with and without RAS via MDR. Fig. 5.3(a) also shows the impact of scheduling the 9th user whenever resources are available. We see an improvement in the ergodic sum rate. The impact on an arbitrary individual user's ergodic channel rate is also shown in Fig. 5.3(b).

As shown, the addition of the 9th user causes a slight drop in the individual channel rate. This drop is expected to continue as more users are added. The process of adding users is stopped when the overall sum rate drops or when no more free resources are available.

5.8.3 Comparison of BAS Algorithms and Impact of RAS

In Fig. 5.4, a total of four BAS algorithms that do not perform rate evaluations during user selection are compared against the algorithm in [33], which uses direct-BD precoding during user selection. Selection via exhaustive search is also given for comparison. The four BAS algorithms are (a) GIS-BAS (see Section 5.3.2.3.3), (b) GDS-BAS (see Section 5.3.2.3.4), (c) PMP-BAS [79] and (d) NBS-BAS (see Section 5.3.2.1). The pairwise mutual projection BAS (PMP-BAS) scheme is based on a pair-wise metric that jointly measures the correlation and row vector norm. The simulation involves choosing 4 users out of a pool of 8 users, each equipped with 4 antennas.

As shown, the differences in performance for GDS-BAS, GIS-BAS, PMP-BAS and NBS-BAS compared to Shen's algorithm in [33] are 1.5%, 4.4%, 7.2% and 7.5% respectively when intra-terminal correlation is 0.0. We see that the tradeoff in performance is not great compared to the reduction in complexity. This is especially so when RAS is applied on the chosen user subset. This makes norm-based (or channel gain based) NBS schemes more attractive in practice. The difference in performance becomes negligible for GDS-BAS and GIS-BAS when compared to Shen's algorithm in [33] and exhaustive search.

5.8.4 Combining Decoupled User Selection, RAS and SSRM

In Fig. 5.5, user selection is done first, that is, all antennas of each chosen user are included. An RAS exercise is done next and this is followed by the simultaneous scheduling and sum rate maximization (SSRM) process. The comparison is made for Shen's algorithm in [33], the user selection algorithms based on the GDS and GIS algorithms from [34] and the norm-based algorithm NBS. As shown, the NBS algorithm benefits the most from RAS and SSRM. In this way, RAS and SSRM improves the feasibility of deploying NBS, which lowers implementation complexity as well as reducing the CSI feedback requirement.

5.8.5 Comparing Single-Antenna Selection (SAS) with Decoupled User Selection, RAS and SSRM

In Fig. 5.6, compares the single-antenna selection approach with the decoupled user-antenna selection approach. It shows the sum rate performance as well as the number of users scheduled. The algorithms from [36] and [38] are compared with the GDS-MDR-SSRM combination. As shown, [36] achieves high sum rates at the expense of low individual channel rates because the number of users scheduled is high. The block selection approach is achieves a lower sum rate but provides higher individual rates by scheduling a lower number of users. For comparison with [38], the incremental RAS algorithm from [34] is used for antenna ranking. The results show that the antenna ranking approach has potential for good performance with lower complexity than [36].

5.9 Summary

Efficient and near-optimal algorithms for RAS/SMS are developed for the case where the total number of receive antennas or spatial modes is within the block diagonalization pre-coding constraint. The algorithms provide spatial channel ranking and can therefore be used to provide a systematic method for resource allocation to meet the individual QoS needs of the scheduled group. This rate-loss minimization approach provides a systematic approach to address the impact on all other users when resource allocation is done at any one user. In addition, method does away with the need to make *a priori* decisions on the number of antennas/modes at each terminal. It also solves the combinatorial problem that presents itself when subsets of antennas/modes are to be chosen at some users.

Efficient algorithms for joint user selection and RAS/SMS to maximize the sum rate. To allow joint selection, the concepts of "block antenna selection (BAS)" and "block mode selection (BMS)" are introduced, which account for differences in intra- and inter-terminal processing in block diagonalized systems. A novel approach is based on the modification of existing RAS algorithms is proposed. It has good performance and low complexity, which is realized by avoiding repeated use of BD pre-coding during selection. It allows for decremental selection, which has potential for better performance than incremental selection. An equivalent method for "simultaneous scheduling and sum rate maximization" or SSRM is developed to allow scaling with $M \log\log KN$ (1.4). This method gives significant sum rate improvement for channel-gain based user selection, which have lower processing complexity and significantly lower CSI feedback requirements during user selection.

Fig. 5.1 Direct-BD and Nu-SVD with and without RAS/SMS

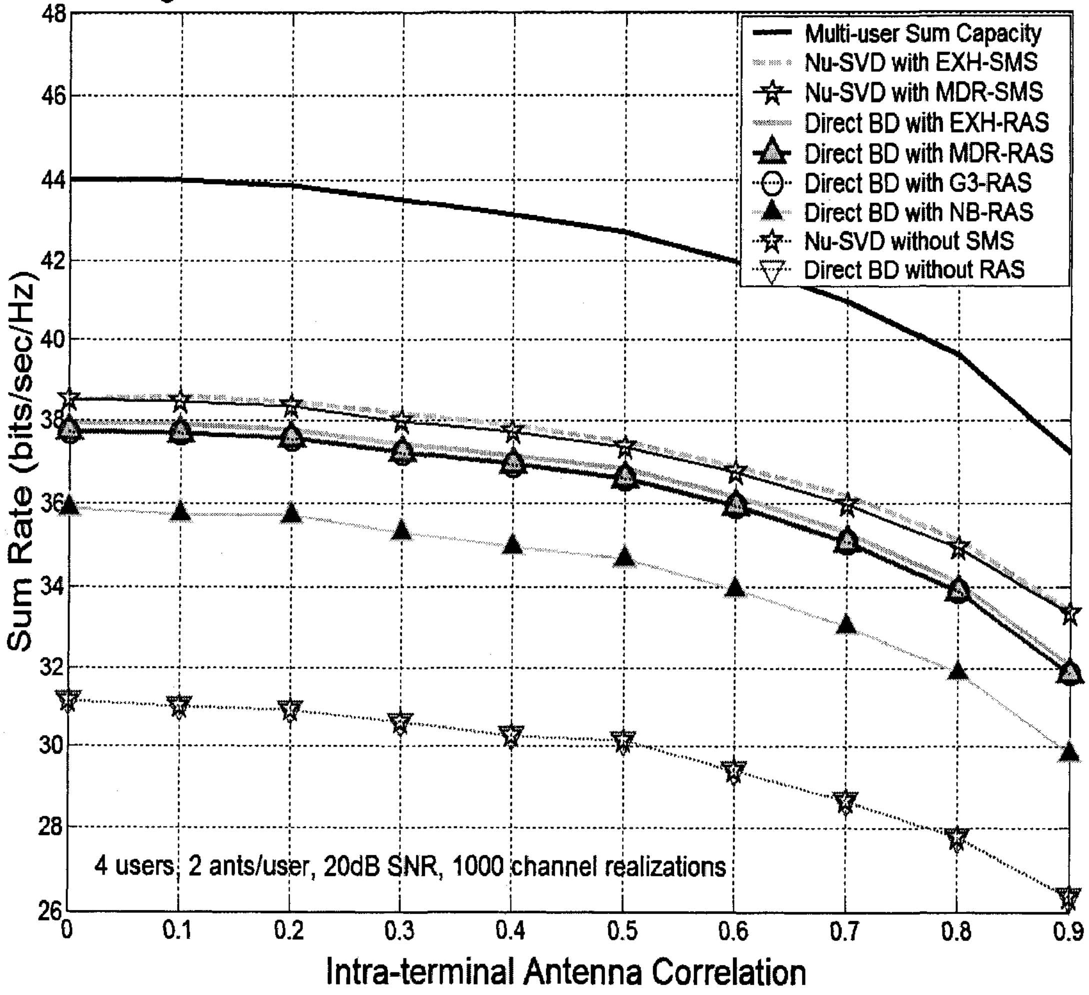
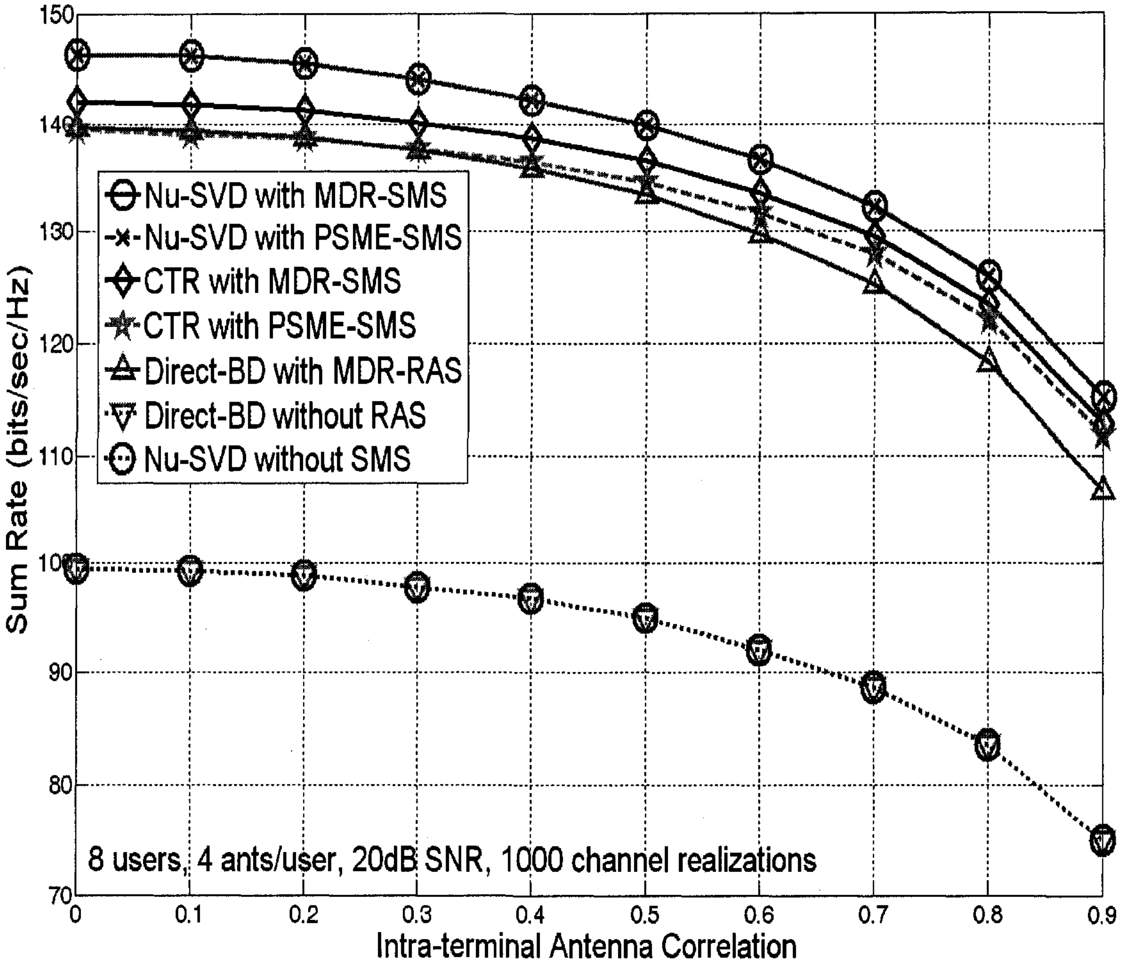


Fig. 5.2 Direct-BD, Nu-SVD and CTR with and without RAS/SMS



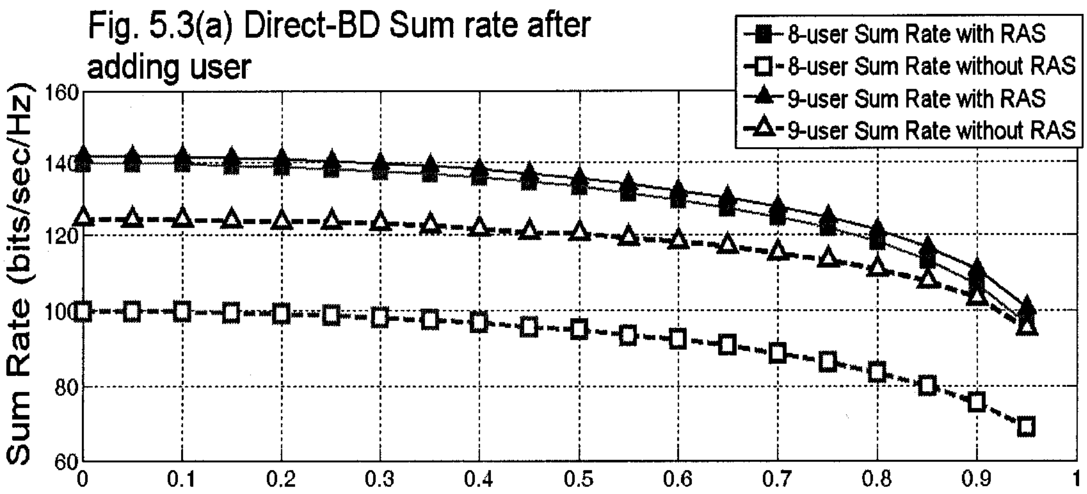


Fig. 5.3(a) Direct-BD Sum rate after adding user

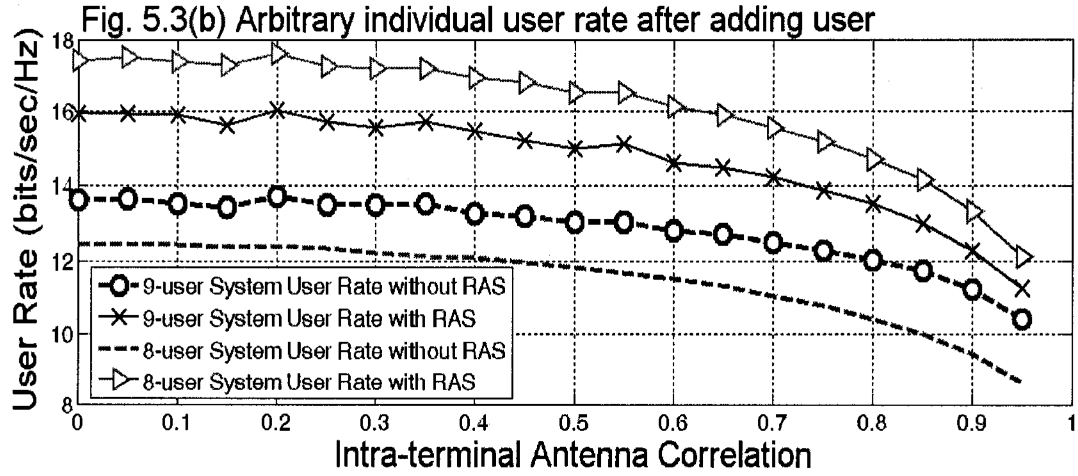


Fig. 5.3(b) Arbitrary individual user rate after adding user

Fig. 5.4 BAS Algorithms: Comparison & Impact of RAS

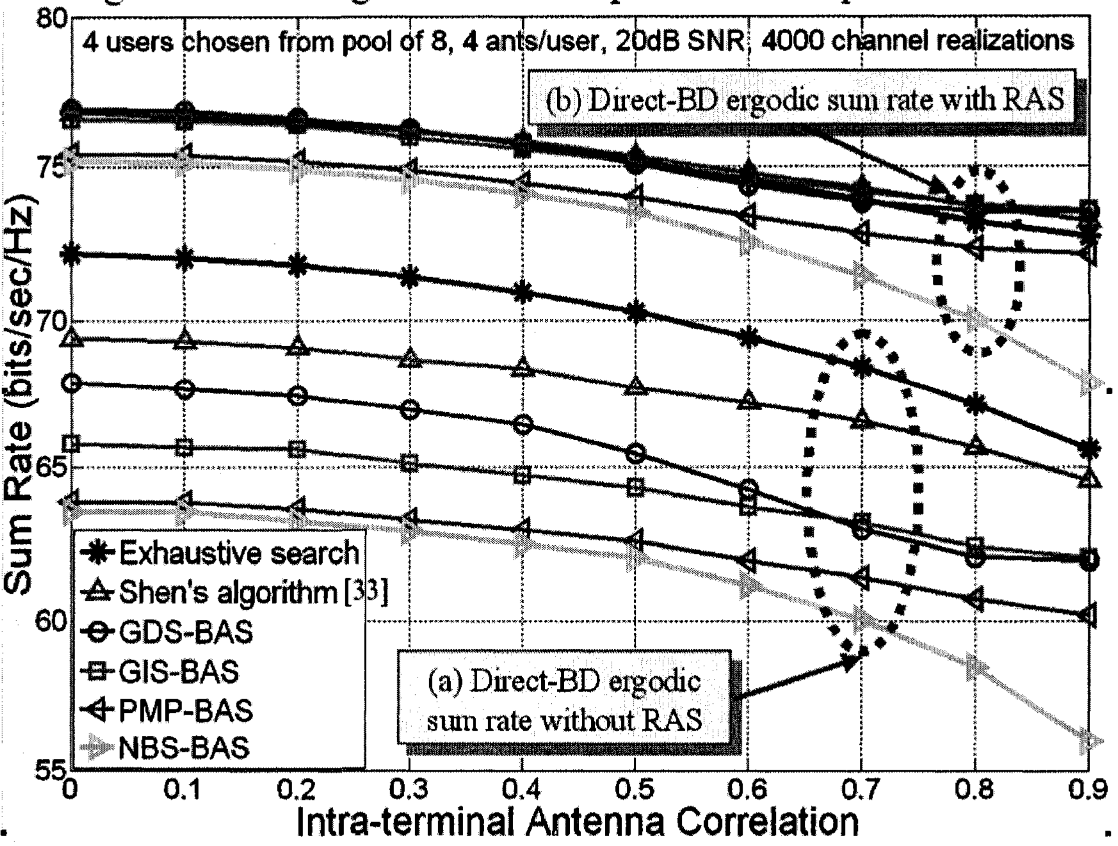


Fig. 5.5. Decoupled USEL-RAS with SSRM*

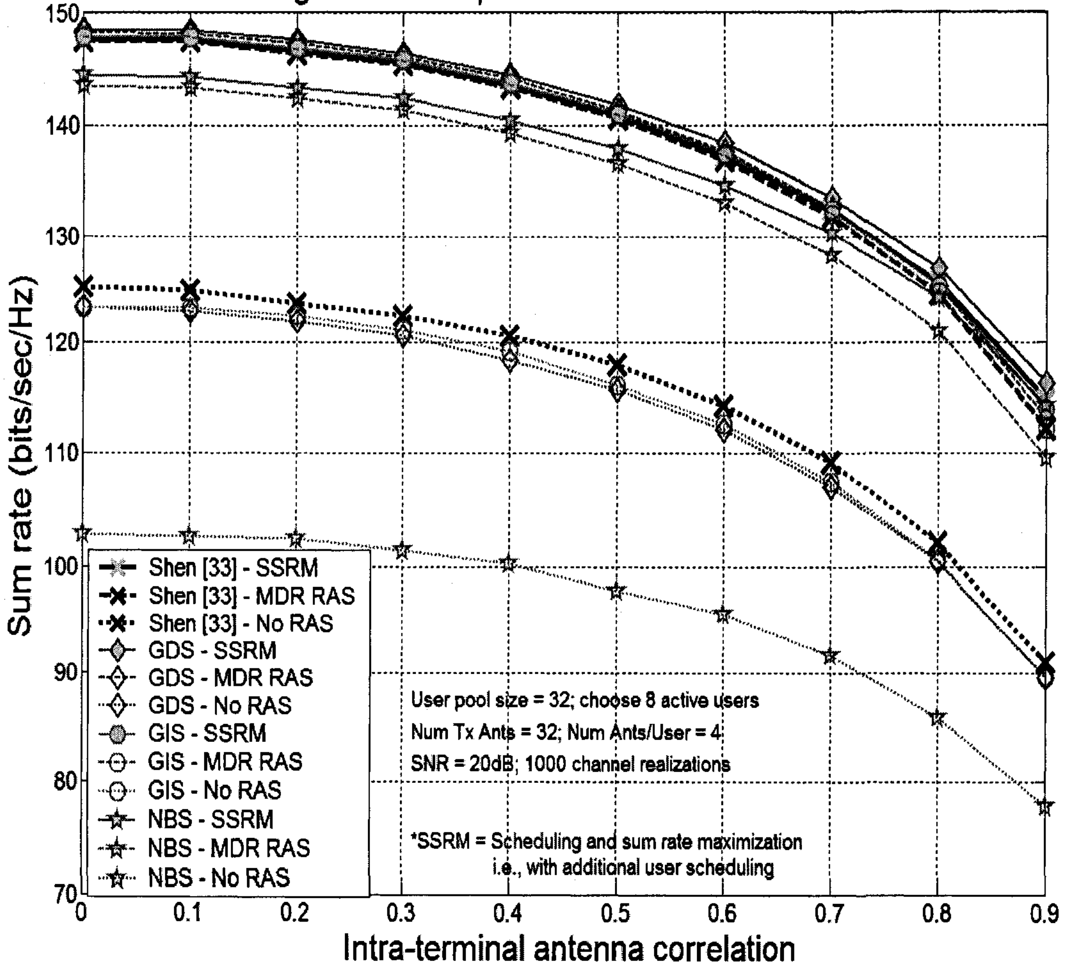
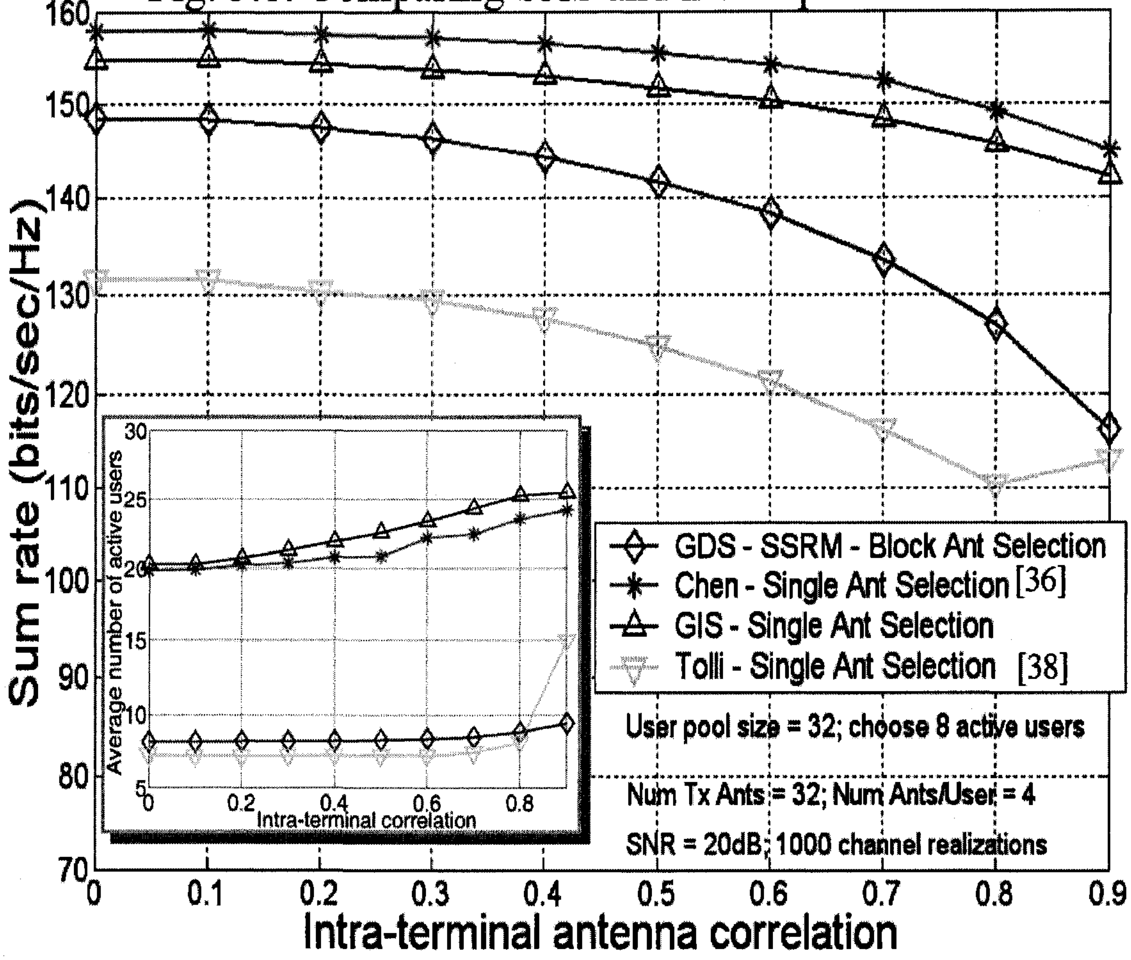


Fig. 5.6. Comparing SAS and Decouple BAS-RAS



Chapter 6

CONCLUSIONS AND FUTURE WORK

In line with the intent to enhance the feasibility of deploying multi-user MIMO techniques in practical downlink systems, the focus of this thesis has been on issues relating to transmit zero-forcing beamforming (TZFBF). Transmit ZFBF is a linear processing technique and is a potential candidate for multi-user downlinks given its low complexity compared to other processing schemes. However, its lower complexity is accompanied by a setback in terms of a sum rate performance gap compared to optimal schemes like dirty paper coding (DPC). Like other spatial multiplexing techniques, it also faces challenges in user selection, resource allocation and system overhead demands.

It is shown in this thesis that receive antenna selection (RAS) is necessary for sum rate maximization when multi-antenna terminals are served via block diagonalized space-division multiplexing (BD-SDM). Sole reliance on user selection to exploit multi-user diversity from a large potential user pool does not achieve the best sum rates. It is shown that optimal user selection is actually subsumed under an exhaustive RAS process. The introduction of RAS to block diagonalized systems helps achieve higher sum rates that narrows the performance gap with DPC even when the user pool sizes are small. Specifically, RAS provides significant sum rate improvements to BD systems since it helps them to scale with M (the number of transmit antennas) even when the potential user pool is small. Hence, the incorporation of RAS enhances the feasibility of deploying block diagonalized systems in practical systems.

For single-antenna terminals, the user selection and RAS processes become identical. For BD systems that use receive-weight matrices as a means of spatial mode allocation in multi-antenna terminals, RAS becomes spatial mode selection (SMS). By analyzing the conditions for sum rate increase, efficient selection algorithms for joint user selection and RAS/SMS are developed for both multi-antenna and single-antenna terminals that work under homogeneous and heterogeneous channel conditions. Invoking the Sato upper bound, a class of low complexity selection algorithms for the multi-user environment is derived from existing RAS algorithms that are meant for single-user MIMO systems. The block antenna selection concept is introduced to enhance the performance of this approach. The analysis also provides novel expressions for ergodic sum rate bounds that jointly reflect the impacts of user selection, RAS/SMS, SNR levels and number of transmit antennas.

On the challenge posed by resource allocation, a key issue centers on mode selection for each user since any selection done at one terminal affects the rates achieved at all other terminals. This causes a departure from the best sum rate when a sum rate maximization process has already been done. A systematic method is introduced that performs resource allocation to meet the individual user throughput requirements while minimizing rate loss at the overall- and individual levels. The method does away with the need to make *a priori* decisions on the number of antennas/modes at each terminal. It also solves the combinatorial problem that presents itself when subsets of antennas/modes are to be chosen at some users. Ordering resource allocation after sum rate maximization via RAS/SMS is expedient since the poorer antennas/modes that were committed to lower rate returns have been removed. In addition, lower complexity is achieved by exploiting the antenna/mode ranking arising from the sum rate maximization process.

On the challenge of reducing system overheads, this thesis has focused on aspects of partial channel state information (CSI) feedback. Specifically, selection algorithms that are based on channel gains are attractive since only a scalar feedback to report the SNR level is needed from each user for the purpose of selection. Detailed channel matrices are then required only from the chosen users. However, the sum rate performance of channel-gain based selection is usually poor due to the fact that inter-user and intra-terminal correlations are not taken into account. To address this, an iterative scheme is introduced to help channel-gain based selection algorithms schemes scale closer to $M \log\log KN$. It is significant to note that this iterative scheme for block diagonalized systems is made possible primarily due to the ability of RAS/SMS to remove antennas/modes that are committed to poor rate returns. A streamlined process that integrates the sum rate maximization exercise (via user selection and RAS/SMS), the resource allocation process, and the partial CSI feedback scheme is introduced. Taken together, the streamlined process helps to improve the feasibility of deploying multi-user MIMO techniques in practical systems.

Suggestions on future work include the following:

(a) Consideration of the above techniques for relay networks.

(b) Consideration of the above techniques for cooperative methods in cellular systems. This includes cooperation among base stations and relay nodes.

(c) Improving the analytical bounds for the ergodic sum rate expressions.

(d) Combining the above techniques with other dimensions like time and frequency, that is, in the areas of time-division multiplexing (TDM) and frequency-division multiplexing (FDM), including OFDM.

(e) Consideration of the above techniques in the area of inter-cell interference control. Methods may capitalize on the fact that RAS in block diagonalized systems frees transmission resources that could be used for interference control. The methods could be done cooperatively among base stations to maximize the overall system throughput.

(f) Improvements to the partial CSI feedback schemes proposed so far. One consideration is to exploit the processing ability of the receiver to improve the feedback metric.

(g) Implementing a test-bed for field tests.

REFERENCES

[1] International Telecommunication Union, "ITU/MIC workshop on shaping the future mobile information society," *Document: SMIS/0S 19 February 2004*, Seoul, March 2004.

[2] T. Maseng and R. Landry, "Network-centric military communications," *IEEE Comms. Mag.*, vol. 45, no. 10, pp. 42 – 44, Oct. 2007.

[3] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," US Patent No. 5,345,599, 1993.

[4] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," ATT Bell Laboratories, Tech. Memo., 1995.

[5] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, Autumn 1996.

[6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Commun.*, vol. 10, no. 6, pp. 585–595, 1999.

[7] L. Zheng and D.N.C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[8] D. Soldani and S. Dixit, "Wireless relays for broadband access," *IEEE Comms. Mag.*, vol. 46, no. 3, pp. 58–66, Mar. 2008.

[9] B. Walke, H. Wijaya and D. Schultz, "Layer 2 relays in cellular mobile radio networks," *IEEE VTC-2006 Spring*, Melbourne, Australia, May 2006.

[10] C.B. Peel, et al., "Linear and dirty-paper techniques for the multiuser MIMO downlink," *Space-Time Processing for MIMO Communications*, chapter 6. Chichester, England: Wiley 2005.

[11] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol.49, no. 7, pp. 1691-1706, Jul. 2003.

[12] P. Viswanath and D.N.C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink–downlink duality," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[13] S. Vishwanath, N. Jindal and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," *IEEE Int. Conf. Commun.*, vol. 3, pp. 1444–1450, Apr. 2002.

[14] W. Yu and J.M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[15] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol.29, no. 3, pp. 439–441, May 1983.

[16] C.B. Peel, "On "Dirty-Paper Coding"," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 112–113, May 2003.

[17] H. Weingarten, Y. Steinberg and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[18] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 11–15, Jan. 2007.

[19] C.B. Peel, B.M. Hochwald and A.L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no.3, pp. 537–544, Mar. 2005.

[20] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, Mar. 1971.

[21] H. Harashima and H. Miyakawa, "A method of code conversion for digital communication channels with intersymbol interference," *Transactions of the Institute of Electronics and Communications Engineers of Japan*, pp. 272–273, Jun. 1969.

[22] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: John Wiley & Sons, 2002.

[23] S. Venkatesan and H. Huang, "System capacity evaluation of multiple antenna systems using beamforming and dirty paper coding," Bell Labs, 2003.

[24] N. Jindal and A. Goldsmith, "Dirty-paper coding vs. TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1783–1794, 2005.

[25] C.B. Peel, B.M. Hochwald and A.L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[26] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[27] Q.H. Spencer, A.L. Swindlehurst and M.H. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Sig. Proc.*, vol.52, no.2, pp.461-471, Feb. 2004.

[28] L. Choi and R.D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol.3, no.1, pp.1969-1973, Jan.2004.

[29] Z. Pan, K.K. Wong and T.S. Ng, "Generalized multiuser orthogonal space division multiplexing," *IEEE Trans. Wireless Commun.*, vol.3, no.6, pp.1969-1973, Nov. 2004.

[30] G.G. Raleigh and J.M. Cioffi, "Spatio-temporal coding for wireless communication", *IEEE Trans. Commun.*, vol.46, no. 3, pp.357-366, Mar 1998.

[31] J. Jiang, M. Buehrer and W.H. Tranter, "High-speed downlink packet transmission with spatial multiplexing and scheduling," in *Proc. IEEE Wireless Commun. and Networking Conf.* (*WCNC'04*), vol.4, pp. 2148-2152, Mar. 2004.

[32] A. Bayesteh and A.K. Khandani, "On the user selection for MIMO broadcast channels," in *Proc. IEEE Int'l Symp. Inform. Theory* (*ISIT'05*), pp.2325-2329, Sep. 2005.

[33] Z. Shen, et.al.,"Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *39th Asilomar Conf. On Signals, Systems & Computers*, pp.628-632, Oct. 2005.

[34] A. Gorokhov, D. Gore and A. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: theory and algorithms", *IEEE Trans. Sig. Proc.*, vol. 51, no. 11, pp. 2796–2807, Nov. 2003.

[35] B.C. Lim, C. Schlegel and W.A. Krzymień, "Sum rate maximization and transmit power minimization for multi-user orthogonal space division multiplexing," in *Proc. Globecom-06,* Nov. 2006.

[36] R. Chen, J. Andrews, R. Heath and Z. Shen, "Low-complexity user and antenna selection for multiuser MIMO systems with block diagonalization," in *Proc. IEEE Int'l Conf. Acoustics, Speech and Sig. Proc. (ICASSP 2007),* pp. III-613 – III-616, Apr. 2007.

[37] Z. Shen, et. al., "Sum capacity of multiuser MIMO broadcast channels with block diagonalization," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2040 – 2044, Jun 2007

[38] A. Tolli and M. Juntti, "Scheduling for multiuser MIMO downlink with linear processing," in *Proc. IEEE Int'l Symp. Personal, Indoor and Mobile Radio Commun., (PIMRC '05)*, pp. 156 – 160, Sep. 2005.

[39] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[40] H. Zheng, et. al., "An efficient user selection algorithm for zero-forcing beamforming in downlink multiuser MIMO systems," *IEICE Trans. Commun.*, Vol. E89-B, No. 9, pp. 2641 – 2645, Sep. 2006.

[41] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna arrays with feedback," *IEEE Sig. Proc. Letters*, vol. 1, pp. 150–152, Oct. 1994.

[42] T. Haustein, C. von Helmolt, E. Jorwieck, V. Jungnickel and V. Pohl, "Performance of MIMO systems with channel inversion," in *Proc. IEEE Veh. Tech. Conf.*, vol. 1, pp. 35–39, May 2002.

[43] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: Wiley, 1991.

[44] I. Berenguer, I.J. Wassell and X. Wang, " Opportunistic user scheduling and antenna selection in the downlink of multiuser MISO systems," in *Proc. IEEE Veh. Tech. Conf.*, vol.2, pp. 1138-1142, May 2005.

[45] Z. Tu and R. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Commun. Lett.*, vol. 8, pp. 370–372, 2003.

[46] J. Jiang, R.M. Buehrer and W.H. Tranter, "Greedy scheduling performance for a zero-forcing dirty-paper coded system," *IEEE Trans. Commun.*, vol. 54, pp. 789–793, May 2006.

[47] Q.H. Spencer and A.L. Swindlehurst, "Channel allocation in multi-user MIMO wireless communications systems," in *Proc. IEEE Int. Conf. on Commun.*, vol. 5, pp. 3035–3039, Jun. 2004.

[48] M. Xu and D. Lin, "Low-complexity user selection strategies in the downlink of multi-user channels," in *Proc. IEEE Conf. on Adv. Commun. Tech., ICACT'06*, vol. 1, pp. 204–206, Feb. 2006.

[49] K.P. Jagannathan, S. Borst, P. Whiting and E. Modiano, "Efficient scheduling of multi-user multi-antenna systems," in *Proc. IEEE Conf. on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pp. 1–8, Apr. 2006

[50] G. Dimić and N.D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. on Sig. Proc.*, vol. 53, pp. 3857–3868, Oct. 2005.

[51] Z. Shen, et. al., "Sum capacity of multiuser MIMO broadcast channels with block diagonalization," *Proc. IEEE Int. Symp. on Inform. Theory*, pp. 886–890, Jul. 2006.

[52] R. Knopp and P.A. Humblet, "Information capacity and power control in single-cell multiuser communications," *Proc. IEEE Int. Conf. Commun.*, vol. 1, pp. 331–335, Jun. 1995.

[53] P. Viswanath, D.N.C. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, Jun. 2002.

[54] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 374–377, May 1978.

[55] M. Gharavi-Alkhansari and A. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 339–347, Feb. 2004.

[56] Y.S. Choi, A.F. Molisch, M.Z. Win and J.H. Winters, "Fast algorithms for antenna selection in MIMO systems," in *Proc. IEEE Veh. Tech. Conf.*, pp. 1733 – 1737, Oct. 2003.

[57] Q.H. Spencer and A.L. Swindlehurst, "Channel allocation in multi-user MIMO wireless communications systems," in *Proc. IEEE Int. Conf. on Commun.*, vol. 5, pp. 3035 – 3039, Jun. 2004.

[58] Y. Wu, et. al., "Receive antenna selection in the downlink of multiuser MIMO systems," in *Proc. VTC05-Fall*, pp. 477 – 481, Sep. 2005.

[59] B. Hochwald and S. Vishwanath, "Space-time multiple access: linear growth in the sumrate, " in *Proc. 40th Annual Allerton Conf. Commun., Control, & Computing*, Allerton, IL, Oct. 2002.

[60] H. Viswanathan, S. Venkatesan and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 6, pp. 802–811, Jun. 2003.

[61] P. Almers, et. al., "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, 2007.

[62] H.A. David, *Order Statistics, 2^{nd} Ed.* New York: John Wiley & Sons, Inc., 1981.

[63] A. Edelman, "Eigenvalues and condition numbers of random matrices," *MIT PhD Dissertation*, 1989.

[64] D. Hertz, "Simple bounds on the extreme eigenvalues of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp 175 – 176, Jan. 1992.

[65] E. Jorswieck, G. Wunder, V. Jungnickel and T. Haustein, "Inverse eigenvalue statistics for Rayleigh and Rician MIMO channels," *IEE Seminar on MIMO*, pp. 3/1 – 3/6, Dec. 2001.

[66] V. Jungnickel, T. Haustein, E. Jorswieck and C. von Helmolt, "On linear preprocessing in multi-antenna systems," *IEEE Globecom '02*, vol. 1, pp. 1012 – 1016, Nov. 2002.

[67] Lütkepohl H. *Handbook of Matrices*. John Wiley & Sons: Chichester, 1996.

[68] R. J. Muirhead, *Aspects of Multvariate Statistical Theory*. New York: Wiley, 1982,

[69] A. Paulraj, R. Nabar and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[70] A.M. Tulino and S. Verdu, Random Matrix Theory and Wireless Communications. Hanover: Now Publishers, 2004.

[71] G. Song and Y. Li, "Asymptotic throughput analysis for channel-aware scheduling," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1827 – 1834, Oct. 2006.

[72] V. K. N. Lau, "Proportional fair space-time scheduling for wireless communications," *IEEE Trans. Commun.*, vol. 53, no. 8, Aug. 2005.

[73] N. Jindal, W. Rhee, S. Vishwanath, S.A. Jafar and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[74] B.C. Lim, W.A. Krzymień and C. Schlegel, "Efficient sum rate maximization and resource allocation in block diagonalized space-division multiplexing," Accepted for publication in *IEEE Trans. Veh. Tech.*, Dec 2007.

[75] A.F. Molisch, "MIMO systems with antenna selection – an overview," *Radio & Wireless Conf.,* Boston, MA., USA, pp.167-170, Aug. 2003.

[76] B.C. Lim, W.A. Krzymień and C. Schlegel, "Transmit antenna selection for sum rate maximization in transmit zero-forcing beamforming," in *Proc. 10th Int. Conf. on Commun. Systems, ICCS'06 Singapore,* Oct/Nov. 2006.

[77] M. Gharavi-Alkhansari and A. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Sig. Proc.*, vol.52, no.2, pp.339-347, Feb 2004.

[78] B.C. Lim, C. Schlegel and W.A. Krzymień, "Efficient receive antenna selection algorithms and framework for transmit zero-forcing beamforming," in *Proc. VTC-06 Spring*, May 2006.

[79] B.C. Lim, C. Schlegel and W.A. Krzymień, "Antenna selection and waterfilling for transmit channel-inversion beamforming under poor channel conditions," in *Proc. Wireless 05, Calgary, Canada,* Jul. 2005.

[80] A. Gorokhov, M. Collados, D.A. Gore and A. Paulraj, "Transmit/receive MIMO antenna subset selection," in *Proc. ICASSP*, pp., II-13 – II-16, May 2004.

APPENDIX A

Proof for Equation (3.8)

The Inclusion Principle [Lütkepohl 1996] states that for a (n x n) Hermitian matrix **A** with eigenvalues $\lambda_1(\mathbf{A})\leq \ldots \leq\lambda_n(\mathbf{A})$, a principal ($q$ x q) sub-matrix $\mathbf{A}_{(q)}$ of **A** with eigenvalues $\lambda_1(\mathbf{A}_{(q)})\leq \ldots \leq\lambda_q(\mathbf{A}_{(q)})$:

$$\text{(a)}\quad \lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{A}_{(q)}) \leq \lambda_{n-q+i}(\mathbf{A});\ i=1,\cdots,q\,; \tag{A.1}$$

$$\text{(b)}\quad \lambda_{\min}(\mathbf{A}) \leq \lambda_{\min}(\mathbf{A}_{(q)}) \leq \lambda_{\max}(\mathbf{A}_{(q)}) \leq \lambda_{\max}(\mathbf{A}). \tag{A.2}$$

The matrices $\mathbf{A}_{(q)}=[\mathbf{a}_1\cdots\mathbf{a}_q]$ are the principal sub-matrices of **A**, where $\mathbf{a}_i \in \mathbb{C}^{q\text{x}1}$ and $q=1,\cdots,(n-1)$. It is clear from (A.1) that

$$\lambda_{\min}(\mathbf{A}) \leq \lambda_{\min}(\mathbf{A}_{(q)}) \leq \lambda_2(\mathbf{A}) \leq \lambda_2(\mathbf{A}_{(q)}) \leq \cdots \leq \lambda_{\max}(\mathbf{A}_{(q)}) \leq \lambda_{\max}(\mathbf{A}) \tag{A.3}$$

In the context of this paper, we are interested in the sub-matrices $\mathbf{A}_{ii}$ that are associated with the cofactors A_{ii} of the diagonal elements a_{ii} of **A**. These sub-matrices $\mathbf{A}_{ii}$ are different from the principal sub-matrices $\mathbf{A}_{(q)}$ of **A** and hence it is not clear if (A.3) is applicable to $\mathbf{A}_{ii}$.

> *Lemma A.1*: The Inclusion Principle is applicable to the sub-matrices $\mathbf{A}_{ii}$ that are associated with the cofactors A_{ii} of the diagonal elements a_{ii} of **A**.

Proof: Let $\mathbf{A}=[\mathbf{B}\mathbf{B}^H]$ be a Hermitian matrix arising from a complex matrix **B**. The elimination of row i in **B** corresponds to the elimination of row i and column i in **A**. Let the new sub-matrix be $\mathbf{A}_{ii}$ after the eliminations in **A**. The elements in $\mathbf{A}_{ii}$ constitute the elements for the calculation of the cofactor A_{ii} that is associated with the diagonal element a_{ii} in **A**. Next, any sub-matrix $\mathbf{A}_{ii}$ may be identified as a principal sub-matrix $\mathbf{A}_{(n-1)}$ of **A** that is obtained after moving the i^{th} row and i^{th} column in **A** to their respective last n^{th} row/column positions. This is done by a series of adjacent row/column exchanges until the i^{th} row/column reaches the n^{th} row/column positions. Let this re-arranged $n \times n$ matrix be $\mathbf{A}_x$. It is important to note that:

(a) Each row exchange is matched by another column exchange.

(b) The principal diagonal element associated with the i^{th} row and i^{th} column in **A** remains as a principal diagonal element when shifted to the n^{th} row, n^{th} column position (see example below). Note that the other elements in the original i^{th} row and i^{th} column may be shifted, however, the goal of mapping $\mathbf{A}_{ii}$ as a principal sub-matrix $\mathbf{A}_{(n-1)}$ of **A** is achieved.

(c) All principal diagonal elements of **A** remain as the principal diagonal elements of $\mathbf{A}_x$.

The eigenvalues of $\lambda_i(\mathbf{A}_x)$ are the same as $\lambda_i(\mathbf{A})$ because of the following:

(a) Since the formation of $\mathbf{A}_x$ preserves the elements in the principal diagonal of **A**, then $\text{trace}(\mathbf{A}_x) = \text{trace}(\mathbf{A})$ and this means that

$$\sum_{i=1}^{n} \lambda_i(\mathbf{A}_x) = \sum_{i=1}^{n} \lambda_i(\mathbf{A}). \tag{A.4}$$

(b) Since each row exchange is matched by another column exchange, then $\det(\mathbf{A}_x) = \det(\mathbf{A})$ and this means that

$$\prod_{i=1}^{n} \lambda_i(\mathbf{A}_x) = \prod_{i=1}^{n} \lambda_i(\mathbf{A}). \tag{A.5}$$

The condition under which (A.4) and (A.5) could be simultaneously satisfied must therefore be $\lambda_i(\mathbf{A}_x) = \lambda_i(\mathbf{A})$. Hence the Inclusion Principle as represented by (A.1) and (A.2) can be applied to the sub-matrices $\mathbf{A}_{ii}$ that are associated with the cofactors A_{ii} of the diagonal elements a_{ii} of **A**. □

Example

This shows how a sub-matrix $\mathbf{A}_{ii}$ that is associated with the cofactor A_{ii} of the diagonal element a_{ii} of $\mathbf{A}$ could be mapped as a principal sub-matrix $\mathbf{A}_{(n-1)}$ of $\mathbf{A}$. Given a matrix $\mathbf{A} \in \mathbb{R}^{4\times4}$ with the following entries captured in a table:

1	2	23	5
1	6	2	8
19	4	6	12
13	14	15	16

The cofactor A_{22} for entry a_{22} in $\mathbf{A}$ is desired and the associated matrix is

$$\mathbf{A}_{22} = \begin{bmatrix} 1 & 23 & 5 \\ 19 & 6 & 12 \\ 13 & 15 & 16 \end{bmatrix}$$

The following is a series of row and column exchanges to map $\mathbf{A}_{22}$ as a principal sub-matrix $\mathbf{A}_{(3)}$ of $\mathbf{A}$.

Exchange 2nd and 3rd rows

1	2	23	5
9	4	6	12
1	6	2	8
13	14	15	16

Exchange 2nd and 3rd columns

1	23	2	5
19	6	4	12
1	2	6	8
13	15	14	16

Exchange 3rd and 4th rows

1	23	2	5
19	6	4	12
13	15	14	16
1	2	6	8

Exchange 3rd and 4th columns

1	23	5	2
19	6	12	4
13	15	16	14
1	2	8	6

APPENDIX B

Recursive Matrix Inverse

The objective is to find $\mathbf{A}^{-1}$ given $\mathbf{Z}^{-1}_{(n+1)\text{x}(n+1)} = \begin{bmatrix} \mathbf{A}_{(n\text{x}n)} & \mathbf{B}_{(n\text{x}1)} \\ \mathbf{C}_{(1\text{x}n)} & \mathbf{D}_{(1\text{x}1)} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{q}_{11,(n\text{x}n)} & \mathbf{q}_{12,(n\text{x}1)} \\ \mathbf{q}_{21,(1\text{x}n)} & \mathbf{q}_{22,(1\text{x}1)} \end{bmatrix}$.

Then from [31], $\mathbf{q}_{11} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$, $\mathbf{q}_{12} = -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$, $\mathbf{q}_{21} = -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$ and $\mathbf{q}_{22} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$.

Hence $\mathbf{q}_{11} = \mathbf{A}^{-1} + \underbrace{\mathbf{A}^{-1}\mathbf{B}\mathbf{q}_{22}}_{-\mathbf{q}_{12}}\underbrace{\mathbf{q}_{22}\mathbf{C}\mathbf{A}^{-1}}_{-\mathbf{q}_{21}} / \mathbf{q}_{22}$ and therefore

$$\mathbf{A}^{-1} = \mathbf{q}_{11} - \frac{\mathbf{q}_{12}\mathbf{q}_{21}}{\mathbf{q}_{22}}. \tag{B1}$$

Equation (B1) serves as the basis for recursive inverse when one row is removed at a time in TZFBF. To show this, we examine the applicability of (B1) when $\mathbf{Z} = \mathbf{H}\mathbf{H}^{H}$ where $\mathbf{H}$ is $(n+1)\text{x}(n+1)$. Let $\mathbf{H}_s = \mathbf{P}\mathbf{H}$ where $\mathbf{P}$ is the permutation matrix that switches any two rows in $\mathbf{H}$. Then $\mathbf{H}_s\mathbf{H}_s^{H} = (\mathbf{P}\mathbf{H})(\mathbf{P}\mathbf{H})^{H} = \mathbf{P}\left(\mathbf{H}\mathbf{H}^{H}\right)\mathbf{P}^{H}$, which represents a row and column switch in $\mathbf{H}\mathbf{H}^{H}$. Next, it can be shown that $(\mathbf{H}_s\mathbf{H}_s^{H})^{-1} = \mathbf{P}(\mathbf{H}\mathbf{H}^{H})^{-1}\mathbf{P}^{H}$, which means that the corresponding rows and columns of $(\mathbf{H}\mathbf{H}^{H})^{-1}$ are switched when a pair of rows in $\mathbf{H}$ is switched. For TZFBF user selection, the row to be removed in $\mathbf{H}$ is switched with the last row to form $\mathbf{H}_s$ and the desired $\mathbf{A}^{-1}$ is then computed via (B1).