Downlink Adaptive Resource Allocation for a Multi-user MIMO OFDM System with and without Fixed Relays

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technische Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
M.Sc. Ying Zhang
geboren am 09.02.1980 in Suzhou, V. R. China

Preface

The available work is developed from December 2004 to June 2008 in line with the freelance research work in the Radio System Technology department of Nokia Siemens Networks GmbH & Co. KG in Munich, Germany. The investigation has been conducted under the supervision of Prof. Dr.-Ing. Anja Klein, the director of Communications Engineering Institute at the Technical University of Darmstadt, and the supervision of Dr.-Ing. Elena Costa from Nokia Siemens Networks.

There are many people who have helped me along the duration of this thesis. Without their support this work would have never been accomplished.

My special appreciation is given to Prof. Dr.-Ing A. Klein for encouraging and supervising my work, and for valuable guidance of the thesis.

I would like to thank Dr.-Ing E. Schulz and Dr.-Ing M. Lott for the nice offer to let me be part of the team and for the help in the application of the necessary financial support.

I am deeply indebted to my supervisor Dr.-Ing. E. Costa, not only for fruitful technical discussions, for practical suggestions on presentation and writing, but also for all kinds of support and encouragement in the work and in the life. I can still vividly remember the first time to meet Dr.-Ing. E. Costa in 2004 when she showed me the problem of adaptive resource allocation, which became my thesis topic later. I have learned so much from her since then.

I would like to also express my gratitude to all friends and colleagues in the Radio System Technology department of Nokia Siemens Networks for the interesting and valuable discussions as well as for those happy days we spent together. Especially, I want to thank Mr. R. Halfmann for his magic tips which makes my computer-based work so effective.

Last but not least, I am profoundly grateful to my family for all the unconditional support and love in my life.

Munich, 20. Juni 2008,   Ying Zhang

Kurzfassung

Es hat sich gezeigt, dass adaptive Ressourcenzuweisung eine deutlich bessere Performanz als feste Ressourcenzuweisung erreichen kann, wenn sie an unterschiedliche Kanaldämpfung, Interferenzszenario und Verkehrsbelastung angepasst wird.

Adaptive Ressourcenzuweisung der Abwärtsstrecken in einem Multi-Nutzer MIMO-OFDM-System ist wegen der zusätzlichen Dimensionen der Ressourcen schwierig. Mehrere Nutzer können gleichzeitig über unterschiedliche Subträger oder über orthogonale Beams übertragen, d. h. getrennt im Frequenz- oder räumlichen Bereich. Die vorgeschlagene Ressourcenzuweisungsmethode kombiniert die Ressourcenzuweisung im Frequenz- und räumlichen Bereich, um danach in dieser Kombination zu optimieren. Zwei Optimierungskriterien, insbesondere Minimierung der Sendeleistung und Maximierung der Datenrate, werden in dieser Dissertation untersucht. Beim ersten Kriterium erreicht die kombinierte Ressourcenzuweisungsmethode eine nahezu optimale Lösung, gleichzeitig ist die Komplexität gering. Beim zweiten Kriterium werden mehrere Varianten der kombinierten Ressourcenzuweisungsmethode vorgeschlagen, abhängig von Strategien der Nutzer-Fairness oder Leistungsbedingungen.

Im Gegensatz zu fester Ressourcenzuweisung, benötigt adaptive Ressourcenzuweisung zusätzliche Signalisierung sowohl zum Erwerb der Kanalkenntnis sowie der Übertragungsleistungen als auch zur Aussendung der Zuweisungsergebnisse. Der Overhead der Signalisierung schmälert den Gewinn der adaptiven Ressourcenzuweisung, deshalb muss ein guter Kompromiss zwischen Overhead und Gewinn geschlossen werden.

- Wenn ein Chunk, ein Block nebeneinander liegender Subträger und OFDM-Symbole, als die kleinste Ressourceneinheit betrachtet wird, kann die Signalisierung zur Aussendung der Zuweisungsergebnisse um den Faktor der Chunk-Größe reduziert werden. Um die optimale Chunk-Größe zu finden, wird der Verlust der adaptiven Ressourcenzuweisung als Funktion der Chunk-Größe analytisch beschrieben.

- Beim zeitvarianten Kanal sollte die Kanalkenntnis der Nutzer regelmäßig aktualisiert werden. Ein größeres Intervall führt zu geringerem Overhead bei gleichzeitig schlechterer Performanz aufgrund veralteter Kanalkenntnis. In dieser Dissertation wird der Performanzverlust zunächst anhand einer semianalytischen Methode als Funktion des Aktualisierungsintervalls abgeleitet, womit anschließend das optimale Intervall berechnet wird.

- Zero-Forcing Strahlformung ermöglicht die gleichzeitige Übertragung mehrerer Nutzer über orthogonale Beams unter der Voraussetzung, dass die Kanalmatrizen der Nutzer senderseitig bekannt sind, was zu einem hohen Overhead führt, insbesondere bei sehr hohen Geschwindigkeiten. Erweiterte Eigenstrahlformung basierend auf Kanalkorrelationsmatrizen und fester Strahlformung (engl. Grid-of-beam) sind zwei alternative Techniken, die Räumliche Division Multipler Zugang (engl. Spatial Division Multiple Access, SDMA) ermöglichen. Unter der Annahme, dass instantane Kanalqualität wie z. B. die SINRs zusätzlich vorhanden ist, wird adaptive Ressourcenzuweisung basierend auf diesen zwei SDMA Techniken analysiert.

- Um den Radiozugangspunkt (engl. Access point, AP) über die Übertragungsleistungen der Rückwärtsstrecke zu informieren, senden die Nutzer normalerweise die Bandbreiten-Anfrage über einen Random-Access-Kanal. In dieser Dissertation wird die Random-Access-Performanz analytisch vorgestellt. Anschließend wird ein neuartiger Gruppierungsmechanismus vorgeschlagen, der die Ressourcen beim Random-Access effizienter nutzen kann.

Aus der Literatur ist bekannt, dass die Abdeckung einer Basisstation (BS) durch den Einsatz fester Relay-Knoten zwischen ihr und dem Nutzer vergrößert oder die Kapazität an der Zellengrenze verbessert werden kann. Adaptive Ressourcenzuweisung ist in solchen relay-unterstützten Zellen wegen der Interferenz zwischen mehreren APs komplizierter, einschließlich BS und RNs. Eine vollständige zentrale Methode ist nicht realisierbar aufgrund des hohen Rechnungsaufwands und des massiven Signalisierungsoverheads, der beim Austausch der Kanal- und Interferenz-Kenntnis unter allen APs entstanden ist. Deshalb wird eine zweistufige Methode vorgeschlagen, die viel weniger Signalisierung benötigt. Über einen langen Zeitraum gruppiert jeder AP dynamisch die Nutzer zu so genannten logischen Beams, innerhalb derer die Nutzer eine hohe gegenseitige räumliche Korrelation haben, und die BS weist die Ressourcen zu den logischen Beams auf solche Weise zu, sodass ein Gewinn durch gegenseitige Interferenzdiversität erreicht wird und der Ende-zu-Ende-Durchsatz maximiert werden kann. Über einen kurzfristigen Zeitraum weist jeder AP die Ressourcen den Nutzern zu, welche zu demselben logischen Beam gehören. Dabei kann ein Gewinn durch Multi-Nutzer-Diversität erreicht werden.

Abstract

A downlink (DL) system comprises a centralized base station (BS) communicating to a number of users physically scattered around. The purpose of resource allocation at the BS is to intelligently allocate the limited radio resources, e.g. transmit power, time slots and frequency bandwidth, among users to meet their data rate requirements. Adaptive resource allocation has been shown to achieve significantly higher performance than fixed resource allocation by adapting resource allocation with respect to varying channel fading, interference scenario and traffic load.

This thesis deals with the problem of DL adaptive resource allocation in a multi-user MIMO OFDM system. In a multi-user MIMO OFDM system, multiple users can simultaneously transmit data and be separated in frequency domain or in spatial domain, i.e. via different sub-carriers or via orthogonal beams, respectively. Thus, adaptive resource allocation in such a system is highly challenging because of the high degree of freedom for resources.

Firstly, an approach of jointly optimizing the resource allocation in frequency and spatial domains is proposed in this thesis. Two types of optimization problems, namely power minimization and rate maximization, are addressed. For the power minimization case, the joint approach is shown to achieve a near-optimal solution with low complexity. For the rate maximization case, several variants of the joint approach are proposed in order to take into account different user fairness strategies and different power constraints.

Compared to fixed resource allocation, adaptive resource allocation needs signaling for acquisition of channel and traffic knowledge as well as for delivery of allocation results, which causes additional overhead, thus mitigating the adaptation gain. Hence, the reduction of the signaling overhead is as important as the increase of the adaptation gain in order to maximize the system performance. The following investigations targeting at reduction of signaling overhead are considered in this thesis:

- By defining a chunk as a block of adjacent sub-carriers and OFDM symbols and letting it be the basic resource unit, the signaling for delivery of allocation results from a base station to all users it served can be reduced by a factor of the chunk dimension, but the adaptation gain also decreases with increasing chunk dimension. In order to find the optimal chunk dimension, the adaptation gain as a function of the chunk dimension is analytically derived.

- To cope with the time-variant property of the channel fading, users' channel knowledge needs to be periodically updated. To find the optimal update interval, both the overhead reduction and the performance loss due to outdated channel knowledge should be evaluated. Therefore, the performance is firstly derived as a function of the update interval by means of a semi-analytical method and then the optimal update interval as a function of the velocity can be analytically derived accordingly.

- Zero-forcing beamforming enables multiple users to transmit simultaneously over orthogonal beams, but requires the complete channel matrices, which leads to high signaling overhead especially at very high velocities. Generalized eigenbeamforming and fixed grid-of-beam beamforming are two alternative techniques to enable spatial division multiple access (SDMA) but require only partial channel knowledge and less signaling compared to zero-forming beamforming. Under the assumption that instantaneous channel quality indication is additionally available, adaptive resource allocation based on these two beamforming techniques is investigated and their performance is assessed.

- Random access is commonly used by users to transmit bandwidth requests which inform the BS about the traffic load of the uplink transmission. Typically, slotted ALOHA protocol is used in conjunction with truncated binary exponential back-off algorithm for random access. Its performance in the considered system is firstly analytically analyzed, and then a novel grouping mechanism, yielding a more efficient usage of the resources for random access, is proposed.

Finally, since fixed relay nodes (RNs) has been shown to extend the coverage of the BS or enhance the cell-edge capacity by forwarding data between BS and users, DL adaptive resource allocation in a relay-enhanced cell (REC) is addressed in this thesis. Different from the BS, the RN has no wired connection to the core network, but it also provides radio access to the users, and so both BS and RN are called access points (APs). It is expected that the system performance in such a REC can be enhanced by letting the BS adapt the resource allocation with respect to the interference among the multiple APs in the REC. Since a complete centralized resource allocation approach performed at the BS is not applicable in practical systems due to the extremely high computational complexity and huge signaling for the exchange of channel and interference information among APs, a two-level approach which requires much less signaling is proposed in this thesis. On a long-term basis, e.g. for each super-frame, each AP dynamically groups users with high spatial correlation into so-called logical beams, and then the BS allocates resources to logical beams in such a way that end-to-end throughput is maximized and mutual interference diversity is exploited by allowing logical beams with sufficiently low mutual interference to share the same time-frequency resource. On a short-term basis, e.g. for each frame, each AP exploits multi-user diversity by adaptively selecting one user from each logical beam for each time-frequency resource assigned to that logical beam.

Contents

1 Introduction

1.1 Downlink Adaptive Resource Allocation in a B3G System

Beyond third generation (B3G) mobile communication systems are expected to provide a variety of services such as voice, image and data transmission with different QoS and rate requirements for "anytime-anywhere". As defined by the ITU-R, a peak rate of 100 Mbps for mobile access when users move at high speeds relative to the BS, and 1 Gbps for nomadic access when users are in relatively fixed positions are assumed for B3G systems [ITU03]. The technologies which are being considered as B3G are WiMax, WiBro, iBurst, 3GPP Long Term Evolution and 3GPP2 Ultra Mobile Broadband.

The transmission data rate envisioned for beyond third generation (B3G) mobile communication systems is much higher than that provisioned by current 3G (third generation) systems. With increasing data rate, the symbol duration becomes shorter and shorter. If the symbol duration becomes smaller than the delay spread of the multi-path channel, or equivalently the channel exhibits selectivity in frequency domain, the whole system will heavily suffer from inter-symbol interference (ISI) [Pro01]. An approach to prevent ISI is parallel data transmission, known as multi-carrier (MC) modulation [Sal67, WE71]. It converts a high-rate data stream into a number of low-rate sub-streams that are transmitted over a number of sub-carriers simultaneously. The resulting MC symbol duration, defined as the symbol duration of each sub-carrier, linearly increases with the number of sub-carriers. The increased symbol duration reduces the impact of multi-path time dispersion. Orthogonal frequency division multiplexing (OFDM) transmission technique has become quite popular in the last decades. OFDM is a low complexity technique to bandwidth efficiently modulate parallel data streams to multiple carriers, because the modulation of parallel data streams to a number of sub-carriers in OFDM is performed by inverse discrete Fourier transform (IDFT) which can be implemented very efficiently by inverse Fast Fourier Transform (IFFT), and the sub-carriers in OFDM are orthogonal with a sub-carrier spacing equal to the Nyquist bandwidth [NP00].

In addition to OFDM, multi-input multiple-output (MIMO) transmission technique has also been intensively investigated in the last decades. Multiple antennas installed at transmitter and receiver can improve the transmission reliability and/or the system throughput

by utilizing the spatial dimension [GSsS$^+$03, Fos96].

The high transmission data rate envisioned for B3G mobile communication systems also creates serious power concerns as it is well known that for a given transmit power level, the symbol energy decreases linearly with increasing transmission rate. Moreover, even if the spectral efficiency can be significantly improved by advance technologies such as adaptive transmission, spatial processing and dynamic spectral sharing [PSSH04, CS00], B3G systems still require wider bandwidths than existing systems to meet the more demanding data rate requirements. Since wider bandwidths are in general only available at higher carrier frequencies, the spectrum that will be released for B3G systems will be almost certainly located well above the 2 GHz band used by 3G systems. The propagation in such a high operation frequency band experiences higher attenuation and is mostly dominated by non-line-of-sight (NLOS) conditions.

In cellular systems, each base station (BS) is wirely connected to the core network and provides radio access to the user terminals (UTs) over a certain range. In B3G systems, the reduced symbol energy and the increased attenuation result in lower BS coverage range, or equivalently, less capacity at the cell edge, as well as wider shadowed area. Thus, complete coverage of e.g. urban areas using conventional cellular infrastructures is expected to be very costly owing to the high number of BSs and fixed core network connections required. For this reason, the introduction of fixed relay nodes (RNs) has been widely accepted to cost efficiently extend the coverage of the BS and/or enhance the cell capacity, especially at cell edges [PWSea04]. The cell configured with fixed RNs is referred to as relay-enhanced cell (REC). In the REC, besides being directly connected with the BS, UTs can alternatively exchange data with the BS via certain RN through multi-hop communications.

In summary, it has been widely agreed [BaRT02, KJC$^+$03] that key technologies to enable such a high data rate at low deployment cost for B3G systems include

- **Orthogonal frequency division multiplexing (OFDM)** to efficiently exploit the channel frequency selectivity by converting a high-rate data stream into a number of low-rate streams [NP00],

- **Multi-input multiple-output (MIMO)** to attain high spectral efficiency and/or high transmission reliability by fully utilizing the spatial domain through multiple antennas installed at transmitter and receiver [GSsS$^+$03],

- **Relaying** to extend the BS coverage and/or enhance the cell-edge capacity by means of multi-hop and/or cooperative transmission via fixed RNs [PWSea04].

Therefore, a B3G system can be characterized as a multi-user MIMO OFDM system with and without fixed RNs.

In a downlink (DL) wireless system, a centralized BS communicates to a number of users distributed around it. The purpose of the resource allocation is to allocate the limited

resources, e.g. total transmit power, available time slot and frequency bandwidth, to users to meet the users' quality of service (QoS) requirements, e.g. data rate and delay.

By taking the variance of channel fading, interference scenario and traffic load into account, adaptive resource allocation yields higher system performance than fixed resource allocation, and is becoming more important in wireless communication systems while the user data rate requirements keep increasing [ZKA01].

This thesis is focused on DL adaptive resource allocation in a B3G system, i.e. a multi-user MIMO OFDM system with and without fixed RNs.

Since OFDM enables parallel data streams to be transmitted over multiple sub-carriers, multiple users can share the OFDM symbol by transmitting on different sub-carriers in a frequency division multiple access (FDMA) fashion [BaRT99], referred to as orthogonal frequency division multiple access (OFDMA) hereafter, and so an OFDM system supporting OFDMA is called an OFDMA system. Due to the independent channel fading experienced by multiple users, the system performance can be enhanced by always assigning the sub-carriers to the user with the highest channel gain [RC00] as well as performing bit and power loading across all sub-carriers [KR00].

When the BS is equipped with multiple antennas, users can be separated in spatial domain, leading to spatial division multiple access (SDMA). Simultaneous data streams of different users are transmitted on orthogonal or semi-orthogonal beams so as to eliminate or reduce the inter-user interference at the receivers [FN94].

Other multiple access schemes such as time division multiple access (TDMA) and code division multiple access (CDMA), which separate users in time and code domain, respectively, can be used in combination with OFDMA and SDMA [Rap02]. Thus, given the freedom of separating users in the time, frequency, code, or spatial domain, adaptive resource allocation becomes more flexible yet more challenging, because a multi-dimensional optimization is desired.

Compared to fixed resource allocation, additional signaling is required by adaptive resource allocation to measure and report channel conditions, to broadcast the allocation results, etc.. With increasing user velocities, the channel varies more rapidly and the adaptive resource allocation has to be performed more often. As a consequence, the overhead caused by the additional signaling becomes high, which mitigates the performance gain achieved through adaptive resource allocation. Hence, the optimization of adaptive resource allocation relies on both adaptation gain enhancement and signaling overhead reduction.

By introducing fixed RNs, the data transmission between BS and users can also be relayed by RNs in addition to a direct connection [PWSea04]. Different from the BS, the RN has no wired connection to the core network, but it also provides radio access to the users, and so both BS and RN are called access points (APs) in this thesis. Thus, even in the single cell scenario, the BS is not the only transmitter of DL transmission anymore, and so

3

the transmissions of all APs, including both BS and RNs, may interfere with each other if the used resources are not orthogonal. As the central control unit of the cell, the BS should handle the interference among APs belonging to the cell in an intelligent way.

1.2 State-of-the-art and Open Questions

In this section, the state-of-the-art of DL resource allocation in a MIMO OFDMA system is summarized and the open questions are raised accordingly.

Adaptive resource allocation at a single AP has been widely investigated for various multi-user systems under the assumption that ideal channel knowledge is available at the transmitter. Main contributions are summarized in Table 1.2 and briefly reviewed in the following.

Table 1.1: Summary of previous contributions to DL adaptive resource allocation at a single AP.

	References	Remarks
OFDMA systems	[WCLM99, KLL03]	power minimization
	[LL04, JL03]	rate maximization
	[VTL02, BBG$^+$00]	rate maximization, PF, TDMA
	[KH05, KPD06, ALS$^+$03]	rate maximization, PF, OFDMA
	[Bon04]	rate maximization, score-based fairness
	[RC00, SAE05]	rate maximization, MMF
	[SW$^+$07, COE05]	rate maximization, generalized PF
narrow-band MIMO systems	[Zha02, FGH05]	user partitioning
	[SS04, MK07a]	user selection
	[DS05, FN96, MK07b, YG06]	user selection, greedy user insertion
	[RPS$^+$06]	user grouping, greedy bit insertion
MIMO OFDMA systems	[MA06, LN04]	no SDMA
	[PJKL04]	power minimization, optimal solution
	[Wil06, ZL05, MK07b]	rate maximization, decoupled optimization in frequency and spatial domains

Adaptive resource allocation in multi-user OFDMA systems, known as adaptive OFDMA, assigns the sub-carriers to the users and performs bit and power loading across the sub-carriers. It is usually formulated as an optimization problem, e.g. minimizing the total transmit power under minimum user data rate constraints [WCLM99, KLL03] or maximizing the total throughput under a transmit power constraint [JL03, LL04, SAE05].

The formulated optimization problems are often very difficult to solve and sub-optimal algorithms with low computational complexity have been proposed.

In [WCLM99], a Lagrangian relaxation approach has been used to solve the power minimization problem, in which the binary variable indicating the sub-carrier assignment is allowed to take any value between zero and one during the optimization, and the solution is then appropriately discretized. As the computational complexity of the approach in [WCLM99] is still quite high, a computationally efficient method is proposed in [KLL03]. It reduces the complexity by dividing the problem into two stages, the determination of the number of sub-carriers assigned to each user and the assignment of best sub-carriers to each user [KLL03].

For the rate maximization problem, it has been proven that the total rate is maximized when each sub-carrier is assigned to the user with the highest channel gain and the transmit power is then distributed according to the water-filling algorithm [LL04, JL03]. However, such approach does not ensure any fairness among users, because it always selects users supporting the highest data rate and may leave out users with bad channel conditions, typically located at the cell edge.

One of the most popular strategies to balance the exploitation of multi-user diversity and user fairness is proportional fair (PF) strategy, by which the user with the highest data rate relative to its average achieved data rate is selected [VTL02]. The PF strategy has firstly been proposed for TDMA systems [BBG$^+$00] and has then been applied to OFDMA systems as well [KH05, KPD06, ALS$^+$03]. When users experience asymmetric channel fading, the PF strategy has been observed to be biased, i.e. the vector consisting of achieved users' data rates is not proportional fair [Hot01]. To avoid such a bias, a so-called score-based strategy has been proposed in [Bon04].

Alternatively to the PF strategy, the max-min fairness (MMF) strategy is studied in [RC00], where all users are assured to achieve a similar data rate by maximizing the lowest user's data rate. Considering that different users may require different data rates due to different services, it is proposed in [SAE05] to maximize the total throughput while exactly maintaining proportional user data rates, i.e. the lowest ratio between the achieved and the required data rate is maximized.

Moreover, in [SW$^+$07, COE05], a so-called generalized proportional fair strategy has been presented, in which the user fairness level is tunable by adjusting the parameters used in the objective function. It can be regarded as a generalized formulation for rate maximization problems with no fairness, with proportional fairness and with max-min fairness.

As multiple antennas at transmitters and receivers have been illustrated to be able to enhance the system performance, adaptive resource allocation in MIMO systems has gained a lot of interest. Under the constraint that only one user is served on each time-frequency resource, i.e. no SDMA is allowed, the optimization problem of adaptive resource allocation in MIMO systems is similar to that in OFDM systems and differs only in the fact that the

single-user spatial processing is optimized in addition to adaptive OFDMA [MA06, LN04].

When SDMA is allowed, a group of users can be simultaneously served on the same time-frequency resource and be separated in spatial domain. In this thesis, such a group is termed an SDMA group. Since the spectral efficiency achieved by an SDMA group depends on the spatial separability among users in it, the system performance can be optimized by placing users with low spatial correlations in an SDMA group, referred to as adaptive SDMA hereafter.

Adaptive SDMA can be realized by a user partitioning approach, in which a partition of a given set of users is firstly performed such that each subset corresponds to an SDMA group and another orthogonal multiple access scheme such as TDMA is applied among different subsets [Zha02, FGH05]. A graph-theoretical solution for user partitioning is presented in [Zha02], but its computational complexity is found to be NP-hard. In [FGH05], a tree-based sub-optimal algorithm for user partitioning is proposed, which reaches a close to optimum solution.

Alternatively to the user partitioning approach, a user selection approach has also been widely discussed [SS04, MK07a, YG06, DS05], in which an SDMA group is constructed for each given time-frequency resource by selecting proper users from all users. The problem of finding the best SDMA group that maximizes the system capacity is recognized as a non-deterministic polynomial time hard problem. Its optimum solution can be only found through exhaustive search [SS04]. However, an exhaustive search has exponential complexity and is prohibitive in a practical system due to unaffordable computational cost. In [MK07a], by allowing continuous values for the binary variables, which indicate whether the users are selected in the SDMA group or not, the problem of finding best SDMA group is formulated as a convex quadratic optimization problem and is then efficiently solved by convex optimization methods. Although a convex optimization problem can be solved with non-exponential complexity, it still might require a considerable iterations of a convex optimization algorithm. Thus, an algorithm, named greedy user insertion, has been proposed for user selection [DS05, FN96, YG06]. A greedy user insertion algorithm starts from an empty SDMA group for a given resource and iteratively inserts a user in it till the performance can no longer be increased by adding one more user. The user to be inserted in each iteration is selected according to certain criterion, e.g. maximization of capacity increase. A number of variants of the greedy user insertion algorithm have been proposed for SDMA grouping, which differ from each other on the selection criterion, e.g. group capacity [DS05] and spatial correlation [FN96]. In [MK07b], a so-called regularized correlation-based algorithm is proposed, in which the user selection criterion takes both spatial correlation and channel gain into account. In [YG06], under the assumption of zero-forcing beamforming (ZFBF) and fixed power allocation for each user in the SDMA group, the user with the highest rate weighted by the average data rate is selected and inserted in the SDMA group such that the achieved data rates are proportional fair.

Similar to the greedy user selection algorithm, a greedy bit insertion algorithm has been

proposed in [RPS⁺06]. Starting from assigning to all users zero rate, the users' data rates are iteratively increased till the performance cannot be increased, and in each iteration it is the user with the highest power efficiency that increases its data rate [RPS⁺06].

The aforementioned approaches for adaptive resource allocation either deal with the optimization in frequency domain by, e.g., adaptive sub-carrier assignment in wide-band OFDM systems, or with the optimization in spatial domain by, e.g., adaptive user selection in narrow-band MIMO systems. However, the B3G system is characterized as a multi-user MIMO OFDMA system with flexible multiple access schemes such as OFDMA and SDMA, cf. Section 1.1. Hence, the adaptive resource allocation in such a system is expected to be optimized by considering both frequency and spatial domains. In [PJKL04], the optimal solution for adaptive resource allocation is obtained by converting the problem into an integer linear problem but the computational complexity is too high to be used in practical systems.

Some low complexity sub-optimal approaches have been proposed in the literature. In [Wil06], the allocation of multiple time-frequency resources available in the current frame and over the whole bandwidth is sequentially performed, and for each time-frequency resource, the resource allocation is optimized by means of approaches proposed for narrow-band MIMO systems. Alternatively in [ZL05], it is proposed to divide the problem into two sub-problems: SDMA groups are built by user partitioning in such a way that the correlation between any pair of the users from different groups is lower than a given threshold, and then for each group adaptive OFDMA is carried out independently among users in the same group. This approach, however, is sometimes difficult to apply, because in a rich-scattered propagation environment, the spatial separability, e.g. measured by spatial correlation, is frequency/time-dependent and the best SDMA group for one resource might not be optimum for another [KRT03]. In [MK07b], it is proposed to build a candidate SDMA group for each user on each resource, and then selecting one SDMA group for each resource by means of adaptive OFDMA.

Because all the existing sub-optimal approaches decouple the optimization problem in frequency and spatial domains, the following question needs to be answered:

1. *How to make joint optimization of resource allocation in frequency and spatial domains with low computational complexity, and how much can the joint optimization gain compared to the existing approaches?*

The investigations discussed so far about adaptive resource allocation focus on the optimization of the resource allocation in terms of power minimization or rate maximization, under the assumption that all required information is available at the transmitter, i.e. the AP. The required information includes channel knowledge and users' data rate requests. Moreover, the allocation results also need to be delivered from the AP to individual users over DL broadcasting or dedicated control channel. All the corresponding signaling shares

the resources with the data transmission, and so the reduction of the signaling overhead is as important as the optimization of the data transmission. The signaling required by adaptive resource allocation can be classified as

- DL signaling for the delivery of allocation results from APs to users,

- Signaling for the acquisition of channel knowledge at AP, e.g. uplink (UL) signaling for channel feedback or pilots for channel measurement.

- UL signaling for the transmission of users' data rate requests.

Main contributions to the topic of signaling reduction are presented in Table 1.2.

Table 1.2: Summary of previous contributions to the reduction of the additional signaling caused by adaptive resource allocation compared to fixed resource allocation.

Signaling type		References	Remarks
Delivery of allocation results		[DTSA05]	grouping of sub-carriers
		[IST05a]	grouping of sub-carriers and OFDM symbols
Acquisition of channel knowledge	Pilots	[MI02]	uplink pilot in TDD mode
	Feedback	[IST05b]	CQI compression
		[NLTW98, DJT03, MSEA03]	CSI quantization
		[CH05]	CSI interpolation
		[SH05, TH05]	feedback over random access
Transmission of users' data rate request		[IEE04]	polling and random access

After the AP completes the adaptive resource allocation, it has to inform all users about the allocation results, namely indicating for each sub-carrier which user is served and which transmission mode, i.e. modulation and coding scheme (MCS), is selected. By observing that one user undergoes similar channel fading on adjacent sub-carriers, a solution to reduce the amount of signaling as well as the computational complexity of the adaptive allocation is to group adjacent sub-carriers together [DTSA05]. Furthermore, in [IST05a] it is proposed to use a chunk, consisting of adjacent sub-carriers and OFDM symbols, as the basic resource unit in adaptive resource allocation. Since the gain of the adaptive resource allocation decreases with the increasing size of the resource unit, such grouping requires a trade-off between the reduction of the signaling reduction and the loss of adaptation gain [DTSA05]. This trade-off has not yet been systematically investigated in literature, thus, the existing open question can be formulated as:

2. *How does the performance of chunk-wise adaptive resource allocation decrease with the increasing chunk dimension and how to determine the optimum chunk dimension?*

In time division duplex (TDD) mode, due to channel reciprocity, the DL channel is identical to the UL channel, and so the AP can directly measure the DL channel based on additional pilots inserted by each user in uplink [MI02]. Otherwise, without exploiting the channel reciprocity in TDD mode or in FDD mode, all users should first measure the channel and then feed back the channel measurement to the AP. Channel feedback can be classified as two types, namely channel state information (CSI) such as transmit beamforming vectors or channel matrix, and channel quality indication (CQI) such as signal-to-interference and noise ratio (SINR) or channel norm.

Many approaches to reduce the channel feedback have been proposed in the literature. For CQI feedback composed of real values, a significant reduction can be attained with different kinds of lossless or lossy compression methods by exploiting the correlation among adjacent sub-carriers [IST05b]. For CSI feedback composed of complex vectors or metrics, instead of sending back each entry of the vectors, the vectors can be quantized using a code book designed for narrow-band MIMO channels, e.g. [NLTW98, DJT03, MSEA03]. Even with vector quantization, the amount of feedback still grows in proportion to the number of sub-carriers. In [CH05], a feedback scheme that combines vector quantization and vector interpolation is proposed. Further on, the spherical interpolator developed in [CH05] exploits parameters for phase rotation in order to satisfy the phase invariance [DJT03] and unit norm properties of the transmit beamforming vectors. As the amount of signaling overhead required still increases with the number of users, in order to scale the feedback for a large number of users, the feedback is also proposed to be transmitted by means of random access [SH05, TH05].

Due to user mobility, the channel is variant over time, and so the acquisition of channel knowledge has to be carried out periodically. A short update interval causes too high overhead that cannot be compensated by the gain from adaptive resource allocation, whilst with a too long update interval the CSI does not match with the current channel status and the adaptation gain degrades. Hence, there exists an optimum update interval that maximizes the system performance, and so it is obvious to raise the following question:

3. *What is the optimum interval for the update of channel knowledge in time-variant channels and how is it related to the users' velocity?*

SDMA schemes based on zero-forcing beamforming [DS05] and block diagonalization [SSH04] require the instantaneous CSI at the transmitter. However, it is expected that the feedback of instantaneous CSI is not feasible when users move at very high velocities. Indeed, there are other SDMA schemes that only require long-term CSI, e.g. the channel correlation matrix. For instance, eigenbeamforming maximizes the useful signal

power by transmitting towards the main signal direction indicated by the dominant eigen-vector of the channel correlation matrix [FN95, SB04a]; generalized eigenbeamforming tries to maximizes the SINR by balancing the reduction of the useful signal reduction and the suppression of the interference [Zet95]. In literature, the adaptive resource allocation is generally formulated as optimization of the beamforming vectors and power allocation subject to users' individual target SINRs [SB04a]. Since the CQI feedback causes much less overhead compared to the CSI feedback, it is reasonable to assume that the instantaneous CQI feedback is still affordable at high velocities to certain extent. Consequently, an open question is:

4. *How to efficiently combine long-term CSI and instantaneous CQI in adaptive resource allocation and what is the performance loss compared to that achieved based on instantaneous CSI?*

In addition to the channel knowledge, user's QoS parameters such as the data rate request are also required by adaptive resource allocation. As the user data rate request can be variable over time due to bursty traffic in packet transmission systems, the APs should be informed about the change of rate request by bandwidth request (BW-REQ) transmission, as described in the IEEE 802.16 standard [IEE04]. Two main schemes, namely polling and random access, can be used in BW-REQ transmission [Tan96]. For random access, the transmission is contention-based and so a collision avoidance algorithm is usually defined [Tan96]. For example, the IEEE 802.16 standard uses truncated binary exponential back-off (TBEB) algorithm for collision resolution [IEE04]. In order to optimize the transmission of bandwidth request, e.g., to minimize the resources required by the transmission of bandwidth request, the behavior of the random access should be firstly analyzed. In other words, open question is:

5. *What is the performance of random access in frame-based systems and how to minimize the resources required by the transmission of BW-REQs.*

In the previous discussions, only one single AP is considered. Indeed, in systems comprising multiple APs, e.g. a cellular system or a REC, co-channel interference caused by frequency reuse among multiple APs is the most restraining factor on the system capacity [Lee93]. The allocation of available frequency bandwidth among multiple APs is generally called channel assignment in literature [Lee93]. Main contributions are summarized in Table 1.3 and briefly reviewed in the following.

The channel assignment in early FDMA/OFDMA cellular systems operates on a long-term basis and frequencies are assigned to different APs on a more or less permanent basis [Lee93]. Inhomogeneity in the traffic load can be taken care of by adapting the number of channels assigned to each AP to the expected traffic carried by that AP. A comprehensive survey of different channel assignment schemes is provided in [KN96]. Two kinds

Table 1.3: Summary of previous contributions to DL channel assignment at multiple APs.

	References	Remarks
OFDMA systems	[Lee93]	FCA
	[KN96]	survey on centralized DCA
	[CS00]	sequential DCA
	[LL03]	semi-distributed DCA
MIMO OFDMA systems	[VTZZ06]	distributed DCA

of approaches for channel assignment have been identified, namely fixed channel assignment (FCA) and dynamic channel assignment (DCA) [KN96]. Due to the temporal and spatial variations of traffic in cellular systems, FCA cannot achieve high spectrum efficiency [KN96]. In case of DCA, all channels are kept in a centralized pool, and channels are dynamically assigned to cells or returned in accordance to the increase or the decrease of the traffic load, respectively [KN96].

The performance of DCA schemes is critically dependent on the rate at which the assignment or re-assignment occurs. To fully utilize the potential of DCA gain, channel reassignments must take place at high speed to avoid rapidly changing signal- and interference-levels in a mobile system. However, channel variations, especially those caused by fast fading, are generally very fast. As a result, centralized DCA schemes adapted to such a fast channel variations cause very high computational complexity as well as huge signaling burden for signal and interference measurements, which are usually infeasible in practical systems. On the other hand, completely distributed DCA schemes require much less signaling compared to centralized ones, but are problematic in practice due to collisions of channel assignment, i.e. the possibility for adjacent APs to independently select the same channel, thus causing interference when transmissions occur. Collisions of channel assignment can be avoided by letting adjacent APs sequentially perform DCA algorithm [CS00], but the resulting cycle of DCA might be too long to adapt to the rapid change of fast fading, which limits the DCA gain.

In [LL03], a semi-distributed DCA scheme, which splits the resource allocation between radio network controller (RNC) and BSs, is proposed for multi-cell OFDMA systems. The RNC makes the decision which resource unit, e.g. chunk, is used by which BS as well as the transmit power on a long-term basis, e.g. at super-frame level. The BSs then make the decision which resource unit is assigned to which user on a short-term basis, e.g. at frame level. As RNC only requires the information on channel slow fading and makes the decision at a super-frame level, the rate of information exchange between RNC and BSs is significantly reduced compared to centralized DCA. Moreover, once the RNC has made the decisions which resource is used by which BS with a given transmit power, the co-channel

interference from each BS to users served by other BSs is not dependent on the resource allocation independently carried out by each BS. Thus, the BS can accurately predict the instantaneous SINRs without knowing the actual resource allocation made by other BSs.

However, when BSs are equipped with multiple antennas, the co-channel interference changes with the used transmit beamforming vectors of the interfering BSs even under a stable channel and fixed transmit power. As the transmit beamforming vector chosen for different users are generally different, the BS cannot predict the instantaneous SINR without knowing the resource allocation of other BSs: some users can be "hit" by the beams of the adjacent BSs, whereas other users can be in a very favorable situation, depending on their channel conditions and the directions of the interferers [VTZZ06]. The uncertainty of the co-channel interference makes the adaptive resource allocation problem in MIMO OFDMA systems with multiple APs more difficult compared to systems without multiple antennas at the transmitter. In [VTZZ06], a distributed DCA approach is proposed by approximating the co-channel interference by the value in the worst case scenario, i.e. the interference when being "hitted" by the beams of the adjacent BSs. However, the approach following worst case analysis won't lead to the optimal solution. So far solutions based on an accurate approximation of the co-channel interference for DCA in multi-user MIMO OFDMA systems have not been presented in literature. The open question is thus formulated as:

6. *How to efficiently benefit from the mutual interference diversity among multiple APs in a REC with affordable computational complexity and signaling overhead, when all APs, including both BS and RNs, are equipped with multiple antennas ?*

1.3 Goals of the Thesis

Concerning the open questions raised in Section 1.2 for the DL adaptive resource allocation in multi-user MIMO OFDMA system, the following goals are pursued in this thesis:

- Optimization of the adaptive resource allocation, namely joint optimization of adaptive OFDMA and adaptive SDMA, at a single AP without considering interference from other AP.

- Investigation of the additional signaling required by the adaptive resource allocation in order to reduce the overhead. Particularly, the following four aspects are addressed.

    - Optimization of the size of the basic resource unit to reduce the signaling for delivery of allocation results.

– Optimization of the update interval for channel knowledge acquisition in a time-variant channel.

– Optimization of the adaptive resource allocation when limited channel knowledge is available.

– Optimization of the random access scheme used by users to deliver information on uplink traffic load to APs.

• Optimization of the adaptive resource allocation in a relay-enhanced cell (REC), in which multiple APs interfere with each other.

1.4 Contributions and Thesis Overview

This section discusses the main contributions of the thesis and how the thesis is organized. In the following, the contents of each chapter are briefly described, along with the contributions presented by each one of them.

Chapter 2 is dedicated to describe the assumed system model, which prepares the discussion and investigations in the rest of the thesis. After introducing the channel model, OFDM is reviewed, transmission with multiple antennas is modeled and discussed, link adaptation including both power adaptation and adaptive MCS selection is outlined. Time-frequency resources are then shown to be structured as chunks, frames and super-frames. Some further assumptions are also given.

Based on the system model described in Chapter 2, all the contributions of the thesis are presented in Chapter 3, Chapter 4 and Chapter 5.

In Chapter 3, with the purpose of answering open question 1, the theme of DL joint optimization of adaptive OFDMA and adaptive SDMA based on ZFBF at a single AP is approached. Two kinds of problem, namely power minimization and rate maximization problems, are separately addressed. The contributions to power minimization problem are presented in Section 3.2, in which after reviewing the formulation of the problem and the calculation of the optimal solution, a low complexity sub-optimal algorithm, referred to as successive bit insertion (SBI), as well as some further modifications to enhance the performance, are proposed and analyzed. The contributions to rate maximization problem are presented in Section 3.3, in which the objective functions with respect to different optimization criteria, i.e. rate maximization with different user fairness strategy, are firstly derived, and then the proposed SBI algorithm are shown to take into account different optimization criteria by applying the derived objective functions. Additionally, in order to solve the rate maximization problem while assuming fixed power sharing among users served in the same resource unit, a sub-optimal algorithm, called successive user insertion (SUI), is proposed and analyzed in Section 3.3.

Chapter 4 focuses on the signaling for adaptive resource allocation. In Section 4.2, with the purpose of answering open question 2, the performance gain from adaptive resource allocation is expressed as a decreasing function of the chunk dimension by analytical derivation. By further formulating the amount of forward signaling, i.e. the signaling for delivery of the allocation results, as a linearly decreasing function of the chunk dimension, the optimal chunk dimension that provides the best trade-off between the loss of performance gain and the reduction of signaling with respect to increasing chunk dimension, can be calculated. In Section 4.3, the performance loss due to mismatched CSI is firstly expressed as a function of the interval for the CSI update based on simulative results, and then the optimum channel update interval that maximizes the effective throughput is analytically derived for arbitrary velocities, thus answering open question 3. In Section 4.4, the sub-optimal algorithm proposed in Chapter 3 for joint optimization of adaptive OFDMA and adaptive SDMA based on ZFBF is modified in order to solve the problem of joint optimization of adaptive OFDMA and adaptive SDMA based on two SDMA schemes requiring only limited channel information, i.e. generalized eigenbeamforming and fixed grid-of-beamforming (GoB), which refer to open question 4. Additionally, in case of generalized eigenbeamforming, as the receive SINR cannot be estimated by users without knowing the used transmit beamforming vectors, two methods are proposed to estimate the receive SINR at the AP based on channel correlation matrix and instantaneous CQI. The performance of the two SDMA schemes with limited channel feedback is evaluated and compared by means of numerical simulation. In Section 4.5, the performance of random access used for such as bandwidth transmission is analytically derived with respect to any given values of the parameters in the back-off algorithm, which enables the selection of proper values for those parameters, thus answering open question 5. Moreover, a grouping mechanism is proposed to provide more efficient usage of resources in the contention period .

Chapter 5 presents a novel two-step approach to optimize the adaptive resource allocation in a REC, thus answering open question 6. On a long-term basis, each AP independently constructs logical beams consisting of users with high spatial correlation, and then the BS allocates resource to logical beams to maximize the throughput by optimizing the resource sharing among logical beams with low spatial correlation. On a short-term basis, each AP independently selects the best user for each resource unit according to instantaneous SINR condition. In particular, an algorithm is proposed for APs to dynamically build logical beams, and two algorithms are proposed for the BS to allocate resource to logical beams aiming at rate maximization as well as resource balancing between the first and second hop transmissions. The second algorithm, iterative independent balancing, guarantees a complete fairness among all users in addition.

Finally, Chapter 6 summaries the main contributions of the thesis.

2 System Model

2.1 Introduction

In this chapter, the deployment scenario of the considered cellular system with fixed relays as well as the system parameters are firstly introduced in Section 2.2. The mathematical description of the radio channel and its statistical properties will follow in Section 2.3, outlining the fundamental limitations placed by the mobile radio channel on the performance of wireless communication systems. OFDM and multiple-antenna transmission techniques, as the key techniques for future wireless systems, are discussed in Section 2.4 and Section 2.5, respectively. After introducing the frame structure of the considered system in Section 2.6, chunk-wise link adaptation technique is described in Section 2.7. Finally, a list of assumptions considered for performance evaluation is given in Section 2.8. The research work presented in this thesis has been carried out in the framework of the IST project IST-2003-507581 WINNER [Win], and thus most of the system parameters and assumptions are in accordance with those defined in the WINNER project [IST06].

Throughout the thesis, signals and channel responses are represented by complex scalars, vectors and matrices. Lower case and upper case letters are used to denote vectors and matrices, respectively. To distinguish from scalars, both vectors and matrices are in bold face. Furthermore, $(\cdot)^*$ and $(\cdot)^{\mathrm{T}}$ designate the complex conjugate and the transpose, respectively. The complex conjugate transpose, also called complex Hermitian, is expressed as $(\cdot)^{\mathrm{H}}$. The matrix inverse is represented by $(\cdot)^{-1}$ and $\| \cdot \|$ denotes the Frobenius norm of vectors and matrices. The operator $[\cdot]_{i,j}$ yields the element in the i-th row and the j-th column of the matrix in bracket, and $\mathrm{diag}[\cdot]$ yields a diagonal matrix composed of elements in the bracket. The expressions $E[\cdot]$ and $\mathrm{tr}[\cdot]$ indicate the expectation and trace operation, respectively.

2.2 Cellular System with Fixed Relays

In this section, the layout and the deployment parameters of a cellular system equipped with fixed relays are presented.

According to the characteristics of the environments where B3G systems are envisaged to be operated, such as propagation conditions and user mobility, three kinds of deploy-

ment scenarios are commonly identified: urban macro-cellular, micro-cellular and in-door [IST06]. The urban macro-cellular deployment scenario is the focus of this thesis. In typical urban macro-cellular scenarios, the cell size is generally large, mobile UTs are at street level and fixed BSs are located clearly above surrounding building heights. As for propagation condition, non- or obstructed line-of-sight is a common case, since the street level is often reached by diffractions over the rooftop.

As depicted in Figure 2.1(a), for the considered urban macro-cellular deployment scenario, each BS is assumed to serve one site consisting of three sectors and the distance between two adjacent BSs, denoted with D, is called site distance. Thus, the hexagonal diameter of the cell is about $2/3D$. Moreover, the considered REC is depicted in Figure 2.1(b). In each REC, six RNs are deployed around the BS at a distance of 2/3 of the cell hexagon diameter. UTs are assumed to be uniformly distributed over the whole coverage area. In the following, the area in which a BS can provide either direct or two-hop communication for UTs is referred to as cell or a REC, while the area in which an AP, either BS or RN, can provide direct communication for UTs is referred to as a sub-cell.

(a) Layout of the BSs.
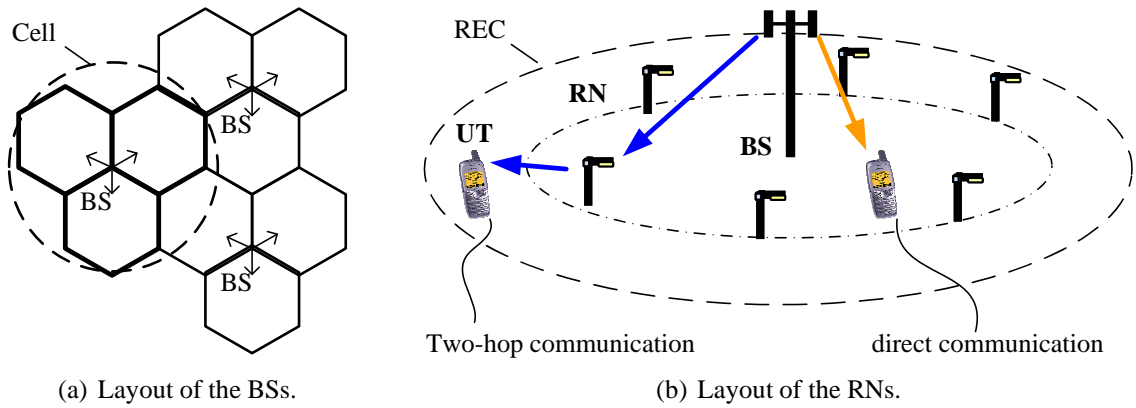
(b) Layout of the RNs.

Figure 2.1: Layout of the cellular system with fixed RNs.

In Table 2.1, the values of the deployment parameters assumed throughout the investigations are given. However, all proposals and algorithms derived later on in Chapter 3, Chapter 4 and Chapter 5 are applicable also for other values of the parameters.

The considered system operates in frequency division duplex mode (FDD) over paired bands of 40 MHz at a carrier frequency of 3.9 GHz and 3.75 GHz, respectively. The site distance D is set to 1 km and each BS is surrounded by six RNs at a distance of 666 m.

All BSs, RNs and UTs are equipped with uniform linear arrays (ULAs) with antenna spacing equal to half wavelength. Each BS has three sectors and each sector of the BS is configured with a ULA consisting of 4 antennas with back-to-front ratio of 20 dB. The maximum total transmit power per sector is 46 dBm. Compared to the BS, a RN is supposed to serve a lower number of sectors with lower maximum transmit power in order to

Table 2.1: Deployment Parameters.

Duplex mode	FDD		
Carrier frequency	3.95 GHz DL / 3.7 GHz UL		
Channel bandwidth	2×40 MHz		
Site distance D	1 km		
	BS	RN	UT
Height	25 m	5 m	1.5 m
No. of sectors	3	1	1
Max transmit power per sector	46 dBm	40 dBm	24 dBm
Antenna configuration	ULA with half wavelength antenna spacing		
Azimuth antenna pattern	$-\min\left[12\left(\frac{\theta}{70°}\right)^2, 20\right]$ dB	0 dB	
No. of antennas per sector	4	4	2
Elevation antenna gain	14 dBi	14 dBi	0 dBi
Receiver noise figure	5 dB	5 dB	7 dB

reduce its cost [IST06]. In the considered system, sectorization is not used at a RN, the maximum transmit power of a RN is assumed to be 40 dBm, and the ULA of a RN is composed of 4 ideal omnidirectional antennas. Each UT is equipped with a 2-element ULA of omnidirectional antennas. The noise figure of 7 dB for UT accounts for cheap mass-market devices, compared to 5 dB for both BS and RN.

2.3 Radio Channel

2.3.1 Introduction

Wireless communications between the transmitter and the receiver are limited by the specific characteristics of the mobile radio channel in the desired frequency range. In general, the transmit signal is affected by

- **Path-loss** due to the distance between the transmitter and receiver [Rap02];

- **Shadowing and diffraction** due to large scale obstacles in the propagation path, which together with path-loss generate the so-called long-term fading [Pro01];

- **Multi-path propagation** due to reflection and scattering at nearby objects which cause the transmit signal to reach the receiver by propagating through different paths

with different delays. This phenomenon represents the so-called channel time dispersion which manifests itself in a varying distortion (or fading) of the receive signal over the transmission band, known as frequency-selective fading. Moreover, if there exists a relative motion between the mobile user terminal and the scatters and/or the reflecting objects, the observed carrier frequency is different from the emitted one. This effect is known as Doppler shift, which makes the phase difference between paths and in turn the distortion (or fading) of the receive signal vary over time. The frequency-selective and also time-variant fading are referred to as short-term fading [Pro01].

In this section, firstly the stochastic modeling of the radio channel is introduced in Section 2.3.2. Note that the configuration of the multiple antennas is also modeled together with the radio channel, as the channel response of one antenna array is dependent on the antenna configuration. Secondly, several useful statistical measures such as delay spread, Doppler spread and angle spread, obtained from the stochastic modeling, are described in Section 2.3.3. The delay spread quantifies the channel time dispersion and in turn the frequency-selectivity; the Doppler spread describes the time variance of the channel; the angle spread is a very important factor determining the spatial correlation between the channel responses on multiple antennas.

2.3.2 Stochastic Channel Modeling

In this section, the stochastic modeling of both long-term and short-term fading is introduced. The most often used stochastic model for long-term fading is exponential path-loss plus log-normal shadowing [Rap02]. Let X_σ denote the log-normal shadowing, which is a zero-mean Gaussian distributed random variable in unit of dB with standard deviation σ (also in unit of dB). Further, by letting d be the distance between the transmitter and the receiver, the attenuation due to the long-term fading in unit of dB is expressed as

$$\text{loss}(d)[\text{dB}] = A + 10\gamma \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma, \tag{2.1}$$

where d_0, A and γ are constant real values. Since the attenuation of a signal is proportional to the square of the propagation distance in free space, the value of γ, known as path-loss exponent, is generally greater than 2. Usually, the values of A, γ and σ are derived from field measurements [Rap02]. According to [IST05c], the model of the attenuation due to the long-term fading for the considered urban macro scenario is obtained from field measurements as

$$\text{loss}(d)[\text{dB}] = 37.49 + 35.74 \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma \quad \text{with} \quad \sigma = 8\,\text{dB}. \tag{2.2}$$

The presence of reflectors and scatters results in multiple versions of the transmit signal that arrive at the receiver, displaced with respect to one another in time and spatial orientation. In the geometric or ray-based model based on stochastic modeling of scatterers, the receive signal is assumed to consist of N time-delayed multi-path replicas of the transmit signal [IST05c]. As shown in Figure 2.2, each of the N paths represents a cluster of M sub-paths. Sub-paths within each path are assumed to have different initial phases but identical delay, because the delay difference among them is too small to be resolvable within the transmission signal bandwidth. Path powers, path delays, and angle properties at both
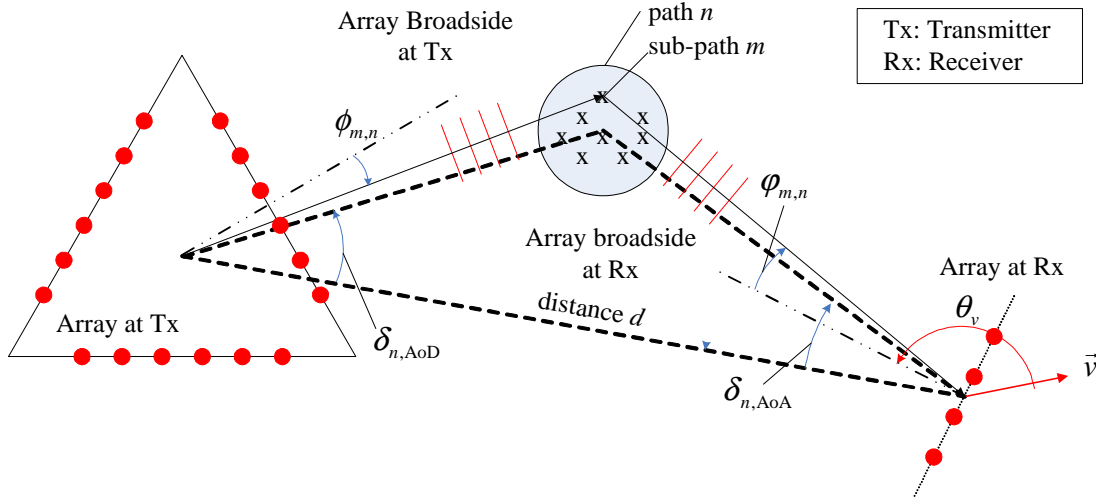
Figure 2.2: Geometric model of multi-path propagation [IST05c].

sides of the link are modeled as random variables defined through individual probability density functions and cross-correlations [IST05c].

To mathematically describe the multi-path propagation, the following notations are introduced. P_n and τ_n denote the power and the delay of the n-th path, respectively. $\phi_{m,n}$ and $\varphi_{m,n}$ represent the angle of departure (AoD) and the angle of arrival (AoA) of the m-th sub-path in the n-th path with respect to the array broadside of the transmitter and receiver, respectively. Further, v is the velocity of the relative motion between the mobile terminals and the surrounding, and its direction with respect to the array broadside of the receiver is represented by θ_v. Note that all defined angles that are measured in a clockwise direction are assumed to be negative in value.

Doppler shift, also referred to as Doppler frequency, is the difference between the observed carrier frequency and the emitted one. It depends on the velocity of the relative motion v, the speed of light c, the carrier frequency f_c, and the angle between the directions of the signal propagation and the relative motion. Since the AoAs of sub-paths differ from each other, different Doppler frequency is observed on each sub-path. The Doppler

frequency for the m-th sub-path of the n-th path is calculated as

$$f_{\mathrm{D},m,n} = \frac{v f_{\mathrm{c}}}{c} \cos(\varphi_{m,n} - \theta_v) \tag{2.3}$$

[Rap02].For a given velocity, the maximum Doppler frequency

$$f_{\mathrm{D,max}} = \frac{v f_{\mathrm{c}}}{c} \tag{2.4}$$

is observed when the direction of a certain sub-path $\varphi_{m,n}$ coincides with the direction of the relative motion θ_v [Rap02].

On each sub-path, by taking the signal transmitted/received at the first antenna element as reference, the signal transmitted/received at each of the other antenna elements experiences a phase shift. For the sake of simplicity, the sub-path and path indices are omitted when presenting the calculation of the phase shift in the following. The phase shift experienced at the i-th antenna element with reference to the first antenna element is given by

$$a_i = e^{j2\pi f_{\mathrm{c}} \tau_i}, \tag{2.5}$$

where τ_i is the time for the signal wave front to pass from the first antenna element to the i-th antenna element [Hay96]. As shown in Figure 2.3, for a given antenna configuration, τ_i only depends on the direction of the incoming wave front, as long as the distance to the source is far enough to make the wave front planar. Thus, the phase shift of the signal on the s-th antenna element with respect to the reference at the transmitter can be formulated as a function of its AoD ϕ and its distance from the reference antenna element d_s, i.e.

$$a_s^{(\mathrm{tx})}(\phi, d_s) = e^{j\frac{2\pi f_{\mathrm{c}}}{c} d_s \sin(\phi)}, \tag{2.6}$$

and the phase shift of the signal on the u-th antenna element with respect to the reference at the receiver can be formulated as a function of its AoA φ and its distance from the reference antenna element d_u, i.e.

$$a_u^{(\mathrm{rx})}(\varphi, d_u) = e^{j\frac{2\pi f_{\mathrm{c}}}{c} d_u \sin(\varphi)}, \tag{2.7}$$

By letting $\psi_{m,n}$ be the initial phase for the m-th sub-path of the n-th path, and G_{Tx} and G_{Rx} represent the antenna gain of the transmitter and the receiver, respectively, the amplitude of the time-variant channel impulse response (CIR) $g_{u,s,n}(t)$ on the n-th path between each antenna pair (u, s) is given by

$$\begin{aligned} g_{u,s,n}(t) = \sqrt{P_n} \sum_{m=1}^{M} \quad & \left(e^{j\psi_{m,n}} \cdot e^{j2\pi f_{\mathrm{D},m,n} t} \dots \right. \\ & \cdot \sqrt{G_{\mathrm{Tx}}(\phi_{m,n})} a_s^{(\mathrm{tx})}(\phi_{m,n}, d_s) \dots \\ & \left. \cdot \sqrt{G_{\mathrm{Rx}}(\varphi_{m,n})} a_u^{(\mathrm{rx})}(\varphi_{m,n}, d_u) \dots \right. \end{aligned} \tag{2.8}$$
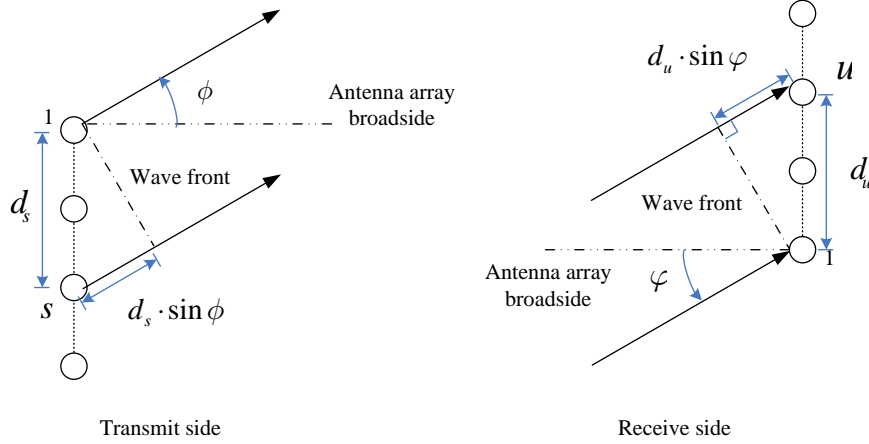
Figure 2.3: Spatial delay incurred when a plane wave impinges on a linear array.

based on (2.3), (2.6) and (2.7).

By denoting with $\delta(\cdot)$ the Kronecker delta function, the CIR between each antenna pair (u, s) at time t is obtained as the superposition of all N paths according to

$$g_{u,s}(\tau, t) = \sum_{n=1}^{N} g_{u,s,n}(t)\delta(\tau - \tau_n). \tag{2.9}$$

2.3.3 Statistical Characterization

As introduced in Section 2.3.2, path powers, path delays, and angle properties at both sides of the link are all considered to be random variables. Their distributions could be different in different environments. For example, due to the higher probability of a larger distance between the transmitter and the receiver, the maximum delay in an outdoor environment is greater than the one in an indoor environment. In this section, several power spectra and statistical parameters that are useful in qualifying statistical channel properties are discussed.

Delay Power Spectrum and Coherence Bandwidth

Since delay power spectrum and coherence bandwidth are discussed in the context of a scalar channel, i.e. a channel between one transmitter antenna and one receiver antenna, for the sake of simplicity, the antenna element indices u and s are omitted in the following.

The auto-correlation function of the time-variant CIR $g(\tau, t)$ in (2.9) is defined as

$$R_g(\tau_1, \tau_2; t_1, t_2) = \frac{1}{2} E[g(\tau_1, t_1) g(\tau_2, t_2)^*] \qquad (2.10)$$

[Pro01]. Under the assumption of a wide-sense-stationary stochastic process, the auto-correlation function $R_g(\tau_1, \tau_2; t_1, t_2)$ in (2.10) does not depend on the absolute time t but depends on the time difference $\Delta t = t_2 - t_1$, i.e.

$$R_g(\tau_1, \tau_2; t_1, t_2) = R_g(\tau_1, \tau_2; \Delta t). \qquad (2.11)$$

[Pro01]. By further assuming uncorrelated scattering, which indicates that the attenuation and phase shift of the path associated with delay τ_1 and with delay τ_2 are uncorrelated unless τ_1 is equal to τ_2 [Pro01], the autocorrelation function $R_g(\tau_1, \tau_2; \Delta t)$ in (2.11) can be simplified to

$$R_g(\tau_1, \tau_2; \Delta t) = R_g(\tau_1; \Delta t)\delta(\tau_1 - \tau_2) = R_g(\tau, \Delta t) \qquad (2.12)$$

[Pro01].

By letting $\Delta t = 0$, the autocorrelation function in (2.12), denoted with $R_g(\tau)$, represents the average power of the channel output as a function of the time delay τ, and is also known as channel multi-path power delay profile or delay power spectrum. It has been proven in [Hoe92] that the delay power spectrum is proportional to the probability density function (PDF) of the multi-path delay τ. The delay spread (DS) σ_τ is defined as the root mean square (RMS) of the multi-path delay τ [Rap02]. A commonly used model assumes that the multi-path delay τ follows negative exponential distribution, called Exponential Power Delay Profile [Pro01]. The exponential power delay profile with a delay spread of σ_τ is given by

$$R_g(\tau) = \frac{1}{\sigma_\tau} \exp\left(-\frac{\tau}{\sigma_\tau}\right), \quad \tau \geq 0. \qquad (2.13)$$

The channel frequency response $h(f, t)$, also called channel transfer function (CTF), is the Fourier transform of the CIR $g(\tau, t)$ with respect to the delay τ [Pro01]. Because of the linearity of the Fourier transform, the CTF $h(f, t)$ has the same statistical characteristics as the CIR $g(\tau, t)$ [OWN96]. Therefore, similar to the autocorrelation function $R_g(\tau, \Delta t)$ of the CIR in (2.12), the autocorrelation function of the CTF $h(f, t)$ does not depend on the absolute frequency and time but depends only on the frequency difference $\Delta f = f_2 - f_1$ and the time difference $\Delta t = t_2 - t_1$, i.e.

$$R_h(f_1, f_2; t_1, t_2) = \frac{1}{2} E[h(f_1, t_1) h^*(f_2, t_2)] = R_h(\Delta f; \Delta t) \qquad (2.14)$$

[Pro01]. It can be shown that the autocorrelation function $R_h(\Delta f; \Delta t)$ in (2.14) is directly related to the autocorrelation function $R_g(\tau; \Delta_t)$ of the CIR by Fourier transform $\mathcal{F}[\cdot]$, i.e.

$$R_h(\Delta f; \Delta t) = \mathcal{F}[R_g(\tau; \Delta t)] \qquad (2.15)$$

[Pro01].

By letting $\Delta t = 0$, the autocorrelation function in (2.14), denoted with $R_h(\Delta f)$, provides a measure of the amplitude correlation as a function of the frequency difference Δf. The coherence bandwidth B_{coh} of the channel can then be defined as the frequency difference at which the absolute value of the correlation function $R_h(\Delta f)$ is reduced to half of its maximum, i.e.

$$|R_h(B_{\text{coh}}, 0)| = 0.5|R_h(0, 0)| \tag{2.16}$$

[Rap02]. With $R_h(\Delta f)$ being the Fourier transform of $R_g(\tau)$ as seen in (2.15), the coherence bandwidth is proportional the reciprocal of the delay spread and can be approximated by

$$B_{\text{coh}} \cong \frac{1}{5\sigma_\tau}, \tag{2.17}$$

as suggested in [Rap02], from which it is inferred that large delay spread results in small coherence bandwidth.

Doppler Power Spectrum and Coherence Time

Since doppler power spectrum and coherence time are also discussed in the context of a scalar channel, for the sake of simplicity, the antenna element indices u and s are again omitted in the following.

The autocorrelation function $R_h(\Delta f, \Delta t)$ in (2.14) represents the correlation of the CTF $h(f, t)$ in both frequency and time directions. By letting $\Delta f = 0$, the autocorrelation function in (2.14), denoted with $R_h(\Delta t)$, provides a measure of the amplitude correlation of the CTF $h(f, t)$ as a function of the time difference Δt. The coherence time T_{coh} of the channel can then be defined as the time duration at which the absolute value of the time correlation $R_h(\Delta t)$ is reduced to half of its maximum, i.e.

$$|R_h(0, T_{\text{coh}})| = 0.5|R_h(0, 0)|. \tag{2.18}$$

In order to relate the Doppler effect to the time variations of the channel, the Fourier transform of $R_h(\Delta t)$ with respect to Δt is defined to be the function $S(f_{\text{D}})$. That is,

$$S(f_{\text{D}}) = \int_{-\infty}^{+\infty} R_g(\Delta t)e^{-j2\pi f_{\text{D}}\Delta t}\mathrm{d}\Delta t. \tag{2.19}$$

In [Hoe92] it has been proven that $S(f_{\text{D}})$ is proportional to the PDF of the Doppler shift f_D. Therefore, $S(f_{\text{D}})$ in (2.19) is known as the Doppler power spectrum [Pro01]. The range of values of f_{D} over which $S(f_{\text{D}})$ is essentially nonzero is called Doppler spread [Rap02]. The Doppler power spectrum $S(f_{\text{D}})$ strongly depends on the type of antenna used. Under the

assumption of vertical monopole antennas and isotropic scatters, the PDF of the Doppler shift follows the Jakes-spectrum

$$S_{f_{\mathrm{D}}}(f_{\mathrm{D}}) = \frac{1}{\pi f_{\mathrm{D,max}}} \cdot \frac{1}{\sqrt{1 - \left(\frac{f_{\mathrm{D}}}{f_{\mathrm{D,max}}}\right)^2}}, |f_{\mathrm{D}}| \leq f_{\mathrm{D,max}} \tag{2.20}$$

[Rap02]. The maximum Doppler frequency $f_{\mathrm{D,max}}$ is determined by the velocity v, carrier frequency f_c and speed of light c, as seen in (2.4). The coherence time can be approximated by the reciprocal of the maximum Doppler frequency $f_{\mathrm{D,max}}$

$$T_{\mathrm{coh}} \cong \frac{1}{f_{\mathrm{D,max}}} = \frac{c}{v f_{\mathrm{c}}} \tag{2.21}$$

[Rap02]. From (2.4) and (2.21), it can be inferred that high velocity v leads to short coherence time T_{coh}, and hence fast time variation of the channel.

Power Angular Spectrum and Spatial Correlation

The spatial correlation between antenna elements is an important parameter, since it is closely related to the channel capacity [GJJV03]. As the attenuation and phase shift of different paths are uncorrelated under the assumption of uncorrelated scattering, the spatial correlation $R_g(i_1, i_2)$ between two antenna elements i_1 and i_2 depends only on the corresponding antenna distance $\Delta d = d_{i_2} - d_{i_1}$, i.e.

$$R_g(i_1, i_2) = R_g(\Delta d). \tag{2.22}$$

Given the PDF of the path angles $p(\theta)$, also known as power azimuth spectrum (PAS), the spatial correlation between two antenna elements spaced Δd meters apart can be determined as follows:

$$R_g(\Delta d) = \int p(\theta) \exp\left(j\frac{2\pi \sin(\theta)}{\lambda}\Delta d\right) d\theta. \tag{2.23}$$

[GSsS$^+$03]. Similarly as for the delay spread, the angle spread (AS) σ_θ is defined as the RMS of the AoD/AoA of each of the multi-path components [Hay96]. In a rich-scattering environment, $p(\theta)$ can be modeled as Gaussian distributed with mean value $\bar{\theta}$ and variance σ_θ, resulting from the Central Limit Theorem [Pap65]. Hence, the corresponding spatial correlation is equal to

$$R_g(\Delta d) = R_g(\Delta, \bar{\theta}, \sigma_\theta) \approx e^{-j2\pi\frac{d}{\lambda}\cos(\bar{\theta})} \cdot e^{-\frac{1}{2}(2\pi\frac{\Delta d}{\lambda}\sin(\bar{\theta})\sigma_\theta)^2} \tag{2.24}$$

[Asz95]. From (2.24), it can be inferred that large antenna spacing and/or large angle spread lead to low spatial correlation and vice versa.

In the considered urban macro scenario, the BS is located well above the rooftop, and so the AS at the BS is relatively small, which causes the channel response of the antenna array at the BS to be highly correlated with each other. According to [IST05c], typical values for the AS at the BS and at the UT are $8°$ and $53°$, respectively, and a typical value for the delay spread is about $310\,$ns.

2.4 Orthogonal Frequency Division Multiplexing

In this section, the benefits of the OFDM transmission technique is reviewed and the modeling of the OFDM transmission system is presented.

With increasing data rate, the symbol duration becomes shorter and shorter. If the symbol duration becomes smaller than the delay spread of the multi-path channel, or equivalently the channel exhibits selectivity in frequency domain, the whole system will heavily suffer from inter-symbol interference (ISI) [Pro01]. An approach to prevent ISI is parallel data transmission, known as multi-carrier (MC) modulation [Sal67, WE71]. It converts a high-rate data stream at symbol rate $1/T$ into N parallel low-rate sub-streams transmitted on N sub-carriers. The resulting MC symbol duration, defined as the symbol duration of each sub-carrier, $T_s = NT$, linearly increases with the number of sub-carriers. The increased symbol duration reduces the impact of multi-path time dispersion.

In the family of MC modulation systems, the parallel data streams are filtered by interpolating filters in general with proper spectral characteristics prior to sub-carrier modulation [BC02]. When all the filters are frequency-shifted versions of one prototype filter and the sub-carriers are equally spaced, the MC modulation can be efficiently implemented through a bank of polyphase filters and discrete Fourier transform (DFT) [Vai92]. Due to the finite steepness of the filter roll-offs, the spacing of the sub-carriers is commonly greater than the Nyquist bandwidth $1/T_s$ so as to eliminate inter-carrier interference (ICI). This leads to inefficient use of the available frequency band.

Orthogonal frequency division multiplexing (OFDM) is a low complexity technique to bandwidth efficiently modulate parallel data streams to multiple carriers. In OFDM, the prototype filter is of a rectangle shape in the time domain of length T_s and the polyphase filtering disappears in the digital implementation [CEO02].

The block diagram of the OFDM transmission system is illustrated in Figure 2.4. A high-rate data stream at symbol rate $1/T$ is firstly converted into N parallel data stream at symbol rate $1/T_s$ by a serial-to-parallel (S2P) converter, and then the data symbols of the N parallel data streams, designated by

$$\mathbf{x} = [x_1, \ldots, x_N]^{\mathrm{T}} \tag{2.25}$$

are modulated by an N-point inverse discrete Fourier transform (IDFT). The data symbols

after IDFT is given by

$$\mathbf{s} = [s_1, \ldots, s_N]^{\mathrm{T}} = \tilde{\mathcal{F}}^{-1}[\mathbf{x}], \tag{2.26}$$

with $\tilde{\mathcal{F}}^{-1}[\cdot]$ representing the IDFT operation. The data symbols $\mathbf{s}$ of the N parallel streams at symbol rate $1/T_s$ are then converted into one data stream at symbol rate $1/T$ by a parallel-to-serial (P2S) converter. $\mathbf{x}$ in (2.25) and $\mathbf{s}$ in (2.26) are usually referred to as the frequency-domain and time-domain transmit signals, respectively. In practice, IDFT can be implemented very efficiently by inverse Fast Fourier Transform (IFFT) [Bla85]. As the consequence of IDFT, the sinc spectra of the individual sub-carriers strongly overlap. However, by letting the sub-carrier spacing be exactly equal to $1/T_s$, at the maximum of each sub-carrier spectrum, all other sub-carrier spectra are zero. It can then be concluded that the sub-carriers in OFDM are orthogonal to each other with a sub-carrier spacing equal to the Nyquist bandwidth $1/T_s$ [NP00].
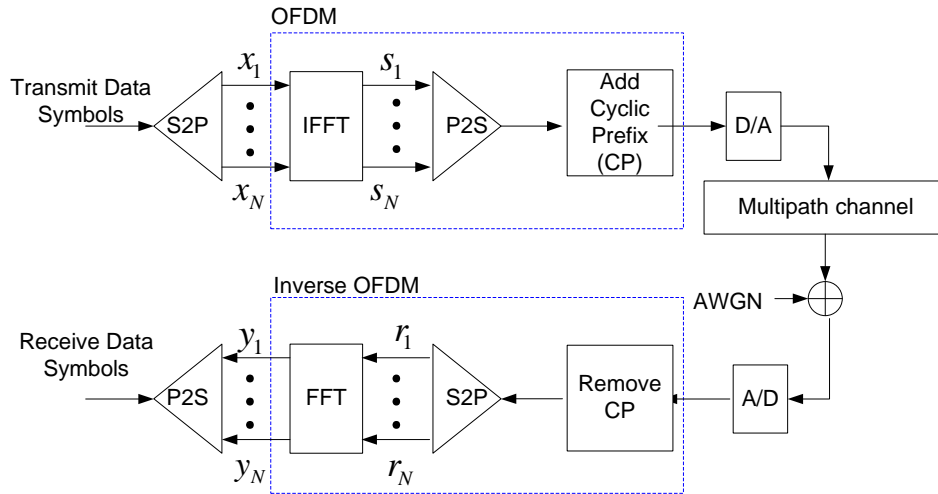
Figure 2.4: Block diagram of OFDM transmission system.

By sampling the channel output at rate $1/T$ at the receiver, each OFDM symbol with duration T_s corresponds to N samples. Letting $\tau_{\max}$ represent the maximum delay spread of the channel, the channel time dispersion is expected to span at most

$$L_c = 1 + \left\lfloor \frac{\tau_{\max}}{T} \right\rfloor \tag{2.27}$$

samples out of the N. For a fixed value of L_c, increasing the number N of sub-carriers reduces the amount of ISI. To completely avoid ISI, a guard interval, which is actually a cyclic prefix (CP), is added to each OFDM symbol [NP00]. The length T_g of the guard interval is chosen to be no less than the channel time dispersion, i.e.

$$T_g \geq L_c T. \tag{2.28}$$

Since only the first L_c samples of the received signal contain ISI, ISI is completely eliminated by removing the guard interval at the receiver.

It is well known that insertion of the CP, transmission over the channel and removal of the CP at the receiver turns the effect of the channel from a linear convolution of the signal and the CIR to a circular one [WG00]. By letting $\mathbf{g} = [g_1, \ldots, g_{L_c}]^T$ denote the discrete CIR sampled at rate $1/T$, the receive signal $\mathbf{r} = [r_1, \ldots, r_N]^T$ after removing the CP is given by

$$\mathbf{r} = \mathbf{s} \circledast \mathbf{g} + \tilde{\mathbf{z}}, \tag{2.29}$$

where $\tilde{\mathbf{z}}$ represents the additive white Gaussian noise (AWGN) and $\circledast$ designates the circular convolution.

As depicted in Figure 2.4, after removing the guard interval, the receive signal $\mathbf{r}$ is demodulated with DFT. Thus, the demodulated signal $\mathbf{y} = [y_1, \ldots, y_N]^T$ is given by

$$\mathbf{y} = \tilde{\mathcal{F}}[\mathbf{r}], \tag{2.30}$$

with $\tilde{\mathcal{F}}[\cdot]$ representing the DFT operation. Let the DFT of the discrete CIR $\mathbf{g}$, also referred to as the discrete CTF, be given as

$$\mathbf{h} = [h_1, \ldots, h_N]^T = \tilde{\mathcal{F}}[\mathbf{g}], \tag{2.31}$$

and the DFT of AWGN $\tilde{\mathbf{z}}$ as

$$\mathbf{z} = [z_1, \ldots, z_N]^T = \tilde{\mathcal{F}}[\tilde{\mathbf{z}}]. \tag{2.32}$$

Recall that, when the Fourier transform is performed with DFT/IDFT, the time-domain circular convolution is equivalent to a frequency domain multiplication [OWN96]. Taking DFT on both sides of (2.29) and substituting (2.30), (2.26), (2.31) and (2.32) results in

$$\mathbf{y} = \tilde{\mathcal{F}}[\mathbf{s} \circledast \mathbf{g} + \tilde{\mathbf{z}}] = \tilde{\mathcal{F}} \left[\tilde{\mathcal{F}}^{-1}[\mathbf{x}] \right] \odot \tilde{\mathcal{F}}[\mathbf{g}] + \tilde{\mathcal{F}}[\tilde{\mathbf{z}}] = \mathbf{x} \odot \mathbf{h} + \mathbf{z}, \tag{2.33}$$

where $\odot$ designates the element-wise multiplication. The transmission of the OFDM system can thus be expressed separately for each sub-carrier as

$$y_n = h_n \cdot x_n + z_n, \tag{2.34}$$

where h_n, x_n, y_n and z_n, being the n-th element of $\mathbf{h}$, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, represent the frequency-domain channel response, transmit signal, receive signal and noise on sub-carrier n, respectively. According to (2.34), the signals received on sub-carriers can be separately equalized by a one-tap filter, which contributes significantly to the reduction of receiver complexity.

The choice of the number of sub-carriers is constrained by the following two aspects. On the one hand, for a fixed guard interval, larger number of sub-carriers increases the spectrum efficiency by increasing the useful symbol duration. On the other hand, in order

to let the channel keep constant at least during one OFDM symbol, the OFDM symbol duration should be much less than the channel coherence time, and so the number of sub-carriers could not be arbitrarily large.

The basic OFDM parameters used in this work are outlined in Table 2.2 [IST06]. By choosing a sub-carrier spacing $\Delta f = 39062.5\,\text{Hz}$, each of the $B = 40\,\text{MHz}$ bandwidth for downlink and uplink is divided into $N = 1024$ sub-channels. The resulting OFDM symbol duration T_s is 25.6 μs. By adding a guard interval of 3.2 μs, the total OFDM symbol becomes $T_s' = 28.8\,\mu\text{s}$.

Table 2.2: OFDM parameters.

Bandwidth B	40 MHz
Sub-carrier spacing $1/T_s$	39062.5 Hz
Number of sub-carriers N	1024
Used sub-carrier	$[-512, 512] \setminus \{0\}$
Useful OFDM symbol duration T_s	25.6 μs
Guard Interval T_g	3.2 μs
Total OFDM symbol duration T_s'	28.8 μs (= 25.6 μs + 3.2 μs)

2.5 DL Transmission in a Multi-user MIMO System

2.5.1 Introduction

As shown in Section 2.4, in an OFDMA system signals transmitted on sub-carriers can be regarded as independent and experiencing flat fading. Thus, the transmission on each sub-carrier in a multi-user MIMO OFDMA systems can be considered separately and viewed as a transmission in a multi-user MIMO system over a flat fading channel. In this section, the DL transmission in a multi-user MIMO system over a flat fading channel is discussed.

Multiple antennas installed at the transmitter and receiver can improve the transmission reliability and/or the system throughput by utilizing the spatial dimension [GSsS+03, Fos96]. Firstly, multiple antennas can be used to combat channel fading by employing diversity techniques at transmitter and/or receiver [GSsS+03]. Secondly, when when both transmitter and receiver are equipped with multiple antennas, a number of parallel channels can be established, over which different sub-streams of one original data stream can be simultaneously transmitted, known as spatial multiplexing gain [Fos96], which increases the spectral efficiency and thus the system throughput. It was proven in [Tel99] that, for

point-to-point Rayleigh fading channels, MIMO channel capacity scales linearly with the minimum number of transmit and receive antennas in high signal-to-noise-ratio (SNR) regime. If the multiplexed data streams are intended for different users, SDMA is configured [FN94]. In this case, effective communication requires suppression of the multi-user interference as well as multi-stream interference.

In the following, the downlink transmission in a multi-user MIMO system is discussed in detail. From the information theory point of view, the downlink multi-user MIMO system is named the MIMO broadcast channel (BC). It was proven in [WSS04] that the MIMO BC capacity region is achieved with dirty paper coding (DPC), which is a multi-user encoding strategy based on interference pre-substraction [Cos83]. Further, the duality relationship between the MIMO BC capacity region and the MIMO Multiple Access Channel capacity region, e.g. uplink, has been derived in [CS03a, VJG03].

Although DPC is the optimal strategy, it is difficult to implement in a practical system due to its high computational effort [CS03b]. An alternative linear precoding approach, beamforming (BF), is generally sub-optimal [JG04] but with reduced complexity relative to DPC. In case of BF, each sub-stream is coded independently and multiplied by a beamforming weighting vector for transmission through multiple antennas [Hay96]. Despite its reduced complexity, BF has been shown to achieve a fairly large fraction of DPC capacity [CS03b] and to approach that of DPC as the number of users goes to infinity thanks to multi-user diversity [SH07]. In this work, the processing on each sub-stream at both transmitter and receiver is restricted to be linear.

The modeling of the linear transceiver for the multi-user DL transmission is introduced in Section 2.5.2, and then the choice of transmit and receive processing are discussed in Section 2.5.3 and Section 2.5.4, respectively. Throughout this section, full CSI is assumed to be available at both transmitter and receiver.

2.5.2 Modeling of Linear Transceiver

In this section, the modeling of the linear transceiver in a MIMO system is presented.

Consider a multi-user downlink system with K users and a single AP, either BS or RN. The AP is equipped with M_t transmit antennas, and user k is equipped with M_{r_k} receive antennas. The total number of antennas at the receivers is then $M_r = \sum_k M_{r_k}$. Further, by letting L_k denote the number of data symbols for user k, the total number of simultaneously transmitted data symbols is equal to $L = \sum_k L_k$. It is assumed that $L \leq M_t$ and $L_k \leq M_{r_k}, k = 1, \ldots, K$.

As depicted in Figure 2.5, the transmit signal intended for user k, denoted by vector $\mathbf{d}_k$ of dimension L_k, is multiplied by an $M_t \times L_k$ transmit filter $\mathbf{W}_k$, also called transmit beamforming matrix. After propagation over the channel and the perturbation by AWGN,

the receive signal of user k, denoted with vector $\mathbf{y}_k$ of dimension M_{r_k}, is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{W}_k \mathbf{d}_k + \mathbf{H}_k \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{W}_j \mathbf{d}_j + \mathbf{z}_k, \tag{2.35}$$

where $\mathbf{H}_k$ denotes the $M_{r_k} \times M_t$ channel matrix from the AP to user k and $\mathbf{z}_k$ represents the M_{r_k}-dimensional noise vector of user k. As the noise $\mathbf{z}_k$ is assumed to be AWGN, its variance is given by $E[\mathbf{z}_k \mathbf{z}_k^{\mathrm{H}}] = \sigma_z^2 \mathbf{I}$. The first term on the right-hand-side of (2.35) is the desired signal for user k, while the second term is the interference seen by user k from the other users' signals.
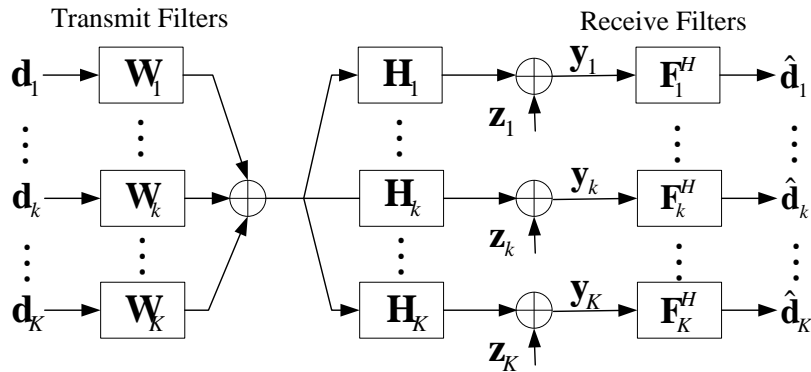
Figure 2.5: Block diagram of downlink transmission in a multi-user MIMO system with linear transceiver.

At the receiver of user k, the estimate $\hat{\mathbf{d}}_k$ of the transmit signal $\mathbf{d}_k$ is obtained by applying a filter represented by an $M_{r_k} \times L_k$ matrix $\mathbf{F}_k$, also called receive beamforming matrix, to the receive signal, yielding

$$\hat{\mathbf{d}}_k = \mathbf{F}_k^{\mathrm{H}} \mathbf{y}_k. \tag{2.36}$$

2.5.3 Optimization of Linear Transmit Filter

In this section, the optimization of the linear transmit filter $\{\mathbf{W}_k\}_{k=1}^{K}$ in a MIMO system is addressed based on the model described in Section 2.5.2.

The transmit beamforming matrix $\{\mathbf{W}_k\}_{k=1}^{K}$ is determined on the basis of the available CSI according to a certain criterion for performance optimization, e.g. the sum capacity maximization [CS03b]. However, finding the optimal transmit beamforming matrix is still a difficult non-convex optimization problem [SB04b]. Here a suboptimal beamforming

strategy, ZFBF is considered. According to ZFBF, the transmit beamforming matrix is chosen in such a way that the interference among transmit data symbols is completely eliminated at the transmitter side by inverting the channel [Hay96].

Assume that L_k data symbols are expected to be received at the selected L_k out of M_{r_k} receive antennas of user k. Moreover, denote with $\mathbf{H}_k^{L_k}$, $\mathbf{y}_k^{L_k}$ and $\mathbf{z}_k^{L_k}$ the corresponding $M_t \times L_k$ channel matrix, the L_k-dimensional receive signal and the L_k-dimensional noise vector at the L_k selected antennas, respectively. By collecting the transmit signals $\{\mathbf{d}_k\}_{k=1}^K$, the beamforming matrices $\{\mathbf{W}_k\}_{k=1}^K$, the channel matrices $\{\mathbf{H}_k^{L_k}\}_{k=1}^K$, the noise vectors $\{\mathbf{z}_k^{L_k}\}_{k=1}^K$ and the receive signals $\{\mathbf{y}_k^{L_k}\}_{k=1}^K$ of all users in $\mathbf{d}$, $\mathbf{W}$, $\mathbf{H}$, $\mathbf{z}$ and $\mathbf{y}$, respectively, as follows

$$
\begin{aligned}
\mathbf{W} &= [\mathbf{W}_1, \ldots, \mathbf{W}_K] & &\in \mathbb{C}^{M_t \times L}, \\
\mathbf{d} &= \left[(\mathbf{d}_1)^{\mathrm{T}}, \ldots, (\mathbf{d}_K)^{\mathrm{T}}\right]^{\mathrm{T}} & &\in \mathbb{C}^{L}, \\
\mathbf{H} &= \left[(\mathbf{H}_1^{L_1})^{\mathrm{T}}, \ldots, (\mathbf{H}_K^{L_K})^{\mathrm{T}}\right]^{\mathrm{T}} & &\in \mathbb{C}^{M_r \times L}, , \\
\mathbf{z} &= \left[(\mathbf{z}_1^{L_1})^{\mathrm{T}}, \ldots, (\mathbf{z}_K^{L_K})^{\mathrm{T}}\right]^{\mathrm{T}} & &\in \mathbb{C}^{L}, \\
\mathbf{y} &= \left[(\mathbf{y}_1^{L_1})^{\mathrm{T}}, \ldots, (\mathbf{y}_K^{L_K})^{\mathrm{T}}\right]^{\mathrm{T}} & &\in \mathbb{C}^{L},
\end{aligned}
\tag{2.37}
$$

the downlink transmission system described by (2.35) can be represented as

$$
\mathbf{y} = \mathbf{HWd} + \mathbf{z}. \tag{2.38}
$$

The Moore-Penrose pseudo-inverse of the channel matrix

$$
\tilde{\mathbf{W}} = \mathbf{H}^{\mathrm{H}} \left(\mathbf{HH}^{\mathrm{H}}\right)^{-1} \tag{2.39}
$$

[Hay96], is one choice of the transmit beamforming matrix $\mathbf{W}$ for ZFBF. $\tilde{\mathbf{w}}_l$, as the l-th column in matrix $\tilde{\mathbf{W}}$ in (2.39), is the transmit beamforming vector for the l-th data symbol. By further letting the beamforming vectors $\tilde{\mathbf{w}}_l$ be normalized by their norms, i.e.

$$
\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_L] \quad \text{with} \quad \mathbf{w}_l = \frac{\tilde{\mathbf{w}}_l}{\|\mathbf{w}_l\|^2}, \quad \forall l = 1, \ldots, L, \tag{2.40}
$$

the l-th data symbol can be regarded as transmitted on the channel with the effective gain G_l

$$
G_l = 1/[(\mathbf{HH}^{\mathrm{H}})^{-1}]_{l,l} \tag{2.41}
$$

without interference [DS05].

ZFBF, also called channel inversion, is in general power inefficient, as explained in the following. When the spatial correlation is high among users, the channel matrix $\mathbf{H}$ is ill-conditioned. The inversion of an ill-conditioned channel matrix significantly reduces the effective channel gain in (2.41). The drawbacks of channel inversion are ultimately caused by the stringent requirement of nulling the interference. Accepting a limited amount of interference at the receiver would enable the enhancement of the signal power and in turn improve the overall SINR. The "regularized" channel inversion proposed in [PHS05]

relaxes the interference-nulling requirement and improves the channel gain by adding a multiple of the identity matrix before inverting, i.e.

$$\mathbf{W} = \beta \mathbf{H}^{\mathrm{H}} \left(\mathbf{H} \mathbf{H}^{\mathrm{H}} + \alpha \mathbf{I} \right)^{-1}, \tag{2.42}$$

where β is a real-valued scalar to fulfill a certain total transmit power constraint and an appropriate selection of the scalar α can maximize the SINR at the receiver [PHS05].

In future systems, it is expected that users may have different bandwidth and/or SINR requirements. One way to meet these requirements is to adjust the amount of power allocated to each user. ZFBF is advantageous over regularized channel inversion when power adaptation is carried out. Indeed, in case of ZFBF power allocation is quite straightforward, because the effective channels seen by different users are independent, while in case of regularized channel inversion, changing the power of one user will change its interference to all other users. In the latter case, a solution enabling the joint optimization of the transmit beamforming matrix and the power allocation would be desirable [BO99, SB04a, RFLT98], but this would be rather challenging.

When there is a sufficient number of users in the system, the power inefficiency problem for ZFBF caused by an ill-conditioned channel matrix due to the high spatial correlation among users can be avoided by only letting users with low spatial correlation transmit together. The spatial correlation of user i and user j can be for instance defined as the inner product of the dominant eigenvectors of the channel correlation matrix, $|\mathbf{v}_{i,1}^{\mathrm{H}} \mathbf{v}_{j,1}|$ [GV96]. It has been shown that in downlink MISO (multiple input single output) systems where UTs are equipped with a single antenna, ZFBF can achieve the same asymptotic sum capacity as that of DPC if the spatial correlation among users is sufficiently low [DS05, JG04].

In downlink MIMO systems, where more than one antenna is installed at user terminals, another issue arising with ZFBF is the following. Signals received at the closely spaced antennas of the UT are usually of high spatial correlation. Thus, ZFBF suffers from a performance loss when trying to completely eliminate the interference among receive antennas belonging to the same user by a ZFBF transmit beamforming matrix at the transmitter. A scheme named block diagonalization is proposed to overcome this problem. Its key idea is to eliminate the multi-user interference [CM04, SSH04], i.e. the interference between any pair of receive antennas of different users, by choosing

$$\mathbf{W}_k \mathbf{H}_j = 0, \forall j = 1, \ldots, K, j \neq k, \tag{2.43}$$

such that for each user, a data symbol is only interfered by the other data symbols dedicated to the same user. The inter-symbol interference can then be handled in an easier way compared to multi-user interference, e.g. by means of joint detection or successive interference calculation [Hay96].

Alternatively to block diagonalization, another solution to combat the high spatial correlation between the signals at the antennas of the same user is to convert the single user

correlated MIMO channel into orthogonal MISO sub-channels, i.e. to firstly remove the interference between antennas of the same user. As $M_t \geq M_{r_k}$, the rank of the $M_{r_k} \times M_t$ channel matrix $\mathbf{H}_k$ of user k is at most M_{r_k}. Let the singular value decomposition (SVD) [GV96] of $\mathbf{H}_k$ be

$$\mathbf{H}_k = \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^{\mathrm{H}}, \tag{2.44}$$

where $\mathbf{U}_k$ is the matrix containing the M_{r_k} left singular vectors of dimension M_{r_k}, $\mathbf{U}_k = [\mathbf{u}_{k,1}, \ldots, \mathbf{u}_{k,M_{r_k}}]$, $\mathbf{V}_k$ is the matrix containing the M_{r_k} right singular vectors of dimension M_t, $\mathbf{V}_k = [\mathbf{v}_{k,1}, \ldots, \mathbf{v}_{k,M_{r_k}}]$, and $\boldsymbol{\Sigma}_k$ is the diagonal matrix containing the M_{r_k} singular value $\{\sigma_{k,i}\}_{i=1}^{M_{r_k}}$. By multiplying the receive signal $\mathbf{y}_k$ by $\mathbf{U}_k^{\mathrm{H}}$ for all K users, the considered MIMO BC system can be viewed as an MISO BC system with $\sum_{k=1}^{K} M_{r_k}$ single antenna users. The channel vector for the i-th equivalent MISO channel of user k, $i = 1, \ldots, M_{r_k}$, is given by

$$\mathbf{h}_{k,i}^{\mathrm{MISO}} = \sigma_{k,i} \mathbf{v}_{k,i}^{\mathrm{H}}, \tag{2.45}$$

where $\sigma_{k,i}$ and $\mathbf{v}_{k,i}$ are the i-th singular value and the i-th right singular vector of the channel matrix $\mathbf{H}_k$, respectively. Due to the orthogonality among the right singular vectors, there is no interference among the receive antennas of the same user.

In this work, ZFBF based on the equivalent orthogonal MISO sub-channels is assumed in the downlink transmission of the considered multi-user MIMO system.

2.5.4 Optimization of Linear Receive Filter

In this section, the optimization of the linear receive filter $\{\mathbf{F}_k\}_{k=1}^{K}$ in a MIMO system is addressed based on the model described in Section 2.5.2.

The purpose of the receive filters $\{\mathbf{F}_k\}_{k=1}^{K}$ in (2.36) is to compensate for channel distortion, e.g. to enhance the power of the desired signal and/or to reduce the interference, so as to achieve high data rate.

It has been shown in [JKG$^+$02] that three kinds of linear filters, matched filter, zero-forcing filter and MMSE (minimum mean square error) filter, can be identified according to different optimization criteria. The matched filter maximizes the power of the desired signal but takes no care of the interference. The zero-forcing filter completely eliminates the interference but suffers from noise enhancement. The MMSE filter minimizes the mean square error between the transmit signal and its estimate, i.e. $E[\|\hat{\mathbf{d}}_k - \mathbf{d}_k\|^2]$, and finds a trade-off between the signal enhancement of the matched filter and the interference suppression of the zero-forcing filter. The MMSE filter of user k is expressed as

$$\mathbf{F}_k = \left(\mathbf{H}_k \mathbf{W} \mathbf{W}^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{H}_k \mathbf{W}_k. \tag{2.46}$$

As seen from (2.46), the MMSE filter $\mathbf{F}_k$ converges to the matched filter in low SNR regime and converges to zero-forcing filter in high SNR regime [JKG$^+$02]. As the MMSE filter

outperforms the zero-forcing filter and the matched filter in the SNR regime of interest, it is used in this work.

2.6 Frame Structure

In this section, the frame structure is presented. The frame structure is important for the analysis of adaptive resource allocation, especially when investigating the signaling overhead, because the frame duration defines how fast the resource allocation can be adapted to the time-varying channel.

As introduced at the beginning of this chapter, the investigations presented in this work are carried out within the framework of the WINNER project, and so the frame structure proposed in the WINNER project is assumed and presented in the following.

The OFDM symbols are organized into frames of fixed length [IST05a]. Moreover, each frame is divided into two time slots of the same duration. In FDD mode, half-duplex UTs can be assigned to one of two groups which transmit and receive alternatively: if group 1 receives in DL in the first slot of the frame and transmits in uplink (UL) in the second slot, group 2 transmits in UL in the first slot and receives in DL in the second one. Without loss of generality each frame can be assumed to comprise a DL slot followed by an UL slot. [1]   In order to enable fast adaptive resource allocation as well as a short channel feedback loop for the purpose of CSI acquisition in high mobility scenarios, a short frame is chosen to consist of 24 OFDM symbols, resulting in a duration of 0.6912 ms. Further, a super-frame is designed to comprise a preamble followed by 8 frames [IST05a], as shown in Figure 2.6. The super-frame is synchronized among all BSs and RNs. Given such a short frame duration, operations such as synchronization and random access are not necessary to be performed once every frame and therefore are carried out once every super-frame in the preamble of the super-frame. Accordingly, the preamble is designed to consist of two synchronization slots, one time slot reserved for the UL contention-based random access channel (RAC), a set of OFDM symbols that carry the DL broadcast control channel (BCH) messages and a guard interval between UL RAC and DL synchronization slot. Given the parameters for the preamble described in Table 2.3, a super-frame has a duration of 5.8896 ms. Note that the detailed design of the preamble will not be further considered in this work. A summary of the aforementioned parameters of chunk, frame and super-frame can be found in Table 2.3.

As well-known, adjacent sub-carriers and OFDM symbols undergo similar channel fading. Hence, instead of individual sub-carriers and OFDM symbols, a block of adjacent sub-carriers and OFDM symbols are chosen as the basic time-frequency resource unit, named

---

[1]In TDD mode, a frame is assumed to comprise a DL slot followed by an UL slot. Therefore, each frame can be assumed to comprise a DL slot followed by an UL slot, regardless of the duplex mode.
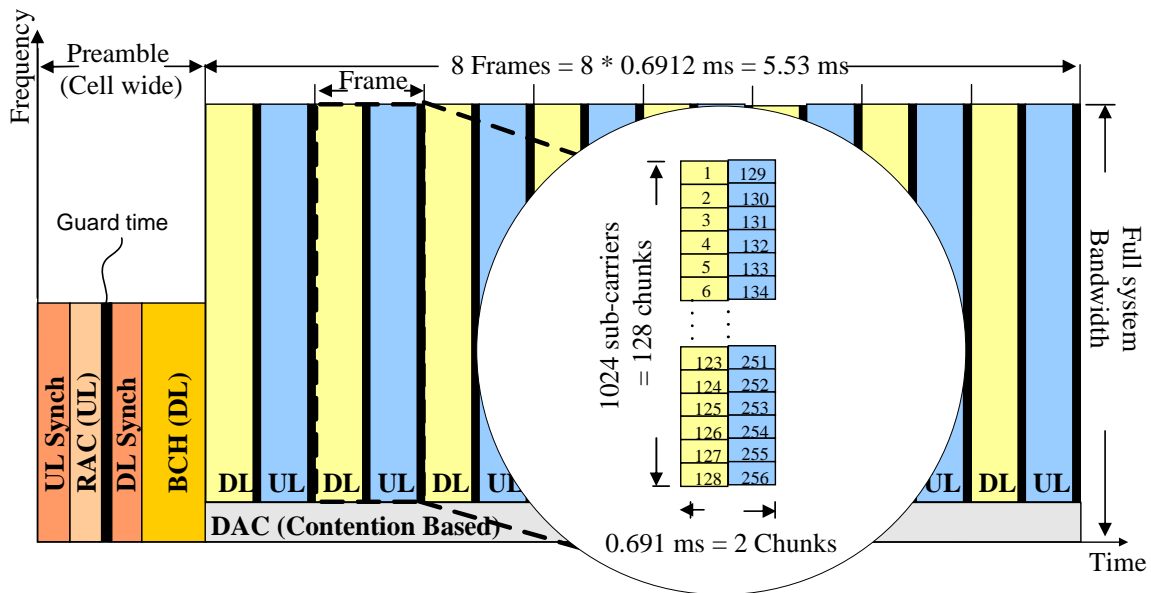
Figure 2.6: Structure of frame and super-frame [IST05a].

Table 2.3: Parameters of Chunk, Frame and Super-frame.

DL/UL Frame dimension	1024 sub-carriers $\times$ 12 OFDM symbols
Frame duration	0.6912 ms
Super-frame duration	5.8896 ms
Preamble duration	0.360 ms
UL Synch.	2 OFDM symbols
RAC (UL)	3 OFDM symbols
Guard time	14.4 μs
DL Synch.	3 OFDM symbols
BCH (DL)	4 OFDM symbols
Chunk dimension	8 sub-carriers $\times$ 12 OFDM symbols
No. of chunks per DL/UL frame	128×2
No. of data symbol per chunk	75

chunk. The chunk-wise resource allocation reduces not only the signaling for the delivery of the resource allocation results, but also the computational complexity of adaptive resource allocation optimization. Each chunk contains data symbols and pilot symbols. It may also contain control symbols that are placed within the chunk. It is assumed hereafter that each chunk is defined as a block of 8 sub-carriers by 12 OFDM symbols and contains 75 data symbols.

As shown in Section 2.5, each time-frequency resource unit can be shared by independent multiple data symbols transmitted over orthogonal or semi-orthogonal beamforming vectors, and the multiple data streams can be dedicated to either single user or multiple users. A chunk layer, defined as a spatial layer of a chunk, is identified by a transmit beamforming vector and carries a single data stream, as depicted in Figure 2.7. The number of offered information bits per chunk layer depends on the utilized modulation and coding scheme.
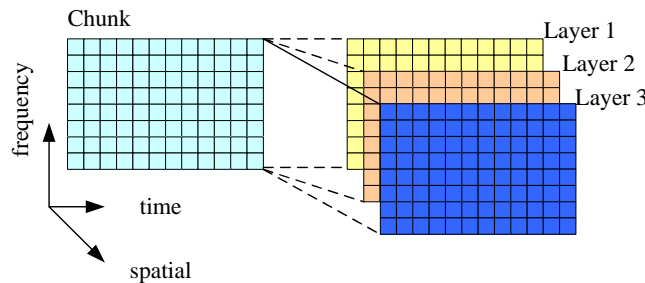
Figure 2.7: Illustration of chunk layers.

2.7 Link Adaptation

In this section, the link adaptation is introduced and the schemes used in this work are described. The availability of information about the current channel condition at the transmitter allows to choose transmission parameters flexibly with respect to the observed link status rather than assuming the average or the worst-case conditions. Compared to a system with fixed transmission parameters, the system performance in terms of data rate, power consumption and reliability can thus be improved. Techniques for performing such kind of adaptation are commonly termed link adaptation. The most important aspects of link adaptation are:

- **Dynamic Fragmentation:** The length of data packets can be varied to account for the tradeoff between packet error probability, increasing with the packet length, and protocol overhead, decreasing with the packet length.

- **Adaptive Modulation and Coding (AMC):** From a set of allowed combinations of modulation format and code rate, i.e. a set of MCS, the most suitable one is selected with respect to the channel condition, e.g. the receive SINR [Cav72, Vuc91]. The selection directly determines the bandwidth efficiency of the transmission. In the context of multi-carrier systems, the MCS used for each sub-carrier may differ from each other due to frequency-selective fading [FH96].

- **Bit and Power Loading:** According to the water-filling principle, the maximum sum data rate over a number of parallel channels under a certain total power constraint is achieved by allocating the strongest channel with the highest power [Gal68]. Bit and power loading, known as water-filling using discrete MCS, consists in adaptively distributing the total available power over multiple channels and/or multiple users in order to achieve the highest power efficiency [TCS04, WCLM99].

As proposed in the WINNER project, the code word length is fixed [IST05a], and so dynamic fragmentation will not be further discussed. Upon the arrival of a certain data packet at medium access control (MAC) layer, a data packet is segmented into segments of a fixed size. Each segment is encoded by means of a forward error correction (FEC) code, yielding an FEC block. In each frame, an integer number of FEC block are transmitted. All FEC blocks of the same user to be transmitted in the frame are interleaved before being mapped onto the assigned chunks/chunk layers, as depicted in Figure 2.8. Both MCSs and
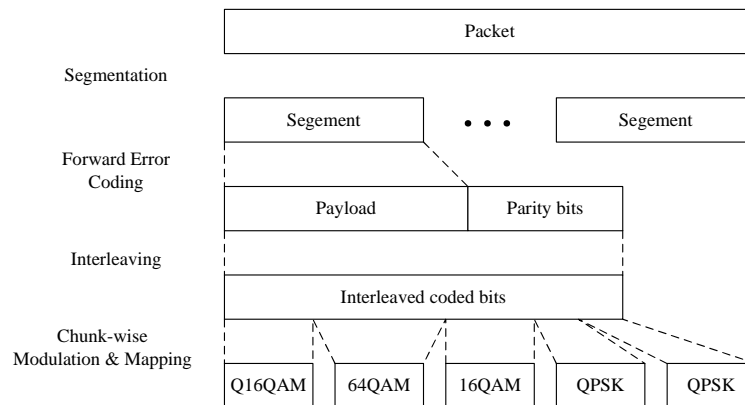
Figure 2.8: Adaptive modulation and coding scheme with fixed code word length [IST05a].

power of the individual chunks can be adapted with respect to the current channel condition if channel knowledge is available at the transmitter.

Consider a system with K users and N_c chunks, and further assume the channel fading on each chunk to be flat. Let $G_{k,n}$ represent the channel gain of user k on resource n, $I_{k,n}$ and $N_{k,n}$ represent the multi-user interference and the AWGN experienced by user k

on chunk n, respectively. By denoting with $r_{k,n}$ the target data rate in unit of bits/symbol allocated to user k on chunk n and with $f(c)$ the required SINR for reliable reception of c bits/symbol, the required transmit power $P_{k,n}$ become

$$P_{k,n} = f(r_{k,n}) \cdot \frac{I_{k,n} + N_{k,n}}{G_{k,n}}, \qquad (2.47)$$

Based on (2.47), bit and power loading aims at the minimization of the required transmit power $P_{k,n}$ for a target data rate $r_{k,n}$, while adaptive modulation and coding targets at the maximization of the achievable data rate $r_{k,n}$ given a certain power allocation $P_{k,n}$.

In case of ZFBF, there exists no multi-user interference, i.e. $I_{k,n} = 0$, and the channel gain $G_{k,n}$ is computed according to (2.41), which depends not only on the channel response of user k but also on the channel response of other users served on chunk n. By denoting with σ_z^2 the power of the Gaussian noise $N_{k,n}$, the one-to-one mapping in (2.47) between the required transmit power and the target data rate of user k on chunk n is expressed as

$$P_{k,n} = f(r_{k,n}) \cdot \frac{\sigma_z^2}{G_{k,n}}. \qquad (2.48)$$

As the data rate is determined by the applied MCS, given the finite set of supported MCSs, the number of supported data rates is also finite. Let

$$\Gamma_r = \{\gamma_{r,1}, \ldots, \gamma_{r,Q}\} \qquad (2.49)$$

denote the set of possible data rates with cardinality Q and further assume its elements are arranged in ascending order, i.e. $\gamma_{r,q} > \gamma_{r,q-1}$.

The function $f(c)$ reflecting the relationship between required SINR and target data rate is specific for each coding scheme and the required reliability of the data transmission, e.g. the targeting BER. For instance, in case of uncoded QAM,

$$f(c) = \zeta(2^c - 1)/3, \qquad (2.50)$$

with ζ depending on the target BER, e.g. $\zeta = 12.1157$ for a target BER of 10^{-3} [Pro01].

A rate-compatible punctured block low-density parity check code (BLDPCC) of mother code rate 1/2 is used in this work. With the four modulation schemes, BPSK, QPSK, 16-QAM and 64-QAM, ten MCSs are supported in the considered system, as summarized in Table 2.4 [IST06]. The code rates 2/3 and 3/4 are obtained by puncturing the mother code.

In order to have an even higher granularity for AMC, i.e. a larger set of supported data rate Γ_r in (2.49), a larger code rate set

$$\mathcal{R} = \left\{ \frac{24}{48}, \frac{24}{44}, \frac{24}{40}, \frac{24}{36}, \frac{24}{32}, \frac{24}{30}, \frac{24}{28}, \frac{24}{26} \right\} \qquad (2.51)$$

Table 2.4: Supported Modulation and Coding Schemes.

MCS	1	2	3	4	5	6	7	8	9	10
Modulation format	BPSK		QPSK			16-QAM			64-QAM	
Code rate	1/2	2/3	1/2	2/3	3/4	1/2	2/3	3/4	2/3	3/4
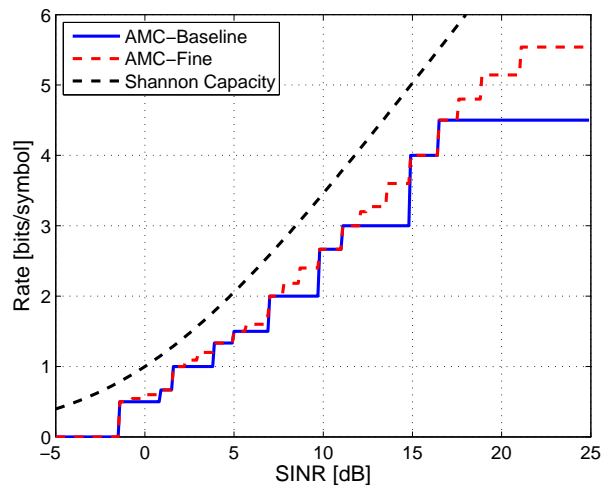
Figure 2.9: SINR thresholds for different modulation formats and coding rates, BLDPCC, code word length = 360 information bits, target CWER = 0.1 [SBC07].

obtained from puncturing can be used instead [SBC07]. To distinguish the two approaches, the previous is referred to as AMC-Baseline and the latter is called AMC-Fine hereafter.

For a code word with fixed information length of 360 bits, the SINR thresholds for switching between different MCSs for a target code word error rate (CWER) of 0.1 is depicted in Figure 2.9 [SBC07].

2.8 Further Assumptions

For performance evaluation, the following assumptions are considered throughout the discussion in this work.

- **Single cell environment**

  Only a single cell is evaluated without considering inter-cell interference. When fixed

relay nodes are introduced within the cell, the BS and the multiple RNs will interfere each other. This kind of interference, referred to as inter-sub-cell interference, are fully modeled in the investigation on resource allocation among BS and RNs in the same REC in Chapter 5.

- **Equal power distribution over the whole bandwidth**

  The total transmit power is usually limited by the capability of the hardware, as stated in Table 2.1. It is further assumed that the total transmit power is equally distributed over the available bandwidth, which simplifies the resource allocation problem. Moreover, the constant power allocation on all sub-carriers allows to predict the inter-cell interference, which is desired for inter-cell interference management. Nevertheless, the power allocation among the chunk layers is still enabled.

- **Infinitely backlogged user queues**

  The user queues are assumed to be infinitely backlogged. This means that, when one user is scheduled for transmission, it always has some data packets to transmit.

- **Mutual Information based Link to System Level Interface**

  The simulation results presented in this work are obtained through system level simulations. The performance is evaluated in terms of data throughput, defined as the number of information bits correctly received at the receiver, i.e.

  $$\rho = \rho_{\mathrm{tx}} \cdot (1 - \mathrm{CWER}), \tag{2.52}$$

  where ρ_{tx} denotes the number of information bits transmitted. The CWER is derived according to the link quality using a mutual information based performance model [BKA05]. A brief description of the approach is provided as follows.

  1. Calculate the receive SINR values for all resource elements used by the FEC code word of interest. The receive SINR depends on power allocation, transmit and receive processing as well as instantaneous channel and interference characteristics.

  2. Compute average mutual information per bit (MIB) according to

     $$\overline{\mathrm{MIB}} = \frac{\sum_{q=1}^{Q} I_{M_q}\left(\frac{\mathrm{SINR}_q}{\beta}\right)}{\sum_{q=1}^{Q} m_q}, \tag{2.53}$$

     where

SINR_q is the SINR on resource element q,

M_q is the identifier of the modulation format applied on resource element q,

m_q is the number of bits represented by the symbol transmitted on resource element q,

I_{M_p} is the mutual information associated with the modulation format M_p as a function of SNR/SINR [CTB96],

β is the optimization parameter to be derived from link level simulations, which is chosen to be 1 in this work [BKA05].

3. Map the average MIB in (2.53) to a CWER. The mapping between MIB and CWER is generally specific to code rate, code type and code word length [BKA05] and is derived from link level simulations.

3 Adaptive Resource Allocation in a Single Cell

3.1 Introduction

As stated in Section 2.4, OFDM is a low complexity technique to bandwidth efficiently modulate parallel data streams to multiple carriers. It is considered as the leading technique for the next generation wireless communication systems [STT$^+$02]. An OFDM system can support the simultaneous transmission to multiple users with different and variable data rate by assigning them a different number of disjoint sub-carriers in a FDMA fashion, referred to as OFDMA. This provides high flexibility and granularity. Since the channel fading is frequency-selective and independent among different users, the system performance can be improved by means of adaptive sub-carrier assignment as well as bit and power loading [WCLM99, RC00], referred to as adaptive OFDMA hereafter.

Moreover, when the AP, either BS or RN, is equipped with multiple antennas, the spectrum efficiency can be increased by SDMA, i.e. by spatially separating multiple users served on the same time-frequency resource by means of orthogonal or semi-orthogonal antenna beamforming [FN94]. The number of users separable by SDMA is generally limited by the number of multiple antenna elements and the achievable capacity depends on the spatial correlation among users. Therefore, the system performance can be optimized by properly selecting the groups of users sharing the time-frequency resources [DS05], whereas is referred to as adaptive SDMA in the following.

Adaptive OFDMA in wide-band OFDM systems has been well studied in [WCLM99, RC00] and many algorithms for adaptive SDMA in narrow-band flat-fading channel are also proposed in the literature [FGH05, JG04, DS05]. However, adaptive resource allocation becomes more challenging in a frequency-selective broadband OFDM multiple antenna system, where both adaptive OFDMA and adaptive SDMA are desired.

In [ZL05], it is assumed that the users' performance will not be affected by sharing resources with users when their spatial correlation is sufficiently low over the whole transmission bandwidth. Thus, the adaptive resource allocation is proposed to be separated into two steps: spatially correlated users are firstly grouped together while ensuring low spatial correlation between users in different groups, and then adaptive OFDMA is independently

carried out within each group. However, it is observed that the user spatial correlation properties are frequency-selective in some cases such as the considered urban-macro scenario, which causes troubles in the application of this two-step approach. Intuitively, the frequency-selective user spatial correlation over the whole bandwidth could be measured by the average, but the approach still will not provide good results compared to joint optimization of adaptive OFDMA and adaptive SDMA.

In [Wil06] spatial compatible users are adaptively selected to build an SDMA user group on each time-frequency resource, i.e. chunk, but no dependency exists among allocation of each chunk. The ordering of the chunks is non-adaptive to the channel conditions, which is referred to as disjoint OFDMA/SDMA hereafter.

Different to the these approaches, joint optimization of adaptive OFDMA and SDMA is introduced by the author of this thesis in [CZT07] and will be discussed in this chapter. The joint approach intends to make joint optimization of the adaptive resource in both frequency and spatial domains, and it is expected to provide better results. As the computational complexity of the optimal solutions is unaffordable in practical systems, cf. Section 3.2.3, alternative sub-optimal algorithms are proposed.

In this chapter, adaptive resource allocation for the DL transmission at one AP in a single cell is studied. This is supposed to be performed at the AP for each frame. Perfect full CSI is available at the transmitter, i.e. the AP, and the channel fading is flat within one chunk. The AP is equipped with multiple antennas and users may share the same chunk by means of ZFBF, cf. Section 2.5. All users are assumed to be equipped with single antenna, but the proposed algorithms are applicable in the case that users have multiple antennas as well, since a user equipped with M_{r_k} antennas can be viewed as M_{r_k} single-antenna users, cf. Section 2.5.

According to the optimization objective function, two classes of optimization problem in the literature have been identified for adaptive resource allocation , power minimization problem [WCLM99] and rate maximization problem [RC00]. The power minimization problem aims at achieving the minimum total transmit power under the constraint on the users' data rate. The rate maximization problem is intent on maximizing the sum of the user's data rate under the constraint on the total transmit power. Correspondingly, the power minimization problem and the rate maximization problem are solved in Section 3.2 and Section 3.3, respectively.

In solving the power minimization problem, several variants of the greedy sub-optimal algorithm, as summarized in Table 3.1, are proposed and compared. Moreover, the optimal solution achieved by exhaustive search is also presented so as to evaluate the performance degradation of the sub-optimal solutions with respect to the optimum.

In solving the rate maximization problem, the user fairness metric is a very important factor [KMT98]. Indeed, rate maximization favors user with good channel quality and thus resulting unfairness among users with different channel quality, known as near-far

Table 3.1: Summary of algorithms for the power minimization problem.

1	Optimal algorithm via integer linear programming (ILP)
2	Successive bit insertion without user priority (SBI-Original)
3	Successive bit insertion with weighted user priority (SBI-WP)
4	Successive bit insertion with first user priority (SBI-FP)

problem [HT04]. Two commonly used user fairness strategies, proportional fairness (PF) and max-min fairness (MMF) are adopted in this work. Moreover, two kinds of constraints on transmitter power, per chunk power constraint (PCPC) and per user power constraint (PUPC), are considered. Hence, as outline in Table 3.2, totally four variants of a greedy algorithm, SBI-WP, SBI-FP, SUI-WP and SUI-FP, are proposed for each combination of power constraints and fairness strategies. In order to illustrate the benefit of joint adaptive OFDMA/SDMA optimization approach, the performance of the disjoint approach proposed in [Wil06] is used as a reference curve in performance assessment.

Table 3.2: Summary of algorithms for the rate maximization problem.

PCPC	PF	Successive bit insertion with weighted priority (SBI-WP)
	MMF	Successive bit insertion with first priority (SBI-FP)
PUPC	PF	Successive user insertion with weighted priority (SUI-WP)
	MMF	Successive user insertion with first priority (SUI-FP)

3.2 Power Minimization Problem

3.2.1 Introduction

This section studies the adaptive resource allocation targeting the power minimization under a minimum rate constraint for all users. In Section 3.2.2, the optimization problem is introduced and mathematically formulated. After presenting the method to obtain the optimal solution in Section 3.2.3, a Greedy-based sub-optimal algorithm is proposed and then modified for improvement in Section 3.2.4. Finally, the performance is assessed on the basis of numeric results in Section 3.2.5.

3.2.2 Problem Statement

In this section, the power minimization problem is introduced and mathematically formulated.

A total of K users are assumed to transmit over the N_c available chunks in each frame. All users have to meet certain minimum data rate requirements. Bit and power loading over all users and chunks is optimized in such a way that the total required transmit power is minimized. Let $P_{k,n}$ and $r_{k,n}$ represent the transmit power and data rate of user k on chunk n, respectively. $r_{k,n}$ is either zero or chosen from the set of all feasible data rates Γ_r defined in (2.49). By denoting with R_k the data rate request of user k in unit of bits for the current frame, the power minimization problem is mathematically formulated as

$$\min \sum_{k,n} P_{k,n} \quad \text{s.t.} \quad \sum_n r_{k,n} \geq R_k \quad \forall k \in [1, K]. \tag{3.1}$$

The relationship between the transmit power $P_{k,n}$ and the target data rate $r_{k,n}$ is determined according to (2.47) or (2.48) when ZFBF is applied.

As maximally M_t users can be served on one chunk by means of ZFBF, with M_t being the number of antenna elements at the AP, and one from Q feasible data rates in Γ_r is chosen by each user, the size of solution space for the problem described by (3.1) is equal to

$$\left((C_K^0 + C_K^1 \cdot Q + \ldots + C_K^{M_t} \cdot Q^{M_t}) \right)^{N_c}, \tag{3.2}$$

with C_K^i denoting the combination of choosing i from K. The solution space in (3.2) is in the order of $(KQ)^{M_t N_c}$. In a practical system, e.g. the considered system described in Section 2.6, the number N_c of chunks is very large, and so the exhaustive search over such a huge solution space is prohibitive due to the extremely high computational complexity.

3.2.3 Optimal Solution

In this section, the optimal solution of the power minimization problem formulated in the previous section is presented.

The optimization problem described by (3.1) is not only nonlinear but also contains integer variables, thus, it is computationally intensive to solve. It is proposed in [KLKL01], by transforming the problem into a linear optimization problem with integer variables, i.e. integer linear problem (ILP), the optimal solution can be obtained with the standard integer programming tools [Wol98].

For the sake of conciseness, the transformation presented in the following is restricted to the case in which every chunk is shared by at most two users and ZFBF is used. Nevertheless, the transformation can be easily extended to other cases, such as more transmit antennas or other beamforming techniques.

Let k and j be the user indices, n be the chunk index, and q be the index of the feasible data rates, the $K^2 N_c Q$ binary variables $\delta_{k,j,n,q}$ defined as

$$
\delta_{k,j,n,q} =
\begin{cases}
1 & \begin{aligned} &\text{if users } k \text{ and } j \text{ share chunk } n \\ &\text{and user } k \text{ transmits at a rate of } \gamma_{r,q} \text{ in unit of bits/symbol} \end{aligned} \\
0 & \text{otherwise}
\end{cases}
\tag{3.3}
$$

can describe the solution of the allocation problem.

The data rate of user k on chunk n can then be given by

$$
r_{k,n} = S_{\text{chunk}} \sum_{q=1}^{Q} \sum_{j=1}^{K} \delta_{k,j,n,q} \gamma_{r,q},
\tag{3.4}
$$

where S_{chunk} denotes the number of data symbols in one chunk.

Let $G_{k,j,n}$ denote the effective channel gain of user k on chunk n given that user j is also served on chunk n, which is calculated according to (2.41). According to (2.48), the required power for user k to share chunk n with user j and transmit at a rate of $\gamma_{r,q}$, i.e. when $\delta_{k,j,n,q} = 1$, is given by

$$
P_{k,n} = f(\gamma_{r,q}) \frac{\sigma_z^2}{G_{k,j,n}}.
\tag{3.5}
$$

According to (3.4) and (3.5), the optimization problem described by (3.1) can be transformed into the following ILP,

$$
\min \sum_{k=1}^{K} \sum_{n=1}^{N_c} \sum_{q=1}^{Q} \sum_{j=1}^{K} f(\gamma_{r,q}) \frac{\sigma^2 \cdot \delta_{k,j,n,q}}{G_{k,j,n}},
\tag{3.6}
$$

subject to

$$
\sum_{n=1}^{N} \sum_{q=1}^{Q} \sum_{j=1}^{K} \gamma_{r,q} \cdot S_{\text{chunk}} \cdot \delta_{k,j,n,q} \geq R_k \quad \forall k \in [1, K],
\tag{3.7}
$$

$$
\sum_{q=1}^{Q} \sum_{k=1}^{K} \sum_{j=k}^{K} \delta_{k,j,n,c} \leq 1 \quad \forall n \in [1, N],
\tag{3.8}
$$

$$
\sum_{q=1}^{Q} (\delta_{k,j,n,q} - \delta_{j,k,n,c}) = 0 \quad \forall k, j \in [1, K], \forall n \in [1, N],
\tag{3.9}
$$

$$
\delta_{k,j,n,q} \in \{0, 1\} \quad \forall k, j \in [1, K], \forall n \in [1, N_c], \forall q \in [1, Q].
\tag{3.10}
$$

47

Constraints expressed by (3.7) guarantee the minimum rate requirement. Constraints expressed by (3.8) impose that every chunk can be assigned to exactly one or two users. Constraints expressed by (3.9) state that if user k is sharing a certain chunk with user j then also user j has to share the same chunk with user k. Finally, constraints expressed by (3.10) represent the integrality constraints for the set of variables.

To compute the optimal solution of this ILP, the ILP solver *glpsol* that is included in the GLPK (GNU Linear Programming Kit) package [Mak] is used.

The required computational effort to solve an ILP increases exponentially with the number of variables and constraints, i.e. with the number of transmit antenna elements, users and chunks [Wol98]. Therefore, finding the global optimal solution of the power minimization problem in (3.1) by transforming it into an ILP is still unaffordable in practical systems. In the next section, sub-optimal algorithms to reduce the complexity while still delivering performance close to the optimum is derived.

3.2.4 Sub-optimal Algorithms

In this section, sub-optimal algorithms for the formulated power minimization problem are proposed and analyzed.

A greedy algorithm is any algorithm that follows the problem solving heuristic of making the locally optimum choice at each iteration with the hope of finding the global optimum [Cor01]. For most problems, greedy algorithms mostly but not always fail to find the globally optimal solution, because they might make commitments to certain choices too early, which prevent them from finding the best overall solution later [Cor01]. Nevertheless, they are easier to implement, most of time quite efficient and often yield very good or even the best possible approximation to the optimum [Cor01]. Hence, greedy algorithms have been chosen to solve the optimization problem in adaptive resource allocation for multi-user OFDMA systems [SAE05, ZL05] and narrow-band MIMO-SDMA systems [SCA$^+$06, DS05, JG04, RPS$^+$06], in frequency and spatial domain, respectively.

In this section, a greedy algorithm is proposed to solve the power minimization problem described in (3.1) for a multi-user OFDMA/SDMA system. The resource allocation in such a system requires the joint optimization of OFDMA and SDMA in both frequency and spatial domain.

The algorithm starts with the assignment to all K users of a null rate on all N_c available chunks by setting the individual power allocation to zero, i.e.

$$r_{k,n}^{(0)} = 0, P_{k,n}^{(0)} = 0, \quad \forall k \in [1, K], \forall n \in [1, N_c]. \tag{3.11}$$

In each iteration, for each user and chunk a certain cost function is evaluated corresponding to the user data rate increase of a given $\overline{\Delta r}$ on that chunk. Let $\Psi_{k,n}$ denote the cost function,

which represents the cost for granting a rate increase $\overline{\Delta r}$ to user k on chunk n. Recall that the supported data rates by all feasible data rates are represented by the set Γ_r defined in (2.49). If the current data rate is $\gamma_{r,q}$, the rate increase $\overline{\Delta r}$ is set to

$$\overline{\Delta r} = \gamma_{r,q+1} - \gamma_{r,q}, \tag{3.12}$$

in order to achieve the next supported data rate. As a lower cost indicates better performance, in each iteration the pair of user and chunk (k^*, n^*) that yields the minimum cost is found from all KN_c combinations, i.e.

$$(k^*, n^*) = \arg\min_{(k,n)} \Psi_{k,n}. \tag{3.13}$$

The rate of the selected user is then increased by $\overline{\Delta r}$ on the corresponding chunk, i.e. in the i-th iteration,

$$r^{(i)}_{k^*,n^*} = r^{(i-1)}_{k^*,n^*} + \overline{\Delta r}. \tag{3.14}$$

The algorithm ends when all users have reached the target data rate requirement. As the data rate is increased by a certain amount in each iteration, this algorithm is referred to successive bit insertion (SBI) hereafter.

Let $\Delta P_{k,n}$ denote the power needed to increase the data rate of user k on chunk n by $\overline{\Delta r}$. This amount of power is not only used to grant user k a higher data rate, but also to provide the additional power required by other users served on that chunk to maintain their data rates. Indeed, the effective channel gain of users served on a given chunk generally decreases when a user is added, as expressed in (2.41). In order to minimize the total transmit power, it is quite natural to define the cost function $\Psi_{k,n}$ as the ratio between $\Delta P_{k,n}$ and $\overline{\Delta r}$, i.e.

$$\Psi_{k,n} = \frac{\Delta P_{k,n}}{\overline{\Delta r}}. \tag{3.15}$$

From (3.15), a small value of the cost function $\Psi_{k,n}$ indicates high power efficiency. As higher power efficiency is achieved when a user experiences better channel quality, the proposed algorithm tends to assign resources to the users with good channel qualities at the beginning. This implies that much more power will then be required for users with relatively poor channel due to less available resources, which consequently increase the total transmit power. A heuristic solution against this problem is to prioritize the users that currently have been provided with less resources or at less data rate. Assume that the algorithm keeps track of the relative data rate

$$B_k = \frac{\sum_{n=1}^{N_c} r_{k,n}}{R_k}, \tag{3.16}$$

after each iteration, which is a measure of how well the user has met its data rate requirement R_k. Users with low value B_k shall be prioritized in the allocation. Two different proposals for enabling such prioritization are presented in the following.

The first proposal is to modify the cost function as

$$\tilde{\Psi}_{k,n} = B_k \cdot \frac{\Delta P_{k,n}}{\overline{\Delta r}}. \tag{3.17}$$

By weighting the power efficiency with the relative allocated data rate, the cost function $\tilde{\Psi}_{k,n}$ in (3.17) states that either lower relative allocated data rate or higher power efficiency indicates lower cost. The resulting algorithm is thus referred to as weighted priority variant of the SBI algorithm, denoted with SBI-WP, to distinguish from that based on the cost function $\Psi_{k,n}$, which is denoted with SBI-Original hereafter.

Alternatively, user prioritization can be enabled by modifying the greedy algorithm as follows. In each iteration, instead of finding the optimum pair of user and chunk as described in (3.13), firstly the user k^* having minimum relative rate B_k in (3.16) is found, and then for the user k^* the chunk n^* yielding the highest power efficiency, i.e. minimizing the cost function $\Psi_{k^*,n}$ in (3.15), is selected, i.e.

$$\begin{aligned} k^* &= \arg\min_k B_k, \\ n^* &= \arg\min_n \Psi_{k^*,n}. \end{aligned} \tag{3.18}$$

In this way, it is guaranteed in each iteration that the user with the lowest relative data rate has the option to increase the data rate. The resulting algorithm is referred to as first priority variant of the SBI algorithm, denoted with SBI-FP, as it firstly considers user priority instead of power efficiency.

3.2.5 Performance Assessment

The performance of the proposed sub-optimal algorithms is evaluated and compared to the optimal solution in this section.

To evaluate the performance achieved by the proposed algorithms for the power minimization problem, one BS sector with 16 users is simulated, and the inter-cell interference is not considered. All users are uniformly distributed in the sector and have the same target data rate. The numerical results are reported in terms of normalized total transmit power as function of the target user data rate. The normalized total transmit power is defined as the total required transmit power normalized with respect to the maximum available power, e.g. 46 dBm for a BS as given in Table 2.1.

Moreover, the average SNR of a user is defined as the average receive SNR while taking only slow-fading into account and assuming that the total available transmit power is equally distributed over the whole bandwidth. Two different scenarios are considered in the simulations. In the first scenario, it is assumed that the average SNR of all users is identically set to a certain pre-defined value. In the second scenario, since the slow-fading including path-loss and shadow-fading is considered and properly modeled, cf. (2.3), all

users experience different channel quality reflected by different average SNR. An experimental complementary cumulative distribution function (CCDF) of the average SNR for 16 users in 16 realizations is depicted in Figure 3.1. As the user priority is more important for
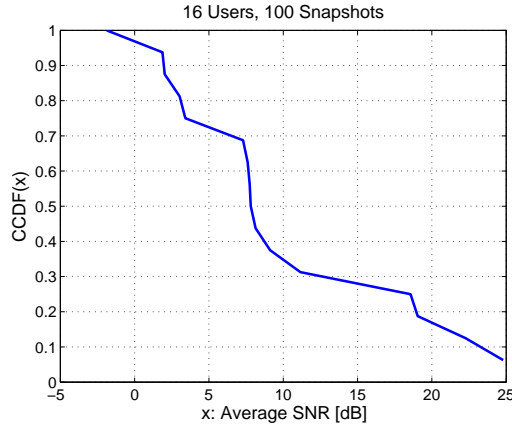
Figure 3.1: CCDF of the average SNR, $K = 16$, 100 independent realizations.

the scenario in which users have varied channel quality, the performance difference among the variants of the proposed greedy algorithms is expected to be higher in the latter scenario than that in the former one.

The performance achieved by the three variants of the proposed algorithm, i.e. SBI-Original, SBI-WP and SBI-FP is compared in Figure 3.2(a) and in Figure 3.2(b) for the two scenarios (a) and (b). In scenario (a), all users have an identical value of average

(a) Identical average SNR of $10\,\text{dB}$

(b) Varied average SNR

Figure 3.2: Comparison of the different variants of the proposed sub-optimal algorithm for the power minimization problem in terms of normalized total required transmit power required, 16 users.

SNR equal to 10 dB, and in scenario (b), all users have different values of average SNR according to the properly modeled slow-fading. As stated in Section 3.2.4, by considering the relative allocated user data rate defined in (3.16) in each iteration, users with poor channel condition is prioritized in resour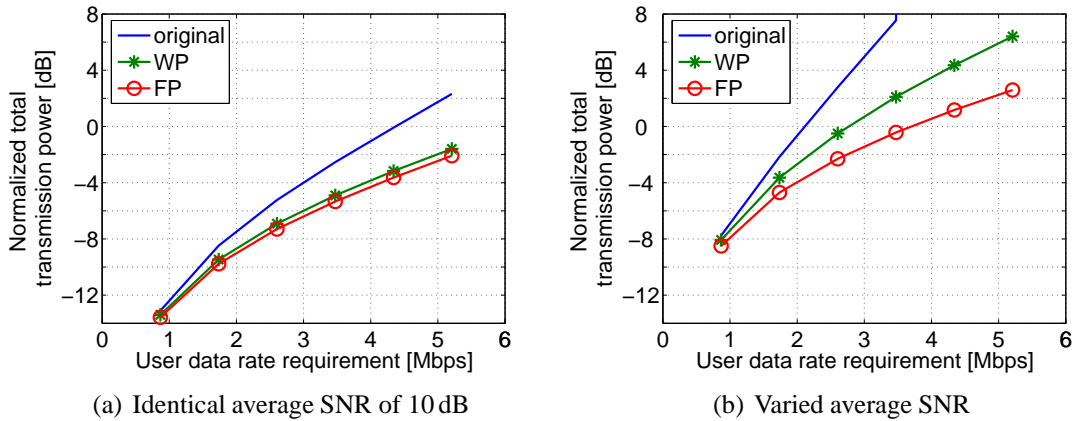ce allocation and ultimately the required total transmit power can be reduced. Hence, as expected, the SBI-WP (solid curve with stars) and the SBI-FP (solid curve with circles) variants perform better, i.e. require less transmit power, than the original (solid curve). The SBI-FP variant, in which user priority is firstly guaranteed, yields the best sub-optimal solution. Furthermore, by comparing Figure 3.2(a) and Figure 3.2(b), it can be inferred that, as expected, the performance gain by introducing user priority is more significant when users experience different channel qualities.

The optimal solution provides a very good reference against which to evaluate the degradation of the sub-optimal solutions. As introduced in Section 3.2.3, the optimal solution can be found by converting the original power minimization problem into an ILP. However, the computational complexity of an ILP is still quite high in case of large number of users and chunks. In order to compare results achieved with the proposed sub-optimal solutions against the optimal one in reasonable time, a relative small scale system is considered in simulations. 8 users share 32 chunks, the BS is equipped with 2 antennas and each UT has a single antenna.
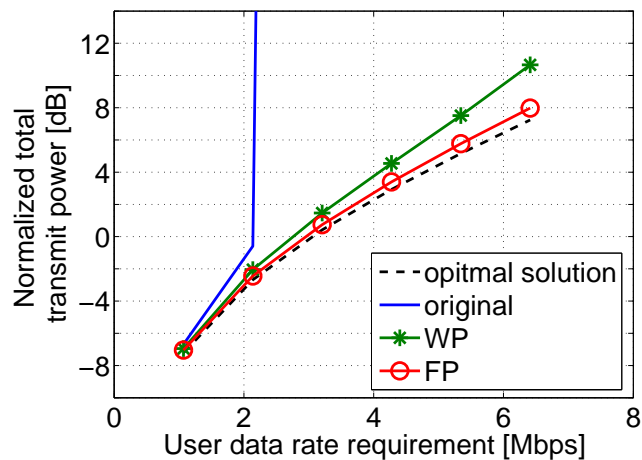
Figure 3.3: Comparison of the different variants of the proposed sub-optimal algorithm w.r.t. the optimal solution for the power minimization problem in terms of normalized total required transmit power, identical average SNR of 10 dB among users, 8 users, reduced system size with 32 chunks, two antennas at BS, and a single antenna at each UT.

In Figure 3.3, the optimal solution obtained by the ILP solver and the sub-optimal solutions obtained by means of SBI-Original, SBI-WP and SBI-FP (solid curve with circles)

are reported. The solution obtained by the SBI-FP variants (solid curve with circles) is not only better than those of the SBI-Original (solid curve) and the SBI-WP (solid curve with stars), but it also approaches very closely to the optimal one (dashed curve). Hence, it could be concluded that the SBI-FP variant is able to deliver a near-optimal solution for the power minimization problem.

3.3 Rate Maximization Problem

3.3.1 Introduction

This section studies the adaptive resource allocation targeting at the rate maximization under a certain transmit power constraint. The problem statement of the rate maximization problem is given in Section 3.3.2, where a mathematical description of the user fairness is also introduced. Sub-optimal algorithms are then proposed in Section 3.3.3. Finally a performance assessment of the proposed algorithms is presented in Section 3.2.5.

3.3.2 Problem Statement

In this section, the rate maximization problem is introduced and mathematically formulated.

In general, the resource allocation in terms of rate maximization aims at maximizing the cell throughput, i.e. the sum of the user throughput, while satisfying certain user fairness properties, under a transmit power constraint. Let T_k denote the average data rate provided to user k. Being tracked with a historical time window of length t_c in unit of frames, the average data rate in frame t, denoted with $T_k(t)$, is computed after resource allocation as

$$T_k(t) = \left(1 - \frac{1}{t_c}\right) T_k(t-1) + \frac{1}{t_c} r_k(t), \tag{3.19}$$

with

$$r_k(t) = \sum_{n=1}^{N_c} r_{k,n}(t), \tag{3.20}$$

representing the overall data rate granted to user k over all N_c chunks in frame t [VTL02].

According to Section 2.8, the total available transmit power P_{tot} is assumed to be equally distributed over all N_c chunks, i.e. PCPC. By letting

$$P_{\text{chunk}} = \frac{P_{\text{tot}}}{N_c} \tag{3.21}$$

denote the power available for each chunk, the PCPC is formulated as

$$\sum_{k=1}^{K} P_{k,n} \leq P_{\text{chunk}}, \forall n = 1, \ldots, N_c. \tag{3.22}$$

In case that the AP is equipped with multiple antennas, multiple users can share the same chunk by transmitting over individual chunk layers, and the power available for a chunk is assumed so far to be adaptively distributed among its chunk layers.

However, in some situations, it might be desirable to further assume equal power sharing among chunk layers. In other words, users served on the same chunk will equally share the power on that chunk, i.e. PUPC By letting $\mathcal{U}_n$ denote the set of users served on chunk n, the PUPC is formulated as

$$P_{k,n} = \begin{cases} P_{\text{chunk}}/|\mathcal{U}_n|, & k \in \mathcal{U}_n \\ 0, & k \notin \mathcal{U}_n \end{cases}, \tag{3.23}$$

with $|\cdot|$ representing the cardinality of a set.

Depending on the degree of user fairness consideration, the rate maximization problem can aim at the maximization of different functions ranging from the overall cell throughput to the minimum user throughput.

The user fairness is commonly measured by Jain's index. In case of K users competing for some resource x, the Jain's index $J(x)$ is defined as

$$J(x) = \frac{\left(\sum_{k=1}^{K} x_k\right)^2}{K \left(\sum_{k=1}^{K} x_k^2\right)} \tag{3.24}$$

[JCH84]. It is equal to 1 when all users get the same amount of resource and tends to $1/K$ in the fully unfair case. In the context of the considered optimization problem, x_k in (3.24) is the average data rate of each user T_k.

For overall throughput maximization, the user with highest achievable rate should always be selected. As it is potentially completely unfair, e.g. no resource will be provided to cell edge users with very poor channel conditions [KMT98], maximizing the overall throughput is not considered further in this work. In the following, two user fairness strategies, namely, PF and MMF, are discussed in detail.

Rate Maximization with PF strategy

PF has been well-studied in the literature , where the resource allocation is usually formulated as an optimization problem to maximize a certain utility function given the constraints

on resources [KMT98, KST04, SV98, VTL02]. A vector of rates for all users is said to be proportionally fair if it is feasible and the aggregate of proportional changes between it and any other vector of feasible rates is non-positive [KMT98]. By denoting with $\{T_k^*\}_{k=1}^K$ the rates achieving PF and with $\{T_k\}_{k=1}^K$ any arbitrary feasible data rates, the following inequality is satisfied,

$$\sum_{k=1}^K \frac{T_k - T_k^*}{T_k^*} \leq 0. \tag{3.25}$$

In other words, any positive change of performance for a user must result in a negative average change for the system performance. It is well known that such criterion is equivalent to the maximization of the sum of logarithmic data rates [KMT98]. Thus, the rate maximization problem with PF strategy can be mathematically formulated as

$$\max \sum_{k=1}^K \log T_k \tag{3.26}$$

subject to a certain power constraint.

In TDMA systems, resource allocation with PF strategy turns out to allocate the current time slot to the user with the largest ratio between the achievable rate in the current time slot and the past average data rate, r_k/T_k [BBG$^+$00]. PF offers a compromise between user fairness and overall throughput. Multi-user diversity is well exploited, as the time slots corresponding to peaks of the achievable data rate are always assigned to each user [Bon04].

Extensions of PF from TDMA systems to systems in which more than one resource units are available in each time slot, such as multi-carrier systems, have been discussed in literature, e.g. [ALS$^+$03, KH05, PJKL03]. In [WG00], several heuristic approaches for proportional fairness in OFDMA systems have been proposed and compared. In [KH05], it is proven that in a system with K users and multiple resource units per time slot, the resource allocation problem with PF strategy described by (3.26) can be equivalently formulated as

$$\max \prod_{k=1}^K \left(1 + \frac{r_k}{(t_c - 1)T_k} \right) \tag{3.27}$$

subject to a certain power constraint.

Rate Maximization with MMF strategy

Different from PF strategy, MMF is formulated as maximizing the minimum user average data rate [RC00], which can be formulated as

$$\max \min_{k=1,...,K} T_k \tag{3.28}$$

subject to a certain transmit power constraint. By taking MMF strategy, it tries to provide all users with the same data rate.

3.3.3 Sub-optimal Algorithms

Similar to the power minimization problem discussed in Section 3.2.4, the optimal solution for the rate maximization problem described in Section 3.3.2 requires too high computational complexity to be affordable in practical systems and hence sub-optimal greedy algorithms are of interest. In the following, the sub-optimal algorithms to solve the rate maximization problems under the two power constraints, PCPC and PUPC, are separately discussed. The relative variants of the algorithm are derived for both PF and MMF user fairness strategies.

Sub-optimal Algorithms under PCPC

The greedy algorithm SBI proposed in Section 3.2.4 for the power minimization problem can be easily adapted to solve the rate maximization problem as described in the following. The greedy algorithm starts with the assignment to all K users of a null rate on all N_c available chunks, cf. (3.11). In each iteration, for each user and chunk a certain cost function $\Psi_{k,n}$ is evaluated and only the rate of the user experiencing the minimum cost function is increased on the corresponding chunk, cf. (3.13) and (3.14). The algorithm ends when no more data increase is possible under the given transmit power constraint.

The cost function $\Psi_{k,n}$ is defined as the power efficiency in terms of the ratio between required power increase $\Delta P_{k,n}$ and performance increase $\Delta\Gamma_{k,n}$ for a given rate increase $\overline{\Delta r}$ of user k on chunk n,

$$\Psi_{k,n}^{\text{SBI}} = \frac{\Delta P_{k,n}}{\Delta\Gamma_{k,n}}. \tag{3.29}$$

From (3.27), the performance to be maximized in the rate maximization problem with PF strategy is expressed as

$$\Gamma = \prod_{k=1}^{K}\left(1 + \frac{r_k}{(t_c - 1)T_k}\right). \tag{3.30}$$

For a certain rate increase $\overline{\Delta r}$ of user k on chunk n, the corresponding performance increase $\Delta\Gamma_{k,n}$ in the i-th iteration, defined as the increase compared to the last iteration, is given by

$$\begin{aligned}
\Delta\Gamma_{k,n}^{(i)} &= \prod_{j=1 j\neq k}^{K}\left(1 + \frac{r_j^{(i-1)}}{(t_c-1)T_j}\right)\cdot\left(1 + \frac{\overline{\Delta r}+r_k^{(i-1)}}{(t_c-1)T_k}\right)\cdots \\
&\quad - \prod_{j=1}^{K}\left(1 + \frac{r_j^{(i-1)}}{(t_c-1)T_j}\right)
\end{aligned} \tag{3.31}$$

As the time average window t_c and the performance in the $(i-1)$-th iteration

$$\Gamma^{(i-1)} = \prod_{j=1}^{K} \left(1 + \frac{r_j^{(i-1)}}{(t_c - 1)T_j} \right) \tag{3.32}$$

are constant, maximizing the performance increase $\Delta\Gamma_{k,n}$ is equivalent to maximizing

$$\Delta\Gamma'_{k,n} = \frac{t_c}{\Gamma^{(i-1)}} \cdot \Delta\Gamma_{k,n}. \tag{3.33}$$

After several simple derivations, $\Delta\Gamma'_{k,n}$ becomes

$$\Delta\Gamma'_{k,n} = t_c \cdot \left(\frac{1 + \frac{\overline{\Delta r} + r_k^{(i-1)}}{(t_c - 1)T_k}}{1 + \frac{r_k^{(i-1)}}{(t_c - 1)T_k}} - 1 \right) = \frac{\overline{\Delta r}}{(1 - \frac{1}{t_c})T_k + \frac{1}{t_c}r_k^{(i-1)}}. \tag{3.34}$$

The denominator of (3.34)

$$T'_k(i) = (1 - \frac{1}{t_c})T_k + \frac{1}{t_c}r_k^{(i-1)}, \tag{3.35}$$

can be viewed as the average data rate T_k defined in (3.19) being updated in each iteration. Thus, the performance increase $\Delta\Gamma'_{k,n}$ is the ratio between the rate increase $\overline{\Delta r}$ and the average data rate T'_k in (3.35), i.e.

$$\Delta\Gamma'_{k,n} = \frac{\overline{\Delta r}}{T'_k(i)}. \tag{3.36}$$

By combing (3.13) and (3.36), the cost function for PF is given by

$$\Psi_{k,n}^{\text{SBI}} = \arg \min_{(k,n)} T'_k \cdot \frac{\Delta P_{k,n}}{\overline{\Delta r}}. \tag{3.37}$$

From (3.37), it can be inferred that the variant of the greedy algorithm for PF strategy compromises power efficiency and user priority (or fairness) by weighting the rate increase $\overline{\Delta r}$ with the average data rate T'_k. Hence, it is hereafter called weighted priority (WP) variant.

For the rate maximization problem with MMF strategy, as described in (3.28), the minimum user data rate shall be maximized. In each iteration of the greedy algorithm, the user with the minimum average data rate is selected, i.e. in the i-th iteration

$$k^* = \arg \min_{k=1,\ldots,K} T'_k(i), \tag{3.38}$$

where $T'_k(i)$ defined in (3.19) is the average data rate of user k which considers the allocation until the previous iteration $(i-1)$. For the selected user k^*, the chunk which requires the least power increase per bit increase is then selected, i.e.

$$n^* = \arg \min_{n=1,\dots,N_c} \frac{\Delta P_{k^*,n}}{\Delta r}. \tag{3.39}$$

Since the user priority in terms of the achieved average data rate is firstly compared and the data rate of the user with highest priority, i.e. with the lowest average data rate, is increased in each iteration, this variants proposed to achieve MMF is called first priority (FP) variant hereafter.

Future systems are supposed to provide users with different types of data services with variable rate. Hence, in most cases users will not have homogeneous requirement on average data rate. By letting $\{R_k\}_{k=1}^K$ denote the set of data rate requirements for all K users, the relative average data rate can be defined as

$$\tilde{B}_k(i) = \frac{T'_k(i)}{R_k}. \tag{3.40}$$

The relative average data rate $\tilde{B}_k(i)$ provides a measure of how well the user k has met its data rate requirement and it can be used in (3.38) to prioritize users instead the absolute average data rate T'_k. Moreover, when a specific traffic model is considered, user priority can be also defined according to other QoS parameters, e.g. the average response time for delay-sensitive services [BDL$^+$02] .

Sub-optimal Algorithms under PUPC

In some situations, it might be desirable to further assume equal power sharing among users served on the same chunk, i.e. to maximize rate under PUPC defined in (3.23), cf. Section 3.3.2. In this case, the proposed greedy algorithm is modified as follows. The algorithm starts by setting $\mathcal{U}_n^{(0)} = \varnothing$ for all n. In the i-th iteration, the user and chunk pair (k^*, n^*) that minimizes certain cost function $\Psi_{k,n}$ is selected and the chunk n^* is assigned to the user k^*, i.e. in each iteration,

$$\mathcal{U}_{n^*}^{(i)} = \mathcal{U}_{n^*}^{(i-1)} \cup \{k^*\}. \tag{3.41}$$

with

$$(k^*, n^*) = \arg \min \Psi_{k,n}. \tag{3.42}$$

Note that selecting user k for chunk n is not considered, if the performance would be decreased by adding one more user. The algorithm ends when the performance of all chunks can not be increased by adding one more user. Different from SBI, here users are iteratively added on chunks. Hereafter, it is referred to as successive user insertion (SUI).

Let $r'(j,n)$ represent the data rate with a power equal to $P_{\text{chunk}}/(|\mathcal{U}_n^{(i-1)}|+1)$ and $\alpha_j(i)$ indicate the user priority of user j in the current i-th iteration. Analog to the cost function $\Psi_{k,n}^{\text{SBI}}$ in (3.37) defined for SBI with PF, the cost function for SUI with PF is given by

$$\Psi_{k,n}^{\text{SUI}} = \frac{P_{\text{chunk}}}{\frac{r'(k,n)}{\alpha_k(i)} \sum_{j\in\mathcal{U}_n^{(i-1)}} \frac{r'(j,n)}{\alpha_j(i)}}. \tag{3.43}$$

The user priority $\alpha_j(i)$ can be the average data rate $T_k'(i)$ in (3.35) or the relative average data rate $B_k'(i)$ in (3.40) in the case that users have different data rate requirements.

Different from PF strategy, the minimum user data rate shall be maximized by following MMF strategy. Hence, to solve the rate maximization problem with MMF strategy, in each iteration of the greedy algorithm, the user k^* with highest priority is firstly selected, i.e.

$$k^* = \arg\min_{k=1,\ldots,K} \alpha_k(i). \tag{3.44}$$

The chunk n^* is then chosen in such a way that the cost function in (3.43) is minimized, i.e.

$$n^* = \arg\min_{n=1,\ldots,N_c} \Psi_{k^*,n}^{\text{SUI}}. \tag{3.45}$$

Complexity Analysis

In the following, the computational complexity of all the proposed algorithms, as listed in Table 3.2, is evaluated and compared. The complexity of the considered resource allocation comes from two folds: one is the computation of the effective channel gain according to (2.41); the other is the complexity of the greedy algorithm, which can be measured by the number of iterations and the number of comparisons required for each iteration.

The computation of the effective channel gain is required by all algorithms, which is not trivial due to the involved matrix inversion, as seen in (2.41). However, if a certain user is spatially highly correlated with users already in the chunk, a low effective channel gain can be anticipated without matrix inversion. Therefore, in order to avoid such an unnecessary computation effort, a user pre-selection procedure can be applied to avoid spatially incompatible users, whose spatial correlation is higher than certain threshold, being served on the same chunk [JG04]. As shown in [JG04], a too low threshold will decrease the multi-user diversity gain due to a too small candidate set. An appropriate threshold can be determined through numeric simulations.

SBI and SUI differ in the number of iterations. In case of SBI, the data rate of a selected user on a selected chunk is increased step by step. Given that totally there are Q feasible data rates available and maximal M_t users can be served on each of the N_c chunks, the maximum number of iterations is in the order of QM_tN_c. In case of SUI, since one user is added on the selected chunk in each iteration, the maximum number of iterations is in the

order of $M_t N_c$. Therefore, compared to SBI, SUI has lower complexity due to the lower number of iterations. However, it is expected that SBI will outperform SUI as it benefits from adaptive power allocation.

The variants proposed for PF and MMF differ from each other in the number of required comparisons in each iteration. In the WP greedy algorithm proposed for PF, the user and the chunk are simultaneously selected from all KN_c combinations, i.e. the number of comparison per iteration is about KN_c, while in the FP greedy algorithm proposed for MMF the user and the chunk are sequentially selected, thus requiring $(K + N_c)$ comparisons.

3.3.4 Performance Assessment

In this section, the performance assessment of the proposed sub-optimal algorithms is presented.

As already mentioned in Section 3.1, the proposed greedy sub-optimal algorithms are compared with the disjoint (DJ) approach presented in [Wil06], and so the disjoint approach is briefly reviewed in the following. The DJ approach evaluates one chunk after anther in an arbitrary order. After a chunk is selected, adaptive SDMA is carried out by starting from an empty set $\mathcal{U}_{n*}^{(0)} = \varnothing$ and successively adding users until the performance can not be increased. In order to achieve PF, in each iteration the user k^* that minimizes the cost function defined in (3.43) is selected, referred to as SUI-DJ-WP hereafter. In order to achieve MMF, in each iteration the user with the highest priority, i.e. with the minimum average data rate is selected, referred to as SUI-DJ-FP hereafter.

One BS sector with 16 users is simulated and statistical results are collected over 100 independent simulations. The positions of the 16 users are randomly generated for each simulation according to uniform distribution over the sector. The performance is reported in terms of average user throughput, 95-percentile user throughput and Jain's Index. The average user throughput is proportional to the cell throughput. The 95-percentile user throughput is very important in the definition of user satisfaction [IST06], as it indicates that 95% users are served with such an average data rate or higher. The Jain's Index in (3.24) ranges from $1/K$ to 1, corresponding to the completely unfair and the fully fair, respectively, cf. Section 3.3.2.

In Table 3.3, the performance of the proposed joint approach as well as the DJ approach is reported.

In the following, the two variants of the DJ approach, DJ-WP and DJ-FP, are firstly compared with the corresponding variants of the joint approach, SUI-WP and SUI-FP. The WP and the FP algorithms are then compared so as to illustrate the impact of user fairness on the resulting allocation. Finally, the two variants, SBI and SUI, are compared in order to quantify the performance gain achieved by adaptive power allocation among users added

Table 3.3: Comparison of the different variants of the sub-optimal algorithm for the rate maximization problem in terms of average user throughput, 95%-tile and Jain's index.

		Average [Mbps]	95%-tile [Mbps]	Jain's Index
SBI	WP	7.99	0.43	0.41
	FP	3.85	2.70	0.98
SUI	WP	5.47	1.15	0.76
	FP	3.14	1.90	0.94
	DJ-WP	5.34	1.10	0.75
	DJ-FP	2.22	1.48	0.96

on the same chunk. In addition to the statistical performance metrics provided in Table 3.3, the corresponding CCDF of user throughput is also reported for each comparison. The CCDF curves can provide a detailed description on the distribution of user throughput.

Comparison between Disjoint and Joint Approaches

In Figure 3.4, the proposed joint approaches are compared with the corresponding disjoint approaches. For both WP (red curves) and FP (blue curves) variants, the proposed
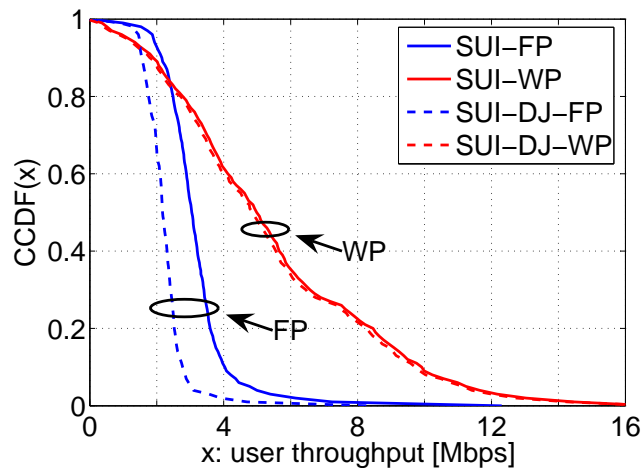

Figure 3.4: Comparison between joint and DJ approaches.

joint approach achieves higher throughput than the DJ one everywhere in the CCDF. The performance loss of SUI-DJ-FP (blue dashed curve) is more significant than that of SUI-DJ-WP (red dashed curve). In order to guarantee MMF, SUI-DJ-FP actually selects in each iteration the chunk and user without considering the channel conditions, therefore, the multi-user diversity gain can not be exploited, which results in significant performance loss compared to the joint approach.

Comparison between PF and MMF User Fairness Strategies

In Figure 3.5, the CCDF curves of the user throughput achieved by SUI-FP and SUI-WP are plotted. By following MMF strategy, the FP variant (red dashed curve) actually sacrifices
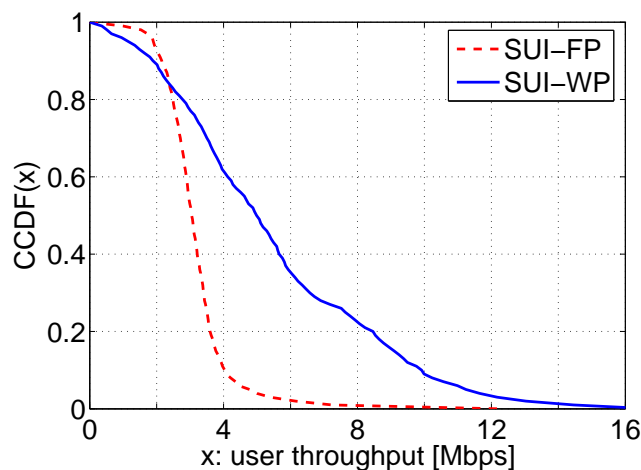
Figure 3.5: Comparison between WP and FP, SUI.

the cell throughput in return for fairness. As expected, the average user throughput of FP is less than that of WP, but the Jain's index of FP is much closer to 1 than that of WP, indicating better fairness, as shown in Table 3.3. Moreover, FP achieves much higher 95-percentile user throughput, which is consistent with the target of the MMF strategy.

Comparison between Adaptive and Equal Power Sharing

In Figure 3.6, the performance of SBI-FP and SUI-FP is compared. SBI enables adaptive power allocation among the co-located users, while SUI supposes equal power sharing among co-located users. It can be inferred that under the same fairness strategy much higher user throughput is achievable by adaptive power allocation (blue solid curve) compared to that by fixed equal power sharing (red dashed curve).
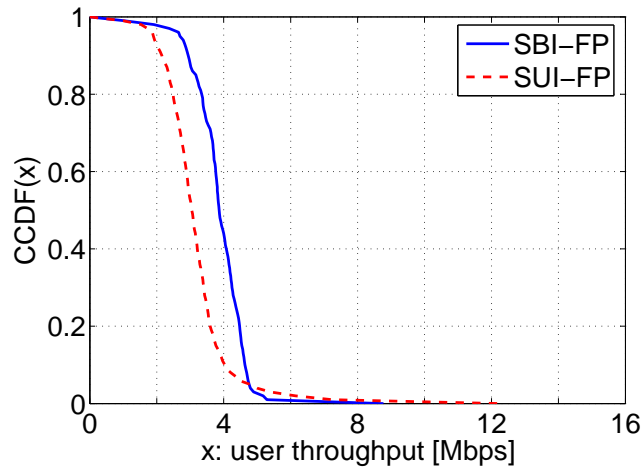
Figure 3.6: Comparison between SBI and SUI, FP.

Impact of the AMC Scheme

The AMC scheme applied in the simulations presented so far is AMC-Baseline, cf. Section 2.7. However, as seen in Figure 2.9, AMC-Fine achieves higher data rate than AMC-Baseline over the whole SNR region due to more feasible data rates. In Figure 3.7, the performance achieved by AMC-Baseline and AMC-Fine is compared for the variants, namely SUI-FP and SBI-FP. As expected, higher data rate is achieved by AMC-Fine (blue dashed
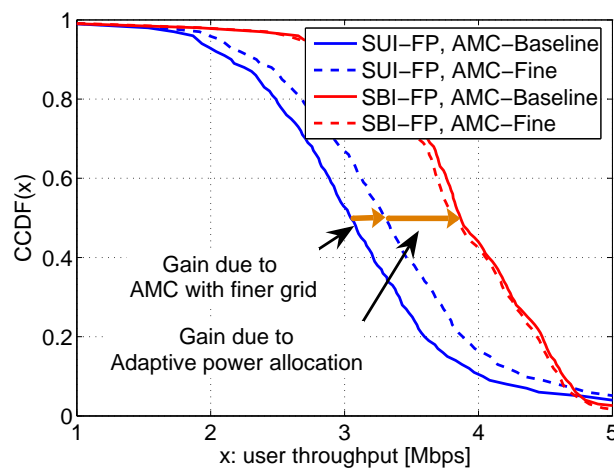
Figure 3.7: Comparison between AMC-Baseline and AMC-Fine.

curve) with respect to AMC-Baseline (blue solid curve) when SUI is applied. However,

when SBI is applied, the performance achieved by adopting AMC-Fine (red dashed curve) and AMC-Baseline (red solid curve) is almost the same. This can be explained as follows. On the one hand, in SBI, adaptive power allocation is carried out by always minimizing the transmit power required for a target data rate. On the other hand, as shown in Figure 2.9, given the same target data rate, AMC-Fine requires the same amount of power as AMC-Baseline. The performance loss of SUI-FP compared to SBI-FP when applying AMC-Baseline, i.e. the difference between solid blue curve and solid red curve, comes from two folds: the less feasible data rates in AMC-Baseline and the equal power sharing assumption. While the performance loss caused by the less feasible data rates in AMC-Baseline is quantified by comparing the performance of SUI-FP with AMC-Baseline and SUI-FP with AMC-Fine, i.e. the gap between the blue solid curve and the blue dashed curve, the performance loss due to the equal power sharing is then quantified by the gap between the blue dashed curve and the red solid curve, as depicted in Figure 3.7.

As summary, following conclusions can be drawn based on the presented results:

- The proposed joint approach generally outperforms the existing DJ approach, especially in the FP variant targeting at max-min fairness.

- Compared to the WP variant targeting at proportional fairness, the FP variant targeting at max-min fairness sacrifices overall cell throughput for better user fairness.

- Compared to SUI, SBI benefits from adaptive power allocation in the spatial domain and thus achieves higher user throughput.

- Under the assumption of equal power sharing, significant performance increase can be obtained by using an AMC scheme with finer grid, i.e. with more feasible data rates.

- Under the assumption of adaptive power allocation in the spatial domain, almost no performance increase can be obtained by using an AMC scheme with finer grid.

4 Signaling Overhead for Adaptive Resource Allocation

4.1 Introduction

Resource allocation plays an important roll in optimizing the system performance. As stated in Section 2.3, the channel fading varies over frequency band due to multi-path propagation and over time due to mobility. Since different users perceive different channel qualities due to independent fading, a resource with deep fading for one user may still be favorable for the others, referred to as selective channel diversity. Moreover, in cellular systems, the variability of the channel quality in terms of SINR comes from the fading of the signal as well as the fading of the interference from adjacent APs, which is referred to as mutual interference diversity. Besides, data service is generally characterized as burst traffic and the arrival time of data service is also different among users, which leads to independent varying of current user data rate requirements, referred to as traffic diversity. Thus, resource allocation adaptive to the traffic status can potentially outperform resource allocation with certain fixed bandwidth assignment. In summary, three kinds of multi-user diversities can be identified as selective fading channel diversity, mutual interference diversity and traffic diversity [LL03]. However, any AP attempting to exploit these multi-user diversities would require the knowledge of both channel and traffic conditions, which may cause additional signaling overhead. In general, high multi-user diversity gain can be obtained with precise channel and traffic knowledge, but from the perspective of system performance, the high signaling overhead due to acquisition of precise information will mitigate the achieved performance gain. Therefore, it is important to balance the signaling overhead and the performance gain achieved by adaptive resource allocation.

In this chapter, the following aspects concerning the signaling for adaptive resource allocation are addressed.

- Different sub-carriers undergo varied channel conditions, and so adaptive modulation and coding improves spectrum efficiency by selecting for each sub-carrier a suitable MCS with respect to its SINR value, cf. Section 2.7. Thus, additional downlink control data is required to inform users about the MCS selection for each sub-carrier. This amount of control data can be reduced by choosing the same MCS for a chunk

consisting of adjacent sub-carriers and symbols at the expense of certain performance degradation. In Section 4.2, the trade-off between the signaling reduction and the performance degradation is illustrated by analytical derivations, based on which a proper chunk dimension can be identified.

- Due to mobility, the channel varies over time and so the channel knowledge shall be measured and/or delivered periodically. In Section 4.3, a semi-analytical method is presented to find the optimum update interval for the channel knowledge, with which the overall system performance is maximized.

- Adaptive resource allocation presented in Chapter 3 is based on the assumption that full CSI is available at the AP and it achieves significant performance gain but also requires considerable channel feedback. Indeed, SDMA can be enabled by other beamforming strategies based on partial CSI. In Section 4.4, two kinds of beamforming strategies requiring reduced channel feedback are address and the corresponding performance degradation is illustrated.

- The AP needs the information about traffic conditions such as the amount of data to be transmitted for the optimization of adaptive resource allocation. In downlink, the AP, as the transmitter, knows the traffic conditions. However, in uplink, the AP can only know the traffic conditions via signaling from users. In Section 4.5, the uplink bandwidth request mechanism, which allows users to deliver the traffic condition such as data rate requirements to the AP, is discussed.

4.2 Optimization of Chunk Dimension

4.2.1 Introduction

In OFDM systems, the channel fading can be regarded as being flat on each sub-carrier but varying among sub-carriers due to frequency selective fading. In multi-user environments, as each user experiences a different fading, it is likely that while one sub-carrier is in deep fade for a particular user, it is in good condition for another. Thus, system performance can be significantly improved by adaptive OFDMA, i.e. dynamically allocating sub-carriers to users. Moreover, the spectrum efficiency can be further enhanced by means of AMC, i.e. employing different MCSs on individual sub-carriers with different SINR levels. Thus, additional downlink control data is required to inform all users about the allocation results, namely indicating for each sub-carrier which user is served and which transmission mode, i.e. MCS, is selected, which accounts for further additional overhead. For a system with N sub-carriers, K users and Q MCSs, the amount of such downlink control data for each OFDM symbol is given by

$$\xi = N \cdot (\lceil \log_2 K \rceil + \lceil \log_2 Q \rceil), \tag{4.1}$$

which is quite considerable as the number N of sub-carriers is usually very large, e.g. equal to 1024 for the considered system as given in Table 2.2.

Given the typical channel coherence bandwidth with respect to the sub-carrier spacing, it is observed that the user allocation as well as the MCS assignment would be most probably the same for adjacent sub-carrier and OFDM symbols, since they experience similar channel fading. Therefore, as already introduced in Section 2.6, the chunk, defined as a block of adjacent sub-carriers and OFDM symbols, is chosen as basic resource unit. In other words, the resource allocation, consisting of both user allocation and MCS assignment, is identical for all sub-carriers and OFDM symbols belonging to the same chunk. By letting the chunk dimension be n_{sub} sub-carriers by n_{symb} OFDM symbols, the amount of the downlink control data can be reduced by a factor of $n_{\mathrm{sub}} \cdot n_{\mathrm{symb}}$.

When targeting at overhead reduction, a large chunk dimension, i.e. large values for parameters n_{sub} and n_{symb}, is desired. However, the performance gain achieved by adaptive resource allocation is expected to decrease with the chunk dimension. In the extreme case when choosing n_{sub} as the total number of sub-carriers N, all sub-carriers are assigned to the same user using the same MCS, therefore no adaptation gain can be achieved. Therefore, the optimization of the chunk dimension requires the trade-off between overhead reduction and performance degradation. In Section 4.2.2, the relationship between the performance degradation and the chunk dimension is analytically derived.

For the upcoming analytical investigation, the following assumptions are invoked:

- An OFDM system with one AP, K users and N sub-carriers is considered.

- The channel fading is modeled as a stationary two dimensional zero-mean Gaussian random process [Pro01]. Without loss of generality, the variance of the channel fading is set to one.

- Full channel knowledge is perfectly known at the AP.

- The system performance is evaluated in terms of Shannon capacity.

- In case of fixed allocation, all users are ordered and chunks are assigned to users in a round robin fashion. In case of adaptive resource allocation, each chunk is assigned to the user which achieves the highest capacity among all.

The remainder of the section is organized as follows. In Section 4.2.2, the system performance achieved by adaptive resource allocation is analytically derived as a function of the chunk dimension. In Section 4.2.3, the optimization of the chunk dimension is discussed.

4.2.2 Derivation of Performance of Chunkwise Adaptive Allocation

In this section, the performance achieved by adaptive resource allocation is analytically derived as a function of the chunk dimension.

By letting $h_{n,t}^{(k)}$ denote the channel coefficient experienced by user k on sub-carrier n of OFDM symbol t, P and σ_z^2 denote the transmit power and the noise power, respectively, the Shannon capacity of user k on sub-carrier n of OFDM symbol t is given by

$$C_{n,t}^{(k)} = \log_2\left(1 + |h_{n,t}^{(k)}|^2 \frac{P}{\sigma_z^2}\right) \quad \text{bits/s/Hz} \tag{4.2}$$

[Pro01].

Let $|h_{n,t}^{\mathrm{S}}|^2$ denote the channel gain on sub-carrier n of OFDM symbol t obtained with resource allocation scheme S. By knowing the PDF of the channel gain $p_{|h_{n,t}^{\mathrm{S}}|^2}(x)$ over all sub-carriers and OFDM symbols, the average capacity achieved by the resource allocation scheme S can be calculated as

$$\bar{C}_{\mathrm{tot}}^{\mathrm{S}} = \int_0^\infty \log_2\left(1 + x\frac{P}{\sigma_\eta^2}\right) \cdot p_{|h_{n,t}^{\mathrm{S}}|^2}(x)dx. \tag{4.3}$$

In case of fixed allocation, as the user assignment is independent of the channel conditions, the channel gain $|h_{n,t}^{\mathrm{fix}}|^2$ obtained with fixed allocation has the same distribution as the channel gain of any user $|h_{n,t}^{(k)}|^2$. As the channel response $h_{n,t}^{(k)}$ is a stationary two dimensional zero-mean Gaussian random process with variance of one, the envelope $|h_{n,t}^{(k)}|$ of the channel response follows the Rayleigh distribution [Pap65] and therefore, the PDF of the channel gain $|h_{n,t}^{(k)}|^2$ is given by

$$p_{|h_{n,t}^{\mathrm{fix}}|^2}(x) = p_{|h_{n,t}^{(k)}|^2}(x) = e^{-x} \; for \, x >= 0. \tag{4.4}$$

Thus, the average capacity over all sub-carriers and all OFDM symbols achieved by fixed allocation is equal to

$$\bar{C}_{\mathrm{tot}}^{\mathrm{fix}} = \int_0^\infty \log_2\left(1 + x\frac{P}{\sigma_z^2}\right) \cdot e^{-x}dx. \tag{4.5}$$

In case of adaptive resource allocation, chunk m is assigned to the user k_m who has the maximum capacity. By denoting with $C_m^{(k)}$ the capacity on chunk m achieved by user k, the chunk m is assigned to user k_m, who fulfills

$$k_m = \arg\max_k C_m^{(k)}. \tag{4.6}$$

Supposing chunk m is composed of n_{sub} sub-carriers starting with the n_m-th sub-carriers and n_{symb} OFDM symbols starting with the t_m-th OFDM symbol, the capacity achieved by user k on chunk m is the arithmetic mean of $n_{\mathrm{sub}}n_{\mathrm{symb}}$ samples of the sub-carrier capacity $C_{n,t}^{(k)}$, i.e.

$$C_m^{(k)} = \frac{\sum_{t=0}^{n_{\mathrm{symb}}-1} \sum_{n=0}^{n_{\mathrm{sub}}-1} C_{n+n_m,t_m+t}^{(k)}}{n_{\mathrm{sub}}n_{\mathrm{symb}}}. \tag{4.7}$$

When $n_{\mathrm{sub}}n_{\mathrm{symb}}$ is sufficiently large, the arithmetic mean can be well approximated by its expectation value [Pap65]. As a result, the capacity of user k over chunk m in (4.7) can be approximated by

$$C_m^{(k)} \approx \int_0^\infty \log_2\left(1 + x\frac{P}{\sigma_\eta^2}\right) \cdot p_{|h_m^{(k)}|^2}(x)dx, \tag{4.8}$$

with $p_{|h_m^{(k)}|^2}(x)$ being the PDF of the channel gain $|h_{n,t}^{(k)}|^2$ within chunk m.

By defining the mean value of the channel coefficient $h_{n,t}^{(k)}$ within chunk m

$$\bar{h}_m^{(k)} = \frac{\sum_{t=0}^{n_{\mathrm{symb}}-1} \sum_{n=0}^{n_{\mathrm{sub}}-1} h_{n+m \cdot n_{\mathrm{sub}},t}}{n_{\mathrm{sub}}n_{\mathrm{symb}}}, \tag{4.9}$$

as the local mean, the local variance of the channel coefficient $h_{n,t}^{(k)}$ within chunk m can be expressed as

$$\begin{aligned} \mathrm{var}_m^{(k)} &= E\left\{ \frac{\sum_{t=t_m}^{n_{\mathrm{symb}}-1} \sum_{n=0}^{n_{\mathrm{sub}}-1} |h_{n_m+n,t_m+t}|^2}{n_{\mathrm{sub}}n_{\mathrm{symb}}} - |\bar{h}_m^{(k)}|^2 \right\} \\ &= E\{|h_{n,t^{(k)}}|^2\} - E\{|\bar{h}_m|^2\}. \end{aligned} \tag{4.10}$$

By recalling that the variance of the channel coefficient $h_{n,t}^{(k)}$ is equal to 1 and letting

$$\Omega = E\{|\bar{h}_m^{(k)}|^2\}, \tag{4.11}$$

the local variance $\mathrm{var}_m^{(k)}$ in (4.10) is equal to

$$\mathrm{var}_m^{(k)} = 1 - \Omega. \tag{4.12}$$

From Appendix A.1, Ω in (4.11) can be calculated by integral of the auto-correlation function of the channel coefficient $h_{n,t}^{(k)}$ over the chunk dimension according to (A.6), i.e.

$$\begin{aligned} \Omega &= \frac{1}{B_{\mathrm{chunk}}} \int_{-B_{\mathrm{chunk}}}^{B_{\mathrm{chunk}}} \left(1 - \frac{|f|}{B_{\mathrm{chunk}}}\right) R_f(f)df \cdots \\ &\quad \cdot \frac{1}{T_{\mathrm{chunk}}} \int_{-T_{\mathrm{chunk}}}^{T_{\mathrm{chunk}}} \left(1 - \frac{|t|}{T_{\mathrm{chunk}}}\right) R_t(t)dt \end{aligned} \tag{4.13}$$

with $B_{\text{chunk}} = n_{\text{sub}}f_s$ and $T_{\text{chunk}} = n_{\text{symb}}T_s$ representing the bandwidth and the time duration of the chunk, respectively. Thus, within chunk m the channel coefficient $h_{n,t}^{(k)}$ can be viewed as a complex Gaussian random variable with non-zero mean value $\bar{h}_m^{(k)}$ and variance $(1-\Omega)$. Consequently, the channel gain $|h_m^{(k)}|^2$ with chunk m as the square value of a non-zero mean Gaussian random variable follows the noncentral chi-square distribution with two degrees of freedom [Pap65]. By letting

$$A_m^{(k)} = |\bar{h}_m^{(k)}|^2, \tag{4.14}$$

the PDF of the channel gain within chunk m, denoted with $p_{|h_m^{(k)}|^2}(x)$, is given by

$$p_{|h_m^{(k)}|^2}(x) = \frac{\exp(-\frac{A_m^{(k)}+x}{1-\Omega})}{1-\Omega} I_0(\sqrt{x}\frac{2\sqrt{A_m^{(k)}}}{1-\Omega}), \tag{4.15}$$

with $I_0(\cdot)$ representing the zero-order modified Bessel function of the first kind.

From (4.8) and (4.15), to assign the chunk to the user with maximum capacity $C_m^{(k)}$ is equivalent to assigning to the user with maximum $A_m^{(k)}$. Hence, the user served on chunk m in the adaptive resource allocation is given by

$$k_m = \arg\max_k C_m^{(k)} = \arg\max_k A_m^{(k)}. \tag{4.16}$$

By denoting with $A_m^{\max,K}$ the maximum $A_m^{(k)}$ of all K users, the PDF of the channel gain obtained with the adaptive resource allocation $|h_{n,t}^{(k_m)}|^2$ is

$$p_{|h_{n,t}^{\text{adapt}}|^2}(x) = p_{|h_{n,t}^{(k_m)}|^2}(x) = \frac{\exp(-\frac{A_m^{\max,K}+x}{1-\Omega})}{1-\Omega} I_0(\sqrt{x}\frac{2\sqrt{A_m^{\max,K}}}{1-\Omega}). \tag{4.17}$$

As a block average of a Gaussian process, the local mean value $\bar{h}_m^{(k)}$ is Gaussian distributed with mean zero and variance Ω [Pap65], and so the PDF and CDF of its square value $A_m^{(k)}$ are given by

$$p_{A_m}(x) = \frac{1}{\Omega}e^{-x/\Omega} \tag{4.18}$$

and

$$F_{A_m}(x) = 1 - e^{-x/\Omega}, \tag{4.19}$$

respectively. Since all K users are supposed to experience independent channel fading, according to order statistic [Wei], the PDF of $A_m^{\max,K}$ can be expressed as

$$\begin{aligned} p_{A_m^{\max,K}}(x) &= K\left(F_{A_m}(x)\right)^{K-1} p_{A_m}(x) \\ &= \frac{K}{\Omega}e^{-x/\Omega}\left(1 - e^{-x/\Omega}\right)^{K-1}. \end{aligned} \tag{4.20}$$

By substituting (4.20) in (4.17) and then (4.17) in (4.3), the average capacity achieved by the adaptive resource allocation is

$$
\begin{aligned}
\bar{C}_{\text{tot}}^{\text{adapt}} \;=\; & \int_0^\infty da \left(\frac{K}{\Omega} e^{-a/\Omega} \left(1 - e^{-a/\Omega} \right)^{K-1} \cdots \right. \\
& \left. \cdot \int_x d_0^\infty \log_2 \left(1 + x\frac{P}{\sigma_\eta^2} \right) \cdot \frac{\exp(-\frac{a+x}{1-\Omega})}{1-\Omega} I_0 \big(\sqrt{x}\tfrac{2\sqrt{a}}{1-\Omega} \big) \right).
\end{aligned}
\tag{4.21}
$$

From (4.21), it can be inferred that the capacity achieved by the adaptive resource allocation $\bar{C}_{\text{tot}}^{\text{adapt}}$ increases with increasing A_m^{max} and decreases with increasing Ω. Moreover, it can be seen from (4.20) that $A_m^{\text{max},K}$ increases with the number K of users , and from (A.6) that the Ω decreases with increasing chunk dimension $n_{\text{sub}}n_{\text{symb}}$. Therefore, as expected, the multi-user diversity gain increases with the number of users and decreases with the increasing chunk dimension.

Moreover, the Ω is calculated by the integral of the auto-correlation function of the channel coefficient $h_{n,t}^{(k)}$, cf. (A.6). Under the simplified assumption, i.e. the assumption of exponential power delay profile and Jakes Doppler spectrum, the auto-correlation function can be fully characterized by the coherence bandwidth and the coherence time. Let's define the normalized chunk dimension, i.e. the normalized chunk bandwidth and the normalized chunk duration, as

$$
\tilde{B}_{\text{chunk}} = \frac{B_{\text{chunk}}}{B_{\text{coh}}} = \frac{n_{\text{sub}}f_s}{B_{\text{coh}}},
\tag{4.22}
$$

and

$$
\tilde{T}_{\text{chunk}} = \frac{T_{\text{chunk}}}{T_{\text{coh}}} = \frac{n_{\text{symb}}T_s}{T_{\text{coh}}}.
\tag{4.23}
$$

From the derivation in Appendix A.1, the Ω and so the capacity of adaptive resource allocation $\bar{C}_{\text{tot}}^{\text{adapt}}$ can be expressed as a function of normalized chunk dimension, $\tilde{B}_{\text{chunk}}$ and $\tilde{T}_{\text{chunk}}$, as seen in (A.9), i.e.

$$
\begin{aligned}
\Omega \;=\; & \frac{1}{\tilde{B}_{\text{chunk}}} \int_{-\tilde{B}_{\text{chunk}}}^{\tilde{B}_{\text{chunk}}} \left(1 - \frac{|f|}{n_{\text{sub}}} \right) \frac{1}{1 + \frac{j2\pi}{5}f} df \cdots \\
& \cdot \frac{1}{\tilde{T}_{\text{chunk}}} \int_{-\tilde{T}_{\text{chunk}}}^{\tilde{T}_{\text{chunk}}} \left(1 - \frac{|t|}{\tilde{T}_{\text{chunk}}} \right) J_0(2\pi t) dt.
\end{aligned}
\tag{4.24}
$$

Hence, it can be concluded that the multi-user diversity gain decreases with the increasing normalized chunk dimension.

As an example, in Figure 4.1 the average capacity achieved by the adaptive resource allocation $\bar{C}_{\text{tot}}^{\text{adapt}}$ is reported as a function of the normalized chunk dimension with system parameters of 16 users and average receiving SNR of 5 dB, i.e. $K = 16$ and $P/\sigma_z^2 = 5\,\text{dB}$.
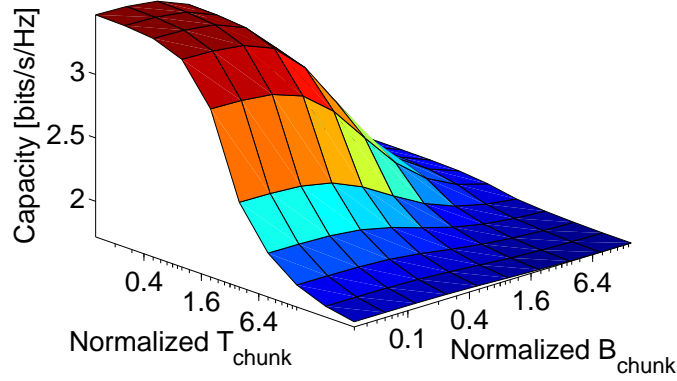
Figure 4.1: Capacity achieved by adaptive resource allocation with different normalized chunk dimensions, $K = 16$, $P/\sigma_z^2 = 5\,\mathrm{dB}$.

4.2.3 Derivation of Optimum Chunk Dimension

In this section, the chunk dimension is optimized based on the analysis presented in the previous section.

It can be seen from (4.1) that the amount of downlink control data for each chunk is $(\lceil \log_2 K \rceil + \lceil \log_2 Q \rceil)$. By denoting with r_{ctl} the rate of the control data and recalling that each chunk is composed of $n_{\mathrm{sub}} n_{\mathrm{symb}}$ symbols, the overhead caused by the downlink control data is given by

$$\mathrm{Ovh} = \frac{\lceil \log_2 K \rceil + \lceil \log_2 Q \rceil}{r_{\mathrm{ctl}} \cdot n_{\mathrm{sub}} n_{\mathrm{symb}}}. \tag{4.25}$$

The effective capacity defined as

$$\bar{C}_{\mathrm{tot,eff}}^{\mathrm{adapt}} = (1 - \mathrm{Ovh}) \cdot \bar{C}_{\mathrm{tot}}^{\mathrm{adapt}} \tag{4.26}$$

can be used as the measure of the system performance. As the first item in (4.26), $(1 - \mathrm{Ovh})$, increases with increasing chunk dimension, but the second item, $\bar{C}_{\mathrm{tot}}^{\mathrm{adapt}}$, decreases with increasing chunk dimension, there exists an optimum choice for the chunk dimension with which the effective capacity is maximized, i.e.

$$(n_{\mathrm{sub}}^*, n_{\mathrm{symb}}^*) = \arg \max_{n_{\mathrm{sub}}, n_{\mathrm{symb}}} \bar{C}_{\mathrm{tot,eff}}^{\mathrm{adapt}}. \tag{4.27}$$

In the following, the effective capacity is numerically evaluated by assuming the system parameters as listed in Table 4.1.

In Figure 4.2, the effective capacity is reported as a function of the chunk dimension with $\tau_{\max} = 3.2\,\mu\mathrm{s}$ and $v = 100\,\mathrm{km/h}$. It can be seen that the optimum chunk dimension is

Table 4.1: Parameters used for the calculation of the effective capacity.

Carrier frequency, f_c	5 GHz
Bandwidth, B	40 MHz
No. of sub-carriers, N	1024
No. of users, K	16
P/σ_η^2	5 dB
rate of control data, r_c	1
No. of MCSs, Q	10

8 sub-carriers by 18 OFDM symbols. The corresponding overhead Ovh in (4.25) and the maximum effective capacity $\bar{C}_{\mathrm{tot,eff,max}}^{\mathrm{adapt}}$ in (4.26) are 5.56% and 3.06 bits/s/Hz, respectively.

It has be seen from (A.9) that the capacity $\bar{C}_{\mathrm{tot}}^{\mathrm{adapt}}$ is a function of the normalized chunk dimension, and a higher capacity is achieved in case of a smaller normalized chunk dimension. Moreover, the normalized chunk dimension is reciprocal to the coherence bandwidth and coherence time, cf. (4.22) and (4.23). Hence, for a given chunk dimension, when the values of the coherence bandwidth and/or the coherence time are large, the performance degradation will not be so critical. It follows that the performance degradation can be traded for overhead reduction, resulting in a large optimum chunk dimension.

In Table 4.2, the optimum chunk dimension, the maximized effective capacity as well as the corresponding overhead are reported for two typical propagation scenarios with different multi-path delay spreads of 1600 ns and 3200 ns and velocity of 100 km/h. As expected, compared to the scenario with $\sigma_\tau = 3200$ ns, the optimum chunk dimension is larger and the corresponding maximum effective capacity is higher for the scenario with the smaller delay spread $\sigma_\tau = 1600$ ns.

Table 4.2: Optimal chunk dimensions in scenarios with different multi-path delay spread.

Max Delay Spread [μs]	Eff. Capacity [bits/s/Hz]	Opt. Chunk Dimension (n_{sub},n_{symb})	Overhead
1.6	3.18	(12,16)	4.17 %
3.2	3.06	(8,18)	5.56 %
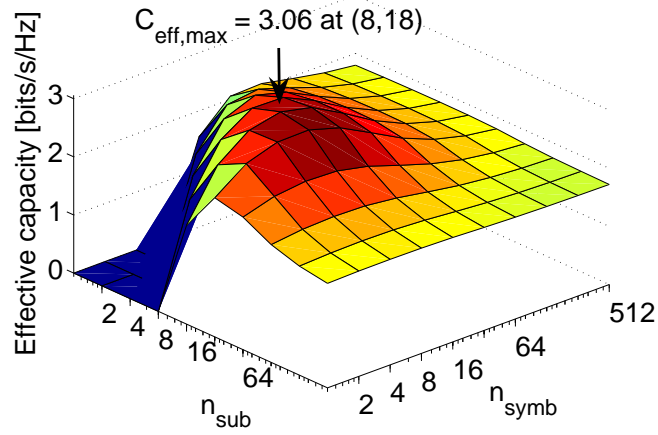
Figure 4.2: Effective capacity achieved by the adaptive resource allocation with different chunk dimensions, $\tau_{\mathrm{max}} = 3.2\,\mu$s and $v = 100$ km/h.

Similarly, the optimum chunk dimension is larger and the corresponding maximum capacity is higher for a scenario with a larger coherence time, i.e. at a lower velocity, as shown in Table 4.3.

Table 4.3: Optimal chunk dimensions in scenarios with different velocities.

Mobility [km/h]	Eff. Capacity [bits/s/Hz]	Opt. Chunk Dimension $(n^*_{\mathrm{sub}}, n^*_{\mathrm{symb}})$	Overhead
30	3.24	(6,44)	3.03 %
100	3.06	(8,18)	5.56 %
200	2.91	(10,10)	8.00 %

4.3 Optimization of Channel Update Interval

4.3.1 Introduction

Adaptive resource allocation as well as AMC allows efficient exploitation of high multi-user diversity gain through frequency-adaptive transmission based on the instantaneous

CSI. However, pilot insertion for the measurement and/or signaling of the CSI leads to additional overhead. Moreover, the measured and/or the signalled CSI feedback may differ from the ideal one. In general, a high accuracy of CSI requires a high overhead, which results in a trade-off between the CSI accuracy that affects the adaptation gain, and the required overhead that reduces the overall throughput. Furthermore, in a time-variant channel, the CSI shall be updated after certain interval when the channel is different from the previous measurement/feedback. A short update interval causes too high overhead that cannot be compensated by the adaptation gain, whilst with a too long update interval the CSI does not match with the current channel status and the adaptation gain degrades.

In this section, a method to identify the optimum update interval, which maximizes the system performance in terms of effective throughput, is discussed. This method has been firstly introduced by the author of this thesis in [ZCL05, ZCL06]. The effective throughput takes also the pilot and/or signaling overhead into account. Let T_{up} denote the CSI update interval in unit of frames. By supposing that the CSI is updated in the current frame, this CSI will be used for the resource allocation in the next T_{up} frames including the current one. The shortest update interval of 1 frame indicates that a new CSI is available in each frame. The optimum update interval is defined as the one for which the effective throughput is maximized.

By letting $\rho_0(T_{up})$ be the throughput as a function of the update interval T_{up} and by assuming one CSI update requires a fraction β of the resources available in one frame, the effective throughput is defined as

$$\rho = (1 - \frac{\beta}{T_{up}})\rho_0(T_{up}). \tag{4.28}$$

The remainder of the section is organized as follows. In Section 4.3.2, the throughput ρ_0 as a function of the update interval is derived by means of simulations. In Section 4.3.3, an analytical method to derive the optimum update interval is presented and the results are discussed.

4.3.2 Derivation of Performance Loss due to Channel Mismatch

In the section, the throughput as a function of the update interval, $\rho_0(T_{up})$, is derived through simulations.

One BS sector with 16 users is simulated. Homogenous users are assumed, i.e. the average SNRs and the velocities of all users are supposed to be identical. As ideal CSI update is further assumed, no channel mismatch exists with an update interval of 1 frame. Moreover, adaptive resource allocation is carried out by using the SBI greedy algorithm with max-min fairness strategy, i.e. SBI-FP, proposed in Section 3.3.3.

In Figure 4.3, the average user throughput ρ_0 is reported as a function of the update interval T_{up} with an average SNR of $10\,\mathrm{dB}$ and a velocity of $5\,\mathrm{km/h}$. As expected, the average user throughput decreases with increasing update interval.

The SBI greedy algorithm aims at maximizing the throughput by proper bit and power loading with respect to certain target CWER. On the basis of a CSI that matches the actual CSI only in the update frame, as the channel changes over time due to mobility, the channel mismatch in the following frames will keep enlarging till the CSI is updated again. Hence, after a CSI update, the CWER achieved in each frame will keep increasing along the time and it will meet the target CWER only when the CSI is updated again. In Figure 4.4, the average CWER of each frame is reported along with the frame index for several different update intervals. Note that the CSI is updated in frame $1, 1 + T_{\mathrm{up}}, 1 + 2T_{\mathrm{up}}, \ldots$. It can be inferred that the shorter the update interval, the earlier the average CWER goes down to the target one, resulting in lower average CWER and consequently higher throughput.

In Figure 4.5(a), the average user throughput ρ_0 as a function of the update interval T_{up} is reported under the assumption of different velocities. It can be seen that the average user throughput decreases more significantly for higher velocity. This is quite reasonable. With the same update interval, the channel mismatch increases with increasing velocity due to more rapidly varying channel.

As the channel time variability can be characterized by the channel coherence time $T_{\mathrm{coh}}(v)$ in (2.21), the relative update interval defined as the update interval divided by the coherence time is introduced. By letting T_{f} denote the frame duration, the relative update interval T'_{up} is defined as

$$T'_{\mathrm{up}} = \frac{t_{\mathrm{up}} \cdot T_{\mathrm{f}}}{T_{\mathrm{coh}}(v)}, \tag{4.29}$$

It can be inferred from Figure 4.5(b) that the relationship between the average user throughput and the relative update interval is almost the same for different velocities.

By letting $\bar{\rho}_0(T'_{\mathrm{up}})$ denote the average throughput as a function of the relative update interval T'_{up} for any arbitrary velocity, the effective throughput in (4.28) is given by

$$\rho = \left(1 - \frac{\beta}{T_{\mathrm{up}}}\right) \cdot \bar{\rho}_0\left(T'_{\mathrm{up}}\right) = \left(1 - \frac{\beta}{T_{\mathrm{up}}}\right) \cdot \bar{\rho}_0\left(\frac{T_{\mathrm{up}} T_{\mathrm{f}}}{T_{\mathrm{coh}}(v)}\right), \tag{4.30}$$

which is a function of overhead factor β, the velocity v and the update interval T_{up}.

4.3.3 Derivation of Optimum Channel Update Interval

In this section, the optimum update interval at any arbitrary velocity is derived. From (4.30) and Figure 4.5, it can be inferred that once the function $\bar{\rho}_0(T'_{\mathrm{up}})$ is obtained by means of curve fitting from the simulation results corresponding to a certain velocity, the optimum update interval at any arbitrary velocity can be calculated.
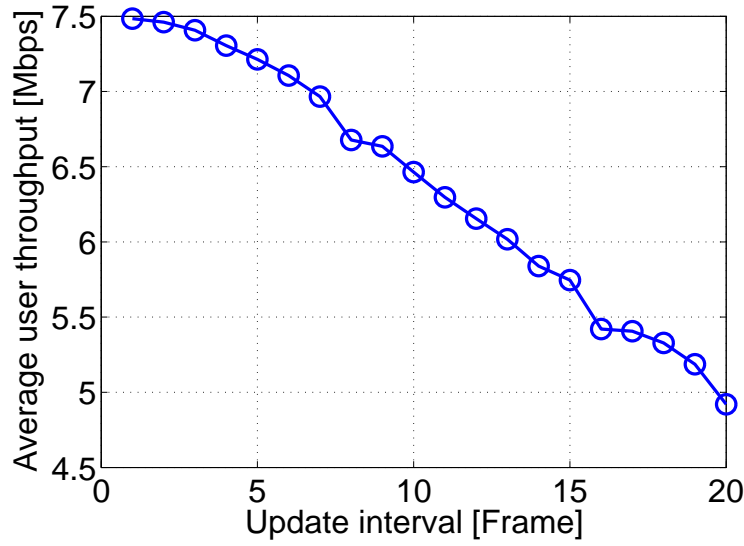
Figure 4.3: Average user throughput achieved by the adaptive resource allocation at different CSI update intervals, average SNR = 10 dB, $v = 5$ km/h.
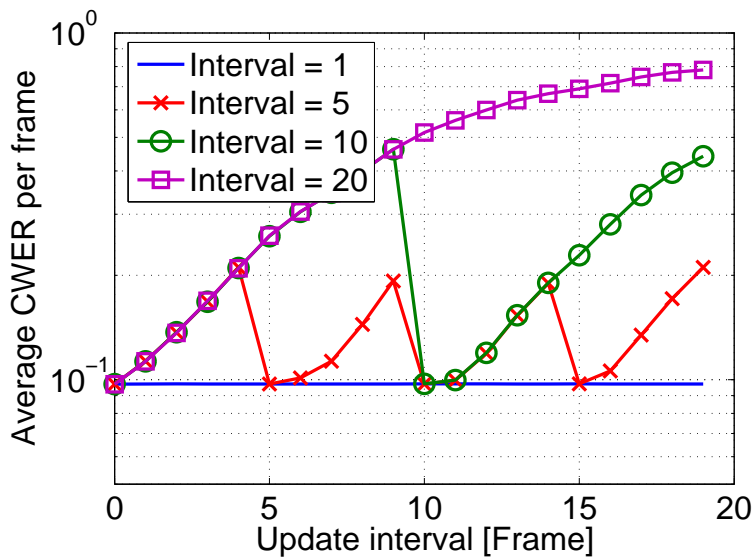

Figure 4.4: Average CWER vs. Frame index, average SNR = 10 dB, $v = 5$ km/h.

(a) x-axis: update interval, T_{up}

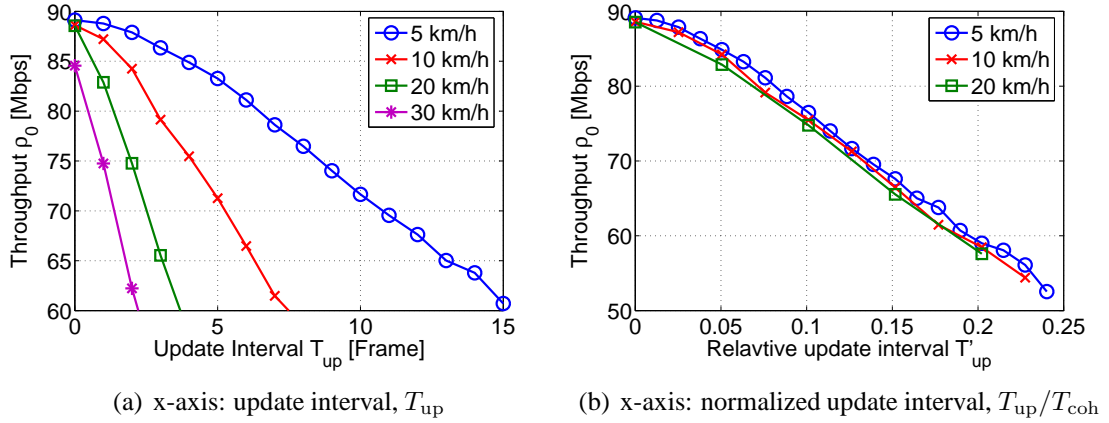(b) x-axis: normalized update interval, $T_{\text{up}}/T_{\text{coh}}$

Figure 4.5: Comparison of average user throughput achieved by adaptive resource allocation at different velocities.

As shown in Figure 4.6, the numerical results of throughput $\bar{\rho}_0(T'_{\text{up}})$ can be approximately regarded as being linear with respect to the relative update interval T'_{up}. By approximating the throughput $\bar{\rho}_0(T'_{\text{up}})$ by the linear function

$$\bar{\rho}_0(T_{\text{up}}) \approx -\xi_1 \cdot T'_{\text{up}} + \xi_0, \tag{4.31}$$

the effective throughput in (4.30) is given by

$$\rho(\beta, v, T_{\text{up}}) = \left(1 - \frac{\beta}{T_{\text{up}}}\right)\left(-\xi_1 \cdot \frac{T_{\text{up}}T_{\text{f}}}{T_{\text{coh}}(v)} + \xi_0\right). \tag{4.32}$$

From (4.32), the optimum update interval $T_{\text{up,opt}}$ as well as the maximum effective throughput ρ_{max} can be derived as a function of the overhead β and the velocity v, as stated in following theorem proved in Appendix A.2.

Theorem 1 *Under the assumption of the throughput being linear with respect to the relative update interval as described in (4.32), the optimum update interval is proportional to the square root of the ratio between the overhead β and the velocity v*

$$T_{\text{up,opt}} = \sqrt{\frac{\xi_0 c}{\xi_1 T_{\text{f}} f_c}} \cdot \sqrt{\frac{\beta}{v}}, \tag{4.33}$$

with c and f_c being the speed of light and the carrier frequency, respectively. The corresponding maximum effective throughput can be expressed as a quadratic function of $\sqrt{\beta v}$, i.e.

$$\rho_{\text{max}} = \rho(T_{\text{up,opt}}) = \xi_1 T_{\text{f}} \frac{f_c}{c}(\sqrt{\beta v})^2 - 2\sqrt{\xi_1 \xi_0 T_{\text{f}} \frac{f_c}{c}}\sqrt{\beta v} + \xi_0, \tag{4.34}$$

which moronically decreases with $\sqrt{\beta v}$.
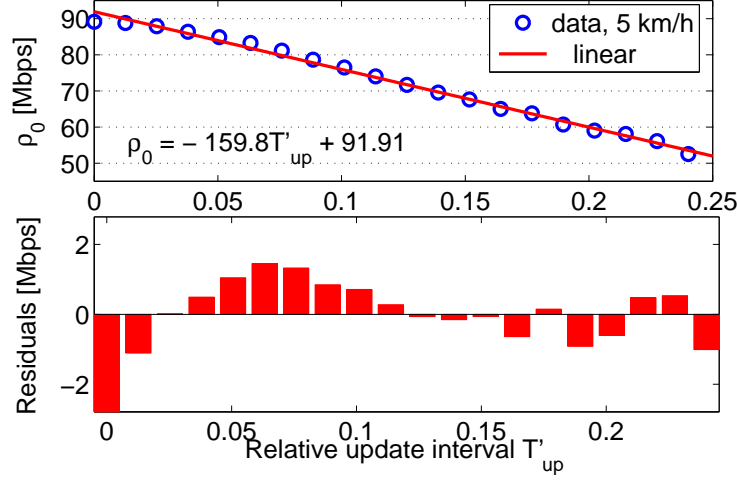
Figure 4.6: Curve Fitting. v = 5 km/h, SNR = 10 dB.

From the theorem, following conclusion can be drawn.

- With increasing velocity v, both the optimum update interval $T_{\text{up,opt}}$ in (4.33) and the maximum effective throughput decreases, as depicted in Figure 4.7(a). In Figure 4.7, the effective user throughput is depicted as the function of update interval for different velocities under the assumption of $\beta = 1/6$ [1].

- By substituting (2.21) and (4.33) into (4.29), the optimum relative update interval is given by

$$T'_{\text{up,opt}} = \sqrt{\frac{\xi_0 T_{\text{f}} f_c}{\xi_1 c}} \cdot \sqrt{\beta v}. \tag{4.35}$$

Thus, with increasing velocity v, the optimum relative update interval $T'_{\text{up,opt}}$ increases, as depicted in Figure 4.7(b). Since the channel mismatch is proportional to the relative update interval, more insights can be inferred from Figure 4.7(b). As explained in Section 4.3.1, optimizing the update interval consists in finding the tradeoff between the overhead caused by the CSI update and the loss in adaptation gain

---

[1]The assumption of β equal to 1/6 comes from the following consideration: under the assumption of TDD operation mode, an estimate of the channel is supposed to be obtained by measuring the pilots periodically inserted in uplink. In the considered system, each frame consists of 128×1 chunks and each chunk consists of 8 sub-carriers and 12 symbols, cf. Table 2.3. It is reasonable to assume that one pilot symbol per chunk per antenna is required for each user to measure the channel [IST05a]. Suppose there are $K = 8$ active users in the system at each user is configured with two antennas, totally $1 \times 8 \times 2$ pilot symbols are required for one channel measurement, and so $\beta = (1 \times 8 \times 2)/(8 \times 2) = 1/6$.
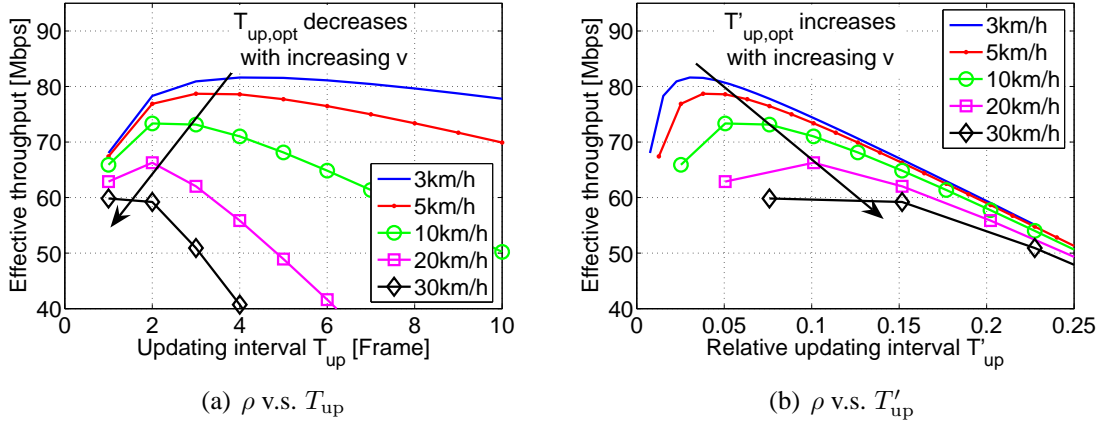
Figure 4.7: Optimum update intervals at different velocities, SNR = 10 dB.

due to the mismatched CSI. At a low velocity, as the channel does not change rapidly, the CSI does not need to be updated very often. It follows that the overhead is not so relevant as at high velocities, and the loss of adaptation gain becomes relatively more critical. Hence, less loss of the adaptation gain should be ensured by decreasing the relative update interval T'_{up} at lower velocities.

- With increasing cost for the CSI update, as expressed by β, the optimum update interval $T_{up,opt}$ in (4.33) increases and the maximum effective throughput decreases, as depicted in Figure 4.8. This can be explained by the following: a larger update interval is desired so as to reduce the overhead when the impact of overhead becomes more dominant due to the higher cost of the CSI update.

4.4 Optimization of Adaptive Resource Allocation with Reduced Channel Feedback

4.4.1 Introduction

Provided with full CSI, the AP equipped with multiple antennas can serve multiple users on orthogonal beams of the same chunk by means of ZFBF, cf. Section 2.5.3. In this way, co-located users share the same chunk but experience no interference from each other. Further on, algorithms enabling joint optimization of adaptive OFDMA and adaptive SDMA have been proposed and evaluated in Chapter 3 under the assumption that ZFBF is applied based on full CSI at the transmitter. In order to let the AP acquire full CSI, however, additional overhead needs to be introduced, which might be unaffordable in practical applications, especially in high user mobility scenarios. There exist other techiniques that enable SDMA
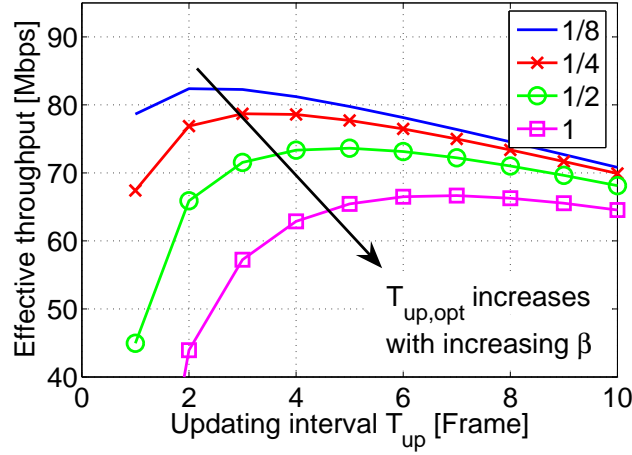
Figure 4.8: Optimum update intervals with different amount of cost for CSI update β. v = 5 km/h, SNR = 10 dB.

but require much less channel information, such as adaptive semi-orthogonal beamforming [FN95, SB04a] and fixed grid-of-beam (GoB) beamforming.

Adaptive semi-orthogonal beamforming enables multiple users to share the same resource by transmitting over semi-orthogonal beams built according to the second-order statistics of the channel matrix, i.e. the channel correlation matrix [FN95, SB04a]. Let $\mathbf{f}_{k,n}$ denote the receive beamforming vector of user k on chunk n and further assume that it is normalized, i.e.

$$\|\mathbf{f}_{k,n}\|^2 = 1. \tag{4.36}$$

Let $\mathcal{U}_n$ denote the set of users served on chunk n, $P_{k,n}$ denote the transmit power of user k on chunk n, $\mathbf{w}_{k,n}$ denote the transmit beamforming vector of user k and σ_z^2 denote the noise power. Thus, from (2.35) and (2.36) the receive SINR of user k on chunk n is derived as

$$\text{SINR}_{k,n} = \frac{P_{k,n}|\mathbf{f}_{k,n}^{\text{H}}\mathbf{H}_{k,n}\mathbf{w}_{k,n}|^2}{\sum_{\substack{j\in\mathcal{U}_n \\ j\neq k}} P_{j,n}|\mathbf{f}_{k,n}^{\text{H}}\mathbf{H}_{k,n}\mathbf{w}_{j,n}|^2 + \sigma_z^2}. \tag{4.37}$$

When using adaptive semi-orthogonal beamforming, the specific choice of one user's transmit beamforming vector becomes a complicated task because it also affects the interference experienced by other co-located users [SB04a]. As shown in [SB04a] and in the references therein, the downlink beamforming problem is generally formulated as optimization of the transmit beamforming vector $\mathbf{w}_k$ and power allocation $p_{k,n}$ subject to users' individual target SINRs. A rapidly converging iterative algorithm has been proposed in [SB04a] by converting the downlink beamforming problem into an equivalent dual uplink problem.

As AMC is allowed in the considered system, users have no fixed target SINR values, but the suitable MCS can be adaptively chosen according to the receive SINR. When AMC is enabled, adaptive power allocation is not critical any more and equal power allocation can yield already good results [BPS98]. Therefore, fixed power allocation in terms of equal power sharing is assumed in the following. Moreover, among all the existing semi-orthogonal beamforming schemes, the so called Generalized Eigenbeamforming is able to balance the useful signal enhancement and the interference reduction [Zet95], and so it is adopted in this work.

It can be seen in (4.37) that the receive SINR of each user on individual chunks cannot be calculated without full CSI. If a short-term channel quality indicator (CQI), e.g. the channel norms of individual chunks, is available, the SINR can be estimated based on the short-term CQI and the channel correlation matrix, as discussed in Section 4.4.2. However, the estimation error results in performance worse than expected.

Fixed GoB beamforming has been proposed in [SH05]. Independent data streams are supposed to be transmitted over the beams simultaneously. Based on periodic pilot transmission over the beams, each user can measure the SINR of each beam and feed back the index of the best beam along with the corresponding SINR. This approach is generally not optimum since the beams are not adapted to the channel conditions. However, compared to adaptive semi-orthogonal beamforming, e.g. generalized eigenbeamforming, fixed GoB beamforming may benefit from precise AMC based on accurate receive SINR.

The remainder of the section is organized as follows. In Section 4.4.2, after reviewing the generalized eigenbeamforming, the methods for estimating receive SINR based on channel correlation matrix and short-term CQI are proposed. In Section 4.4.3, fixed GoB beamforming is reviewed and the estimation of the receive SINR is discussed. Their performance is then presented and compared in Section 4.4.4.

4.4.2 Generalized Eigenbeamforming

In this section, after reviewing the generalized eigenbeamforming, the methods for estimating receive SINR based on channel correlation matrix and short-term CQI are proposed and analyzed.

When the channel varies rapidly due to high user mobility, it becomes impossible for the AP to obtain full CSI due to unaffordable signaling overhead. However, SDMA can still be conveniently constructed through eigenbeamforming techniques based on the channel correlation matrix [SB04a]. By denoting with $\mathbf{H}_{k,n}(t)$ the time-varying channel matrix of user k on chunk n at time t, the channel correlation matrix of user k is given by

$$\mathbf{R}_k = E[\mathbf{H}_{k,n}(t)^{\mathrm{H}}\mathbf{H}_{k,n}(t)], \tag{4.38}$$

where the expectation is taken over frames as well as over all chunks available in the frame.

The generalized eigenbeamforming approach aims at balancing the useful signal enhancement and the interference reduction by choosing the transmit beamforming vector $\mathbf{w}_k$ as

$$\mathbf{w}_k = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^{\mathrm{H}} \mathbf{R}_k \mathbf{w}}{\mathbf{w}^{\mathrm{H}} \left(\sum_{j \in \mathcal{U}, j \neq k} \mathbf{R}_j + \beta_k \mathbf{I} \right) \mathbf{w}} \tag{4.39}$$

[Zet95], where β_k is a scaling factor and is suggested to be chosen as

$$\beta_k = \frac{0.1}{M_t} \mathrm{tr} \left[\sum_{j \in \mathcal{U}, j \neq k} \mathbf{R}_j \right] \tag{4.40}$$

in [SBO06].

It can be inferred from (4.39) that the transmit beamforming vector of a given user also depends on the transmit beamforming vectors of other co-located users. Hence, the computation of beamforming vectors $\mathbf{w}_k$ and thus the estimation of the receive SINR $\mathrm{SINR}_{k,n}$ in (4.37) can only be carried out at the AP.

Two methods are proposed by the author of this thesis in [ZCE07] to estimate the receive SINR under the assumption that the normalized channel correlation matrix,

$$\bar{\mathbf{R}}_k = \frac{\mathbf{R}_k}{\|\mathbf{R}_k\|} \tag{4.41}$$

and the short-term CQI in terms of chunk-specific channel norm,

$$g_{k,n} = \|\mathbf{H}_{k,n}\| \tag{4.42}$$

are available at the AP.

In the first method, by ignoring the receive beamforming vector $\mathbf{f}_{k,n}$, the receive SINR in (4.37) can be expressed as

$$\mathrm{SINR}_{k,n} = \frac{P_{k,n} \mathbf{w}_k^{\mathrm{H}} \bar{\mathbf{R}}_k \mathbf{w}_k}{\sum_{\substack{j \in \mathcal{U}_n \\ j \neq k}} P_{j,n} \mathbf{w}_j^{\mathrm{H}} \bar{\mathbf{R}}_k \mathbf{w}_j + \sigma_z^2 / g_{k,n}^2}. \tag{4.43}$$

Since the use of multiple receive antennas enables signal enhancement and/or interference reduction, by neglecting the actual receive beamforming vector, the SINR prediction in (4.43) is in general pessimistic, thus resulting in conservative AMC. In other words, a MCS supporting higher data rate could be actually supported. For this reason, the generalized eigenbeamforming approach based on this conservative estimate of the receive SINR is referred to as GenEigBF-CONS hereafter.

The second method for the estimate of the receive SINR relies on an approximation of the channel matrix. Based on the eigenvalue decomposition of the channel correlation matrix

$$\bar{\mathbf{R}} = \mathbf{V}_k \mathbf{\Lambda}_k \mathbf{V}_k^{\mathrm{H}}, \tag{4.44}$$

the channel matrix $\mathbf{H}_{k,n}$ can be approximated by the first $M_{r,k}$ dominant eigenvectors and eigenvalues as

$$\hat{\mathbf{H}}_{k,n} = g_{k,n} \cdot \left[\begin{array}{ccc} \lambda_{k,1}\mathbf{v}_{k,1} & \cdots & \lambda_{k,M_{r,k}}\mathbf{v}_{k,M_{r,k}} \end{array} \right]^{\mathrm{H}} \tag{4.45}$$

with $M_{r,k}$ being the number of receive antenna of user k [Hay96]. By knowing the channel matrix $\hat{\mathbf{H}}_{k,n}$, the receive beamforming vector $\mathbf{f}_{k,n}$ can be determined by assuming e.g. MMSE filtering according to (2.46), and the receive SINR can then be computed according to (4.37). Due to the existing approximation error of the approximated channel matrix $\hat{\mathbf{H}}_{k,n}$ in (4.45), the receive SINR is generally overestimated and hence results in an aggressive AMC, referred to as GenEigBF-AGGR.

4.4.3 Fixed GoB Beamforming

In this section, fixed GoB beamforming is reviewed and the estimation of the receive SINR is discussed.

In case of fixed GoB beamforming, the AP constructs Q beams

$$\mathcal{B} = \{\mathbf{b}_q\}, q = 1, \dots, Q, \tag{4.46}$$

and independent data streams are then transmitted over these beams with equally distributed power

$$p_0 = \frac{P_{\mathrm{chunk}}}{Q} \tag{4.47}$$

simultaneously. The Q beams are supposed to be configured in such a way to not only enhance the average beamforming gain but also to reduce the interference among beams. According to (4.37), combined with the transmit power p_0 in (4.47) and the transmit beamforming vector $\mathbf{b}_q$ in (4.46), the SINR of beam q experienced by user k on chunk n is given by

$$\mathrm{SINR}_{k,n,q} = \frac{\|\mathbf{f}_{k,n,q}^{\mathrm{H}}\mathbf{H}_{k,n}\mathbf{b}_q\|^2}{\sum_{j=1,j\neq q}^{Q} |\mathbf{f}_{k,n,q}^{\mathrm{H}}\mathbf{H}_{k,n}\mathbf{b}_j\|^2 + \frac{\sigma_z^2 \cdot Q}{P_{\mathrm{chunk}}}}. \tag{4.48}$$

The AP is supposed to transmit dedicated pilots, i.e. pilots are transmitted over individual beams. In this way, the users are allowed to measure the SINR of each beam without knowing the transmit beamforming vectors. Based on the measured SINR values for all beams on each chunk, users feed back the index of the best beam along with the corresponding SINR for each chunk.

In general, the GoB constructed for different chunks can be different. In [SH05], the GoB is chosen as Q random orthonormal vectors, which are generated independently for each chunk according to an isotropic distribution.

As introduced in Section 2.3, in the considered typical urban macro-cellular scenario, the channel correlation at the BS can be characterized by the main AoD with certain small

angular spread. From (2.6) and by letting h_0 denote the channel response on the first antenna element, the channel response of the M_t-element uniform linear array (ULA) with antenna spacing d for a signal coming from the AoD ϕ is equal to

$$\mathbf{h}_{M_t,d}(\phi) = h_0 \cdot [1, e^{j\frac{2\pi f_c}{c}d\sin(\phi)}, \ldots, e^{-j\frac{2\pi f_c}{c}(M_t-1)d\sin(\phi)}]^{\mathrm{T}}. \tag{4.49}$$

Thus, the corresponding transmit beamforming vector that maximizes the signal coming from the AoD ϕ is

$$\mathbf{a}(\phi) = \frac{\mathbf{h}^{\mathrm{H}}M_t, d(\phi)}{\|\mathbf{h}M_t, d(\phi)\|} \tag{4.50}$$

[Hay96].

For the considered antenna array, i.e. 4-element ULA with half wavelength antenna spacing, cf. Table 2.1, four beams with equally distributed main directions of $[-45°, -15°, 15°, 45°]$ can cover the $120°$ sector, as seen in Figure 4.9(a). Moreover, due to the high interference between neighboring beams, only non-neighboring beams can be simultaneously active in the same chunk. Here it is assumed to simply assign the beam pair (1,3) to odd chunks and the beam pair (2,4) to even chunks, i.e.

$$\mathcal{B}_n = \begin{cases} \{\mathbf{a}(-45°), \mathbf{a}(15°)\}, & n = 0, 2, \ldots \\ \{\mathbf{a}(-15°), \mathbf{a}(45°)\}, & n = 1, 3, \ldots \end{cases} \tag{4.51}$$

Additionally, Chebyshev tapering is introduced to further reduce the side-lobe level and in turn the interference among beams [Dol47], as depicted in Figure 4.9(b).

(a) Without Chebyshev tappering.

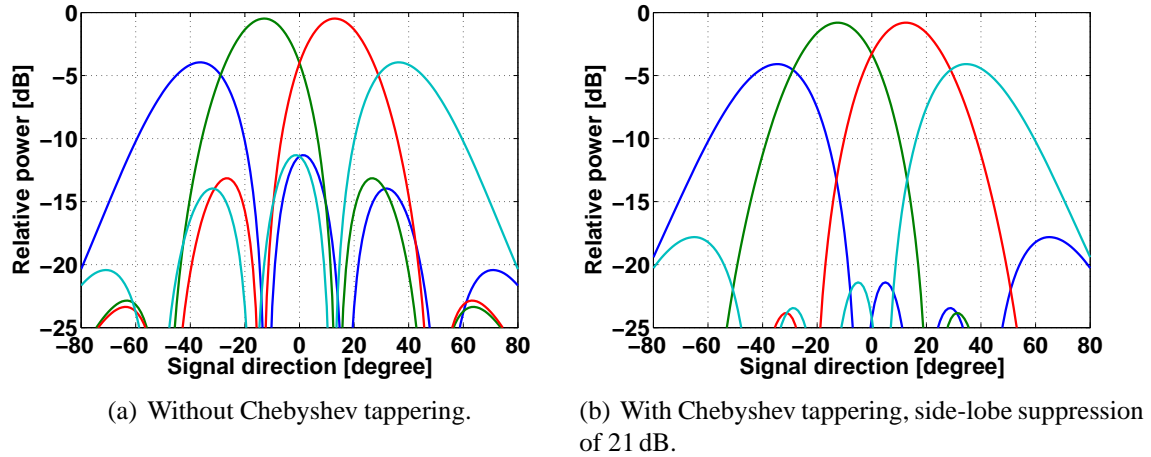(b) With Chebyshev tappering, side-lobe suppression of 21 dB.

Figure 4.9: Antenna pattern of the GoB in a $120°$ sector.

By following the SUI algorithm proposed in Section 3.3.3, the users are successively added on every chunk until performance decreases, and thus the number of added users can vary between one and the number of beams, $|\mathcal{B}_n|$. It has to be noticed that during pilot

transmission each user measures the receive SINR of the best beam by assuming all beams are used, denoted as $\text{SINR}(\mathcal{B}_n)$. As a consequence, when $|\mathcal{U}_n| \leq |\mathcal{B}_n|$, the actual receive SINR of user k on chunk n is upper bounded by

$$\text{SINR}_{k,n}(\mathcal{U}_n) \geq \frac{|\mathcal{B}_n|}{|\mathcal{U}_n|}\text{SINR}_{k,n}(\mathcal{B}_n) \tag{4.52}$$

It is observed that, as illustrated in Figure 4.10, the spatial power spectrum has about the same width as the beam in GoB for the considered typical urban macro-cellular scenario. This indicates that each user tends to select the same beam over the whole bandwidth. Therefore, the selection of the best beam has not to be necessarily chunk-specific. Therefore, it is proposed to select the best beam on a long-term basis, e.g. every several frames, so as to reduce the amount of the feedback.
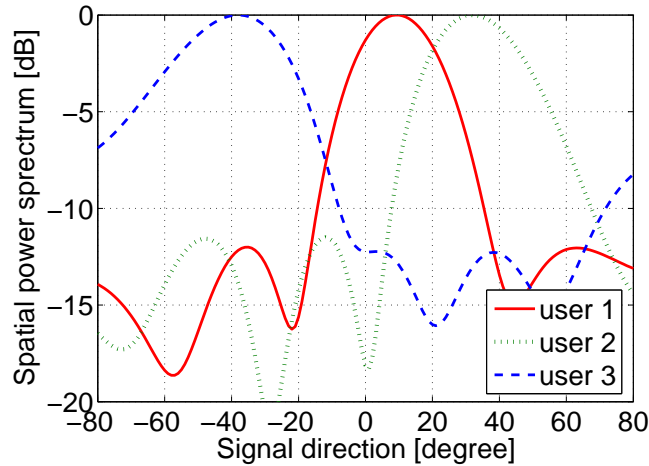
Figure 4.10: Spatial power spectrum in the considered urban-macro scenario.

4.4.4 Simulative Comparison

In this section, the performance achieved by the two SDMA schemes, generalized eigen-beamforming and fixed GoB, is evaluated by means of simulations. One BS sector with 16 uniformly distributed users is simulated. All users are assumed to experience an identical average SNR. Moreover, adaptive resource allocation is carried out by using the SUI greedy algorithm with max-min fairness strategy, i.e. SUI-FP, proposed in Section 3.3.3. The simulation results are reported in terms of average user throughput as a function of the average SNR.

The performance of generalized eigenbeamforming, cf. Section 4.4.2, is depicted in Figure 4.11(a).
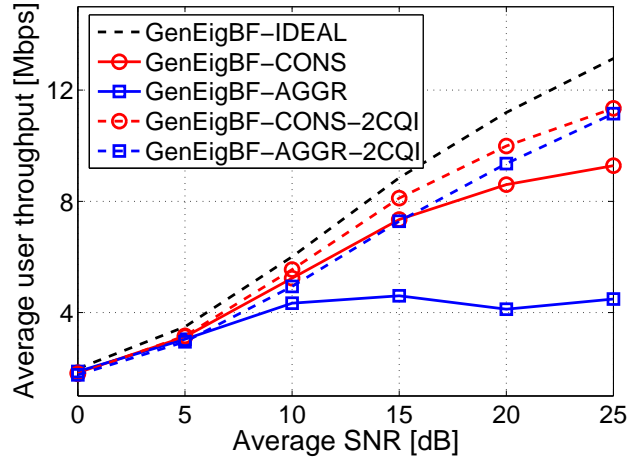
(a) Correctly received
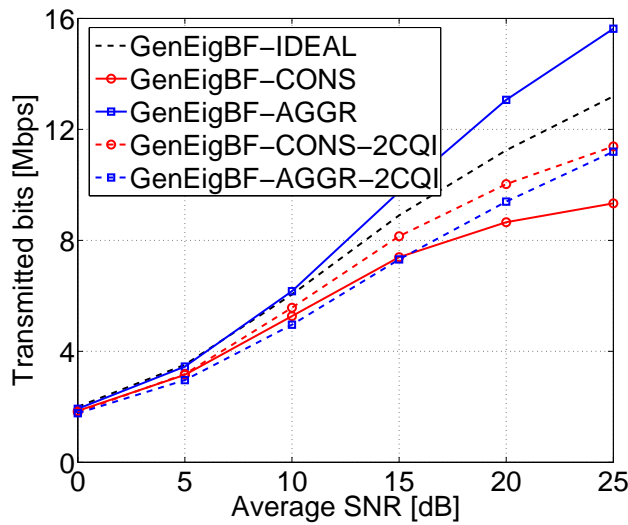
(b) Transmitted

Figure 4.11: Comparison of the different estimation methods for receive SINR achieved by generalized eigenbeamforming in terms of average user throughout, SUI-FP.

GenEigBF-IDEAL (dashed curve) yields an upper bound since it depicts the performance achievable when assuming the receive SINR is perfectly known at the BS. GenEigBF-CONS (solid curve with circles) and GenEigBF-AGGR (solid curve with squares) represent the performance based on the conservative SINR estimate in (4.43) and the aggressive SINR estimate in (4.45), respectively. It can be inferred that GenEigBF-

CONS achieves better performance than GenEigBF-AGGR, indicating that conservative resource allocation is beneficial. This result can be explained by considering also the average rate of transmitted bits reported for all the three variants in Figure 4.11(b). Compared to GenEigBF-IDEAL, GenEigBF-CONS has lower throughput because less bits are transmitted, whilst for GenEigBF-AGGR, although more bits are transmitted, the throughput is still lower, as seen in Figure 4.11(a), which is because of higher CWER due to overestimated the SINR.

It can be inferred from Figure 4.11(a) that the performance degradation due to the inaccurate SINR estimation is quite considerable, especially at high SNR when the interference becomes dominant. The effect of imperfect SINR estimation can be compensated by adding a further CQI feedback. After the resource allocation based on the channel correlation matrix and the first CQI feedback in terms of channel norm has been carried out, the BS transmits dedicated pilots over the adaptively determined beams and all users measure the receive SINR on the assigned beam to signal back the second CQI feedback. This second CQI feedback is then used by the BS to perform AMC, with consequent significant performance improvement, as shown by GenEigBF-CONS-2CQI (dashed curve with circles) and by GenEigBF-AGGR-2CQI (dashed curve with squares) in Figure 4.11(a). However, the second CQI feedback implies additional feedback overhead as well as additional feedback delay.

Indeed, the performance degradation due to imperfect channel knowledge, indicated by e.g. the difference between GenEigBF-IDEAL and GenEigBF-CONS, is caused by two folds, non-ideal resource allocation and non-ideal AMC. As ideal AMC is applied in the case of GenEigBF-CONS-2CQI, the degradation due to non-ideal AMC is indicated by the difference with respect to GenEigBF-CONS. The effect of non-ideal resource allocation is then illustrated by the difference between GenEigBF-IDEAL and GenEigBF-CONS-2CQI. From Figure 4.11(a), it can be inferred that the degradation due to non-ideal AMC in case of aggressive allocation is much greater than that in case of conservative allocation, but the degradation due to non-ideal resource allocation is similar in both cases.

The performance of adaptive resource allocation based on fixed GoB is depicted in Figure 4.12. The solid and dashed curves with crosses represent the performance with chunk-specific and long-term beam selection, respectively. It can be seen that the two curves are overlapping, indicating there is almost no performance degradation due to long-term beam selection, as expected in Section 4.4.3.

The performance of generalized eigenbeamforming is depicted again in Figure 4.12 for the purpose of comparison. It can be seen that generalized eigenbeamforming relying on twice CQI feedback, GenEigBF-CONS-2CQI (dashed curve with circles), achieves better performance than fixed GoB based SDMA (solid curve with crosses). This is because in the case of generalized eigenbeamforming, the beamforming vectors are adaptively determined to maximize the signal power while minimizing the multi-user interference for a given group of users. However, generalized eigenbeamforming with only one CQI feedback,
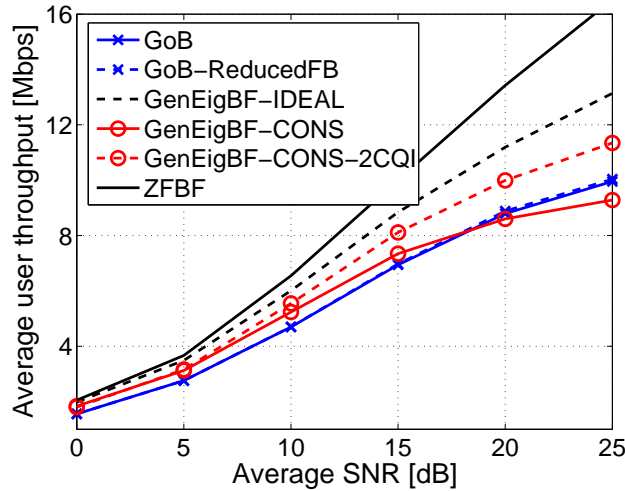
Figure 4.12: Comparison of the SDMA schemes based on different beamforming techniques in terms of average user throughout, SUI-FP.

GenEigBF-CONS (solid curve with circles), suffers from non-ideal AMC and it does not exhibit any gain over fixed GoB in the higher SNR region.

Moreover, the performance of ZFBF is also reported in Figure 4.12. It is obvious that ZFBF (solid curve) based on full CSI yields the highest throughput, especially in high SNR region.

4.5 Optimization of Uplink Bandwidth Request Transmission Mechanism

4.5.1 Introduction

In cellular system, it is the AP that controls the resource allocation in the downlink as well as in the uplink. There is no multiple access problem in the downlink since the AP is the only transmitter, but in the uplink all users see a multiple access channel. The uplink part of the frame is generally divided into a contention period and a period for transmission of data payload, cf. Section 2.6. In general, the user that has packets to transmit asks for allocation of resources by sending a bandwidth request (BW-REQ) to the AP in the contention period and then transmits the packets in the data payload period under the coordination of the AP, as described in, e.g., IEEE 802.16 [IEE04]. In other words, the AP acquires the knowledge of the amount of uplink data through the requests transmitted by users.

For periodically generated packets with fixed length, e.g. packets carrying real-time streaming services, there is no need to send BW-REQ for each packet. The more efficient way is to transmit a BW-REQ to reserve a fixed amount of resources for each frame thereafter and another one to cancel the previous reservation when the data stream ends. For packets carrying data services, since the packet arrival rate as well as the packet length is not constant, reserving resources for each frame is wasteful, and so BW-REQ can be transmitted for each packet. As a result, the BW-REQ is characterized as bursty for both constant bit-rate traffic and variable bit-rate traffic. In this Subsection, the request arrival is assumed to be modeled as Bernoulli process with a parameter λ, i.e. the arrival of a new BW-REQ occurs with probability λ, $0 \leq \lambda \leq 1$, in every frames.

The performance of BW-REQ transmission schemes is commonly measured in terms of delay of the BW-REQ transmission, which can be defined as the number of frames between the arrival and the successful transmission of the BW-REQ. Lower delay is desired in order to enable the AP to obtain instant traffic condition. In general, increasing the contention period results in a lower delay. However, since both contention period and data payload period share the same amount of the uplink resources, the increase of contention period decreases the available resources for data payload transmission and thus reduces the data throughput. Therefore, an efficient usage of the contention period is very important to improve the system performance.

Two principle schemes, namely polling and random access, can be used for the BW-REQ transmission [Tan96]. Polling means that all users transmit the BW-REQ in a collision free manner, i.e. on separated resources [Tan96]. For random access, the transmission is contention-based and so a collision avoidance algorithm is usually defined [Tan96]. IEEE 802.16 standard supports both polling and random access, which adopts slotted ALOHA protocol and truncated binary exponential back-off (TBEB) algorithm [IEE04], as introduced in the following.

Define a transmission opportunity (TO) as the resources required by transmitting one BW-REQ. It is then assumed that there are in total K users in the system and the AP allocates N TOs for the BW-REQ transmission in each frame.

In case of polling, each user is polled by the AP to transmit the BW-REQ, i.e. users transmit the BW-REQ on the N TOs in a round-robin manner. By assuming that the BW-REQ cannot be transmitted in the same frame when it arrives, the minimum delay is one frame. It is then readily derived that the delay is uniformly distributed between 1 and K/N frames. Thus, the average delay for the BW-REQ transmission with polling $\bar{d}^{(\mathrm{Poll})}$ is equal to

$$\bar{d}^{(\mathrm{Poll})} = \sum_{i=1}^{N/K} = \frac{1}{2} + \frac{K}{2N}. \qquad (4.53)$$

In case of random access, users randomly select one TO to transmit. The collisions is

handled by the TBEB algorithm [IEE04]. The TBEB algorithm is parameterized by two values, i.e. the initial back-off window size W_0 and the maximum back-off stage m, which is usually broadcasted by the AP in the control message [IEE04]. The TBEB algorithm is briefly described as follows. Before each transmission attempt, a user uniformly chooses an integer number c_i from the interval $[0, W_i - 1]$, where W_i is the current value of its back-off window. The chosen value c_i, referred to as a back-off counter, indicates the number of TOs the user has to wait before the transmission attempt. For the first transmission attempt, the back-off window is W_0. When a collision happens, the user doubles its back-off window value, and so the back-off window after i collisions W_i is given by

$$W_i = 2^i W_0, \tag{4.54}$$

which is usually named the window size of the i-th back-off stage [Tan96]. The back-off window is not doubled if it reaches the maximum value $2^m W_0$. In the case of the successful transmission the back-off window is reset to the minimum value W_0. The standard [IEE04] does not define any relationship between the initial back-off window size W_0 and the number N of TOs in the frame . However, it is noticed that, if $W_0 < N$, some TOs will never be used during the first transmission attempt. Hence, by denoting with L a natural number, let the initial window size W_0 be multiple of the number N of TOs, i.e.

$$W_0 = LN. \tag{4.55}$$

As a result, the transmission attempt is uniformly distributed over the N available TOs.

In the following analysis, ideal channel condition is assumed, i.e. if exactly one user transmits in a TO, the transmission is successful. Furthermore, the user is assumed to receive a feedback, indicating whether the transmission is successful, from the AP at the beginning of the next frame right after the transmission.

The remainder of the section is organized as follows. The analytical derivation of the performance for random access is presented in Section 4.5.2. In Section 4.5.3, the optimum parameterizations of the random access is discussed, and then by knowing the average delay achieved by polling and random access schemes, the better BW-REQ scheme that results in lower delay can then be identified for different scenarios. Finally, a novel approach enabling more efficient use of the contention period for BW-REQ transmission is proposed in Section 4.5.4.

4.5.2 Analytical Derivation of Performance of Random Access

In this section, the performance of random access for BW-REQ transmission is analytically derived.

The TBEB algorithm has been extensively analyzed for Wireless LAN in [Bia00]. The key assumption of the analysis for random access is that, at each transmission attempt, regardless of the number of retransmissions suffered, each BW-REQ transmission collides

with constant and independent probability p. This assumption leads to more accurate results as long as the number K of users and the back-off window W_i get larger [Bia00]. The probability p is referred to as conditional collision probability, meaning the probability of a collision seen by a request transmitted.

The average delay measured in numbers of frames can be expressed as a function of the conditional collision probability p as follows. Since the back-off counter at the i-th stage is uniformly chosen between 0 and $W_i - 1$, and the number of frames the user has to wait before transmission is uniformly distributed between 0 and $L_i - 1$, the resulting average delay at stage i, denoted with $\bar{d}_i$, is given by

$$\bar{d}_i = \frac{1}{2^i L} \sum_{j=0}^{2^i L - 1} j = \frac{2^i L - 1}{2}. \tag{4.56}$$

As the collision happens with the probability of p, the average delay over all stages has a geometrical distribution and is calculated as

$$\bar{d} = 1 + (1-p) \left(\sum_{i=0}^{m} (p^i \sum_{j=0}^{i} \bar{d}_j) + \sum_{i=m+1}^{\infty} (p^i \sum_{j=0}^{m} \bar{d}_j + (i-m)\bar{d}_m) \right). \tag{4.57}$$

By substituting (4.56) into (4.57), the average delay $\bar{d}$ of the BW-REQ transmission using random access can be expressed as the function of the conditional collision probability p as

$$\bar{d} = \frac{1}{2(1-p)} + \frac{L(1 - p - 2^m p^{m+1})}{2(1-2p)(1-p)}. \tag{4.58}$$

In the following, the analytical derivation of the conditional collision probability p is presented. Adopt the discrete time scale t, with t and $t+1$ representing the beginning of two consecutive TOs. Let $c(t)$ be the stochastic process representing the back-off counter, and $s(t)$ be the stochastic processes representing the back-off stage of a given user. It has been shown in [Bia00] that the bi-dimensional process $\{s(t), c(t)\}$ can be modeled with the discrete-time Markov chain depicted in Figure 4.13. With

$$P\{i_1, n_1 | i_0, n_0\} = P\{s(t+1) = i_1, c(t+1) = n_1 | s(t) = i_0, c(t) = n_0\}, \tag{4.59}$$

the Markov chain is described by the following transition probabilities,

$$P\{i, n | i, n+1\} = 1, \quad i \in [0, m], n \in [0, W_i - 2]; \tag{4.60}$$

$$P\{0, n | i, 0\} = \frac{1-p}{W_0}, \quad i \in [0, m], n \in [0, W_i - 1]; \tag{4.61}$$

$$P\{i, n | i-1, 0\} = \frac{p}{W_i}, \quad i \in [1, m], n \in [0, W_i - 1]; \tag{4.62}$$

$$P\{m, n | m, 0\} = \frac{p}{W_m}, \quad n \in [0, W_i - 1]. \tag{4.63}$$

Figure 4.13: Markov chain model of the TBEB algorithm [Bia00].

 [Bia00]. (4.60) accounts for the fact that the back-off counter $c(t)$ is decreased by 1 at the beginning of each TO. (4.61) describes the fact that following a successful transmission with probability $(1 - p)$, a new transmission starts with back-off stage 0 and the back-off counter is uniformly chosen from 0 to $W_0 - 1$. (4.61) describes that following a collision with probability p, the back-off stage is increased by 1 to i and the back-off counter is uniformly chosen from 0 to $W_i - 1$ accordingly. Let

$$b_{i,n} = \lim_{t \to \infty} P\{s(t) = i, c(t) = n\}, \quad i = 0, \ldots, m, n = 0, \ldots, W_i - 1 \qquad (4.64)$$

denote the stationary distribution of the Markov chain. It has been shown [Bia00] that all the values $b_{i,n}$ can be expressed as functions of the value $b_{0,0}$ and of the conditional collision probability p, i.e.

$$b_{i,n} = \frac{W_i - n}{W_i} b_{i,0}, i \in [0, m], n \in [1, W_i - 1], \qquad (4.65)$$

$$b_{i,0} = \begin{cases} p^i b_{0,0}, & i \in [1, m-1] \\ \frac{p^m}{1-p} b_{0,0}, & i = m \end{cases}. \qquad (4.66)$$

By summarizing the probabilities of the states in which $c(t)$ is equal to zero, the probability of a user to transmit in a frame, denoted with p_{tx}, is given by

$$p_{\text{tx}} = P\{c(t) = 0\} = \sum_{i=0}^{m} b_{i,0} = \frac{b_{0,0}}{1 - p}. \qquad (4.67)$$

So far, the Markov chain model for the TBEB algorithm proposed in [Bia00] has been reviewed. However, this model cannot be directly applied to the considered system which is a frame-based system, as explained in the following.

In the considered frame-based system, there are two situations in which the back-off process will not continue, and so additional idle states have to be introduced correspondingly.

Firstly, after a transmission attempt, the user does not immediately know whether the transmission is successful or not until the beginning of the next frame, and so it will wait till the next frame to continue the back-off process. That is, suppose the transmission attempt happens in the n-th TO of the total N, the back-off process of that user will wait for $N - n$ TOs before continuing. Because the back-off counter is uniformly chosen from W_i values which is equal to multiples of N, the distribution of the number of TOs that the user will wait before the re-start of the back-off process is uniformly distributed over $[0, K - 1]$. Thus, the author of this thesis suggests, in [VZLT05], introducing $(N - 1)$ idle states, depicted in rectangles, in the Markov chain, as shown in Figure 4.14.
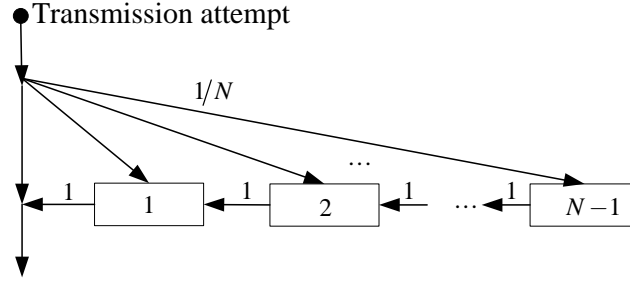
Figure 4.14: Markov chain model of the idle states after transmission attempt in the frame-based TBEB algorithm.

Let $x(t)$ denote the number of TOs to wait before the back-off process continues, the corresponding transition probabilities are

$$
\begin{cases}
P\{x(t+1) = n | x(t) = n+1\} = 1, & n = 1, \ldots, N-2 \\
P\{x(t+1) = n | \text{transmission attempt}\} = \frac{1}{N}, & n = 1, \ldots, N-1.
\end{cases}
\tag{4.68}
$$

With the probability of transmission attempt p_{tx}, the stationary distribution of the $(N-1)$ idle states after a transmission attempt

$$
x_n = \lim_{t \to \infty} P\{x(t) = n\}, \quad n \in [1, N-1]
\tag{4.69}
$$

is derived from (4.68) as

$$
x_n = \frac{N-n}{N} p_{\text{tx}}.
\tag{4.70}
$$

Secondly, after a successful transmission, the back-off processing will not restart if there is no BW-REQ to be transmitted. It is assumed that the BW-REQ is generated or updated in the beginning of each frame and include all the packets arrived in the last frame. Since new packet arrives in each frame with probability λ, after a successful transmission, the probability of a new generated BW-REQ is equal to λ. Thus, when there is no BW-REQ after a successful transmission, whose probability is $(1 - \lambda)$, the back-off process will be stopped for the current frame, i.e. the next N TOs. Therefore, N idle states, depicted in hexagons, are inserted after a successful transmission, as shown in Figure 4.15.

By letting $y(t)$ denote the number of TOs to wait before the back-off process continues after a successful transmission, the transition probability among the idle states in Figure 4.15 is

$$
\begin{cases}
P\{y(t+1) = N-1 | \text{successful transmission}\} = 1 - \lambda, \\
P\{y(t+1) = N-1 | y(t) = 0\} = 1 - \lambda, \\
P\{y(t+1) = n | y(t) = n+1\} = 1, \quad n \in [0, N-2]
\end{cases}
\tag{4.71}
$$
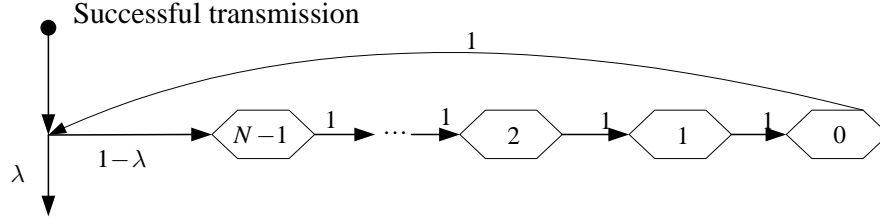
Figure 4.15: Markov chain model of the idle states after successful transmission in the frame-based TBEB algorithm.

With the probability of a successful transmission being equal to $p_{\text{tx}}(1 - p)$, the stationary distribution of the N idle states after a successful transmission

$$y_n = \lim_{t \to \infty} P\{y(t) = n\}, \quad n \in [0, N - 1] \tag{4.72}$$

is given by

$$\begin{cases} y_n &= y_0, \quad n \in [1, N - 1], \\ y_{N-1} &= (1 - \lambda) \cdot (y_0 + p_{\text{tx}}(1 - p)). \end{cases} \tag{4.73}$$

From (4.73), it is obtained that

$$y_n = \frac{1 - \lambda}{\lambda} p_{\text{tx}}(1 - p), \quad n \in [0, N - 1]. \tag{4.74}$$

By substituting the expression of p_{tx} in (4.67) into (4.70) and (4.74), the values x_n and y_n can be expressed as functions of the value $b_{0,0}$ and of the conditional probability p, i.e.

$$x_n = \frac{N - n}{N} p_{\text{tx}} = \frac{N - n}{N} \frac{b_{0,0}}{1 - p}, \quad n \in [1, N - 1] \tag{4.75}$$

$$y_n = \frac{1 - \lambda}{\lambda} b_{0,0}, \quad n \in [0, N - 1]. \tag{4.76}$$

Recall that the values $b_{i,n}$ can be also expressed as functions of the value $b_{0,0}$ and of the conditional probability p, as seen from (4.65) and (4.66). By imposing the normalization condition, i.e. the sum of the stationary probability of all states in the Markov chain is 1

$$1 = \sum_{i=0}^{m} \sum_{j=0}^{W_i - 1} b_{i,j} + \sum_{n=1}^{N-1} x_n + \sum_{n=0}^{N-1} y_n, \tag{4.77}$$

$b_{0,0}$ can then be expressed as a function of conditional collision probability p, i.e.

$$b_{0,0} = \frac{2(1 - p)}{(W_0 + 1) + pW_0 \frac{1 - (2p)^m}{1 - 2p} + (N - 1) + 2N \frac{1 - \lambda}{\lambda}(1 - p)}, \tag{4.78}$$

By substituting (4.78) into (4.67), the probability that a user transmits in a randomly chosen TO, p_{tx}, is expressed as a function of the conditional probability p, i.e.

$$p_{\text{tx}} = \frac{b_{0,0}}{1-p} = \frac{2}{(W_0+1) + pW_0\frac{1-(2p)^m}{1-2p} + (N-1) + 2N\frac{1-\lambda}{\lambda}(1-p)}. \qquad (4.79)$$

Furthermore, as the conditional collision probability p is equal to the probability that at least one of the $(N-1)$ other users transmit, it yields

$$p = 1 - (1 - p_{\text{tx}})^{N-1}. \qquad (4.80)$$

From equations (4.79) and (4.80), the two unknowns, the conditional collision probability p and the transmission probability p_{tx}, can be solved. Once the conditional collision probability p is known, the performance of the BW-REQ transmission in terms of the average delay can be calculated according to (4.58).

In order to verify the presented analytical derivation of the performance for random access, the analytically computed results are compared with the results obtained by means of numerical simulation. The performance in terms of average delay is depicted as a function of the arrival rate of the BW-REQ, λ, in Figure 4.16, for two different sets of parameters adjustable in TBEB algorithm, L and m. The analytical results are represented in curves, and the simulated results represented with markers are values averaged over 5000 frames. It can be seen in Figure 4.16(a) that, the analytical results meet the simulative ones quite well over different BW-REQ arrival rate λ ranging from 0 to 1.0. It is obvious that at high arrival rate of BW-REQ, polling is desired compared to random access. In Figure 4.16(b), the performance at the low arrival rate is reported in detail.

4.5.3 Impact of Parameters in Random Access on Performance

As shown in Section 4.5, the random access scheme is parameterized by L and m which determine the initial window size and the maximum back-off stage, respectively. In this section, the impact of parameters in random access is discussed and the method for optimizing the performance, i.e. minimizing the average delay of the BW-REQ transmission, is presented.

With the knowledge of the number K of users , the number N of TOs and the arrival rate λ of BW-REQ, the parameters for TBEB algorithm, L and m, should be selected in such a way that the average delay is minimized and the corresponding parameters are referred to as optimum parameters.

In Figure 4.17, the optimum parameters $(L_{\text{opt}}, m_{\text{opt}})$ are reported versus different BW-REQ arrival rate λ for a given K and N. It can be inferred that L_{opt} is always equal to
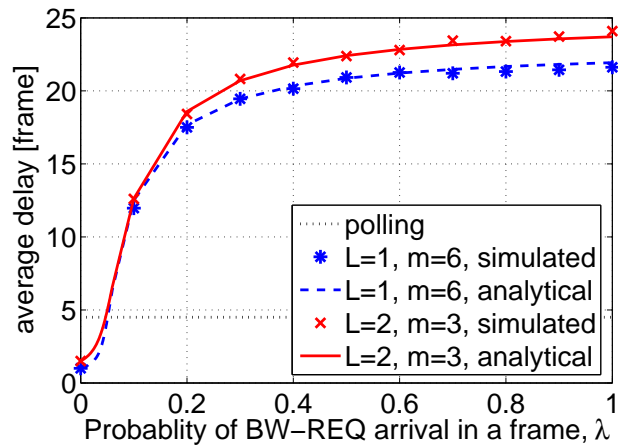
(a) The region with all possible BW-REQ arrival rates.
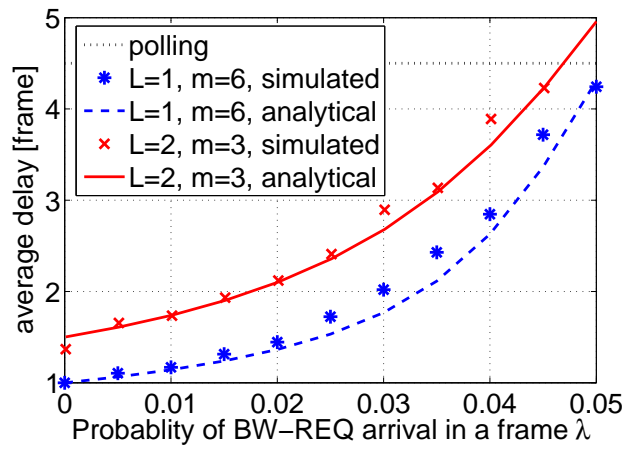
(b) The region with low BW-REQ arrival rates.

Figure 4.16: Comparison of the analytical results and the simulative results in terms of average delay, $K = 48$, N=6.

one and m_{opt} increases with the arrival rate λ. Thus, it can be concluded that in order to minimize the average delay, the initial window size should be always selected as small as possible, i.e. $L = 1$, while the maximum back-off stage should be selected according to the given scenario characterized by K, N and λ.
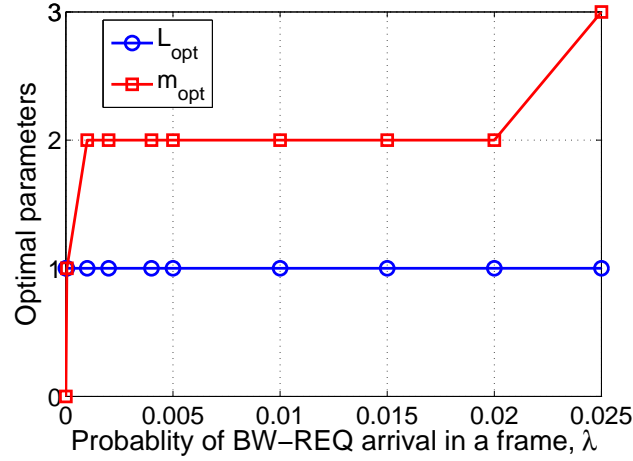
Figure 4.17: Optimum parameters of the TBEB algorithm for different BW-REQ arrival rates, $K = 48$, $N = 3$.

From (4.53), it can be seen that the average delay achieved by polling scheme only depends on the ratio between the number of users and the number of TOs, K/N, regardless the absolute values of N and K. In Figure 4.18, the average delay achieved by random access are reported for four cases:

(1). $K = 48$ and $N = 3$,

(2). $K = 16$ and $N = 1$,

(3). $K = 36$ and $N = 3$,

(4). $K = 12$ and $N = 1$.

In the first and the second cases, the ratio K/N is equal to 16, reported in solid curves in Figure 4.18, and in the third and the fourth cases, the ratio K/N is equal to 12, reported in dashed curves in Figure 4.18. It can be interfered from Figure 4.18 that the performance of random access with different K and N is different even with the same ratio K/N, however, the difference is quite small and could be neglected.
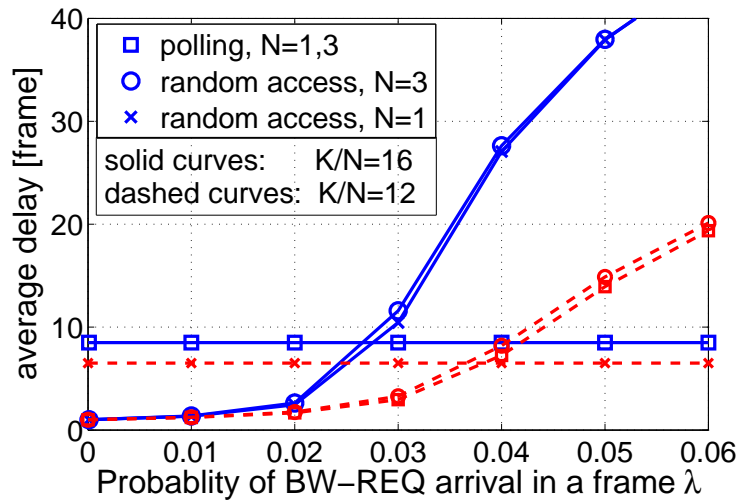
Figure 4.18: Performance comparison of random access with fixed ratio K/N, $L = 1$, $m = 4$.

4.5.4 An Efficient Grouping Mechanism for Random Access

In the section, a novel approach for random access proposed by the author of this thesis in [ZC06] is presented.

It is well known that high spectrum efficiency can be achieved by means of AMC, i.e. by adapting MCS to the user's channel condition, cf. Section 2.7. With different MCS, the transmission of one BW-REQ requires different amount of resources. Suppose that the AP is aware of the supportable MCSs of all users. In case of polling, as it is the AP that invites users to transmit, the AP can then correspondingly assign the required amount of resource for the user. However, in case of random access, all users have to use the same MCS, since they share the same contention period. In order to receive the BW-REQ from the cell border users which are usually penalized by large path-loss, the most robust MCS should be assigned for the contention period. The use of the same MCS by those users close to the BS, who could support a more efficient MCS, unnecessarily reduces the spectrum efficiency. In the following, a novel approach for random access is presented, which efficiently overcomes this limitation.

The main idea is that of dividing the total available resources in the contention period into several parts, e.g. groups of symbols, and assigning different MCSs to each part. The symbols in each group correspond to a certain number of TOs on the MAC layer as usual. In other words, all TOs are organized in groups and each group can be assigned with a certain MCS. To transmit a BW-REQ, each user first decides for the suitable MCS by e.g.

estimating its channel quality from the received broadcast signals, and then transmits the BW-REQ in the corresponding group of TOs. In this way, all users are allowed to transmit the request with the most efficient MCS they can support. Since all users are automatically divided into groups to perform BW-REQ transmission according to their channel qualities without any control from the AP side, no additional signaling is needed by the AP to group users.

Suppose that the contention period consists of R symbols and the length of one BW-REQ is B bits. Let then R_q be the number of symbols assigned for transmission at rate r_q by using the q-th MCS of totally Q. The number of TOs in the q-th part of the contention period is

$$N_q = \frac{R_q r_q}{B}, q = 1, \dots, Q. \tag{4.81}$$

It is inferred form Section 4.5.3 that the performance of random access in terms of the average delay is almost identical while the ratio between the number of users and the number of TOs is the same. In order to maintain the fairness among users, the average delay for all users are the same or at least similar. Therefore, by denoting with K_q the number of users transmitting at rate r_q, the following relationship should be maintained,

$$\forall q = 1, \dots, Q, \frac{K_q}{N_q} = \text{constant}. \tag{4.82}$$

According to relationships (4.81) and (4.82), the division of the contention period consisting of R symbols is obtained as

$$R_q = R \cdot \frac{\frac{K_q}{r_q}}{\sum_{j=1}^{Q} \frac{K_j}{r_j}}, q = 1, \dots, Q. \tag{4.83}$$

By denoting with $\tilde{K}_q$ the percentage of users that select the q-th MCS,

$$\tilde{K}_q = \frac{K_q}{K}, q = 1, \dots, Q, \tag{4.84}$$

and substituting it into (4.83), the number of symbols assigned for the q-th part of the contention period is given by

$$R_q = R \cdot \frac{\frac{\tilde{K}_q}{r_q}}{\sum_{j=1}^{Q} \frac{\tilde{K}_j}{r_j}}, q = 1, \dots, Q. \tag{4.85}$$

Following numerical example is presented to illustrate the advantage of the proposed approach. Consider a system with $K = 48$ users and half of them support 1/2-rate 16-QAM while half of them only support 1/2-rate QPSK, i.e. $\tilde{K}_1 = \tilde{K}_2 = 0.5$ and $r_1 = 1$,

$r_2 = 2$. It is then assumed that the contention period consists of $R = 144$ data symbols and the BW-REQ consists of $B = 48$ bits.

In case of conventional approach, the most robust MCS, i.e. 1/2-rate QPSK, is applied for the transmission, and so the contention period is organized in $N = 3$ TOs, shared by the 48 users. In the proposed approach, according to (4.85), the contention period is divided into two parts, $R_1 = 2R/3$ and $R_2 = R/3$, each organized into $N_1 = N_2 = 2$ TOs shared by $K_1 = K_2 = 24$ users.

In Figure 4.19, the proposed approach is compared to the conventional one in terms of average delay. The performance achieved by means of polling is also reported, which is independent to the arrival rate of the BW-REQ. From Figure 4.19 it can be interfered that for a given amount of resources, the proposed approach (solid curve with cycles) results in better performance, i.e. lower delay, than the conventional one (dashed curve with cycles).
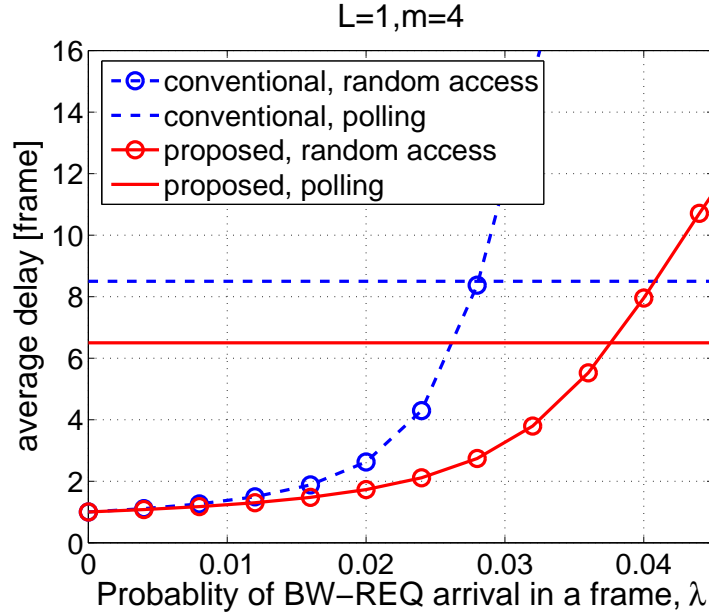
Figure 4.19: Performance comparison between the conventional and the proposed BW-REQ transmission schemes

Indeed, the proposed approach can not only reduce the average delay given a fixed amount of resources assigned for contention period, but also shorten the contention period while still fulfilling a certain delay requirement, as explained in the following. Let N be the number of required TOs to meet certain delay requirements for K users. In case of conventional approach, the data rate supported by all users $r_{\min}$ is used and so the number

of required resources in unit of data symbols R_0 is given by

$$R_0 = N \cdot \frac{B}{r_{\min}}. \tag{4.86}$$

In case of the proposed approach, by following the relationship

$$\frac{K_q}{N_q} = \frac{K}{N}, q = 1, \ldots, Q \tag{4.87}$$

in order to guarantee the delay requirements, the required amount of resources is equal to

$$\sum_{q=1}^{Q} R_q = \sum_{q=1}^{Q} N_q \cdot \frac{B}{r_q} = \sum_{q=1}^{Q} \tilde{K}_q N \cdot \frac{B}{r_q}. \tag{4.88}$$

Thus, compared to the conventional approach, the total required resources by the proposed approach can be reduced by

$$\frac{R_0 - \sum_{q=1}^{Q} R_q}{R_0} = 1 - r_{\min} \sum_{q=1}^{Q} \frac{\tilde{K}_q}{r_q} \tag{4.89}$$

while still achieving the same average delay.

For the previous example with parameters $\tilde{K}_1 = \tilde{K}_2 = 0.5$ and $r_1 = 1$, $r_2 = 2$, the required resources can be reduced by

$$1 - r_{\min} \sum_{q=1}^{Q} \frac{\tilde{K}_q}{r_q} = 25\%. \tag{4.90}$$

5 Adaptive Resource Allocation in a Single Relay-enhanced Cell

5.1 Introduction

In this chapter, a single REC consisting of one BS and several RNs is considered. Both BS and RN are referred to as AP, as they provide access to the users, but the BS is supposed to control all of the transmission in the cell.

In the current cellular system, a radio network controller (RNC) is the governing element and responsible for control of the BSs which are connected to it. The REC differs from the current cellular system in that whereas in current cellular system multiple BSs are wired-connected to the governing element, i.e. the RNC, in REC the multiple RNs communicate with the governing element, i.e. the BS, over wireless link. However, in both cases, there are multiple APs in the system. In this sense, the REC is actually very similar to the current cellular system.

As the number of APs in the system becomes large within fixed available radio frequency spectrum, the number of simultaneous links will become larger than the number of orthogonal resources the available bandwidth can provide. In order to provide service for a large population of users, the bandwidth used by the APs and users has to be reused in some clever way at the cost of co-channel interference. Such a system is said to be bandwidth- or interference-limited [Rap02]. Unlike thermal noise which can be overcome by increasing the SNR, co-channel interference cannot be combated by simply increasing the transmit power. This is because an increase in transmit power on one link increases the interference to other links using the same resources as well.

First-generation cellular systems used fixed frequency reuse, also referred to as fixed channel assignment (FCA). Cells sharing the same channel are separated by a minimum distance to provide sufficient isolation in order to reduce the co-channel interference [ZKA01]. Second-generation cellular systems use either FCA or random channel assignment [ZKA01]. The latter case achieves interference averaging by means of frequency hoping such as in GSM systems [MP92, Car94] or spread spectrum spread such as in DS/CDMA-based IS-95 systems [Sta93]. Also third generation systems, e.g. WCDMA systems, use interference averaging through spread spectrum technique [HT04].

Besides FCA and random channel assignment, dynamic channel assignment (DCA), which enables interference avoidance, has been proposed and studied for cellular systems. The principle of DCA is to track the channel-to-interference ratio and thereby assign users to channels with adequate quality either in a centralized or distributed way [ZKA01]. In centralized DCA schemes, a channel from the central pool is assigned to a user for temporary use by a centralized controller. In distributed schemes, channel assignment is performed independently in each AP. A comprehensive survey of different DCA schemes is provided in [KN96]. It was shown in [Pot95] that interference averaging techniques can perform better than fixed channel assignment techniques, whereas interference avoidance techniques can outperform interference averaging techniques by a factor of 2-3 in spectrum efficiency.

The performance of DCA schemes is critically dependent on the rate at which the resource assignment or re-assignment occurs [ZKA01]. To fully utilize the potential of DCA, channel reassignments must take place very frequently to track rapid changes of signal and interference levels in a mobile system. Indeed, channel variations, especially those caused by fast fading, are usually very fast. As a result, centralized DCA schemes adapted to such fast channel variations cause very high computational complexity as well as huge signaling burden for signal and interference measurements, which are generally infeasible in practical systems [ZKA01]. Completely distributed DCA schemes require much less signaling compared to centralized ones, but are problematic in practice due to collisions of channel assignment, i.e. the possibility for adjacent APs to independently select the same channel, thus causing unexpected interference when transmissions occur [ZKA01]. Collisions of channel assignment can be avoided by letting neighboring APs sequentially perform the DCA algorithm [CS00], but the resulting cycle of DCA will be too long to adapt to the rapid change of fast fading, which limits the DCA gain. When the DCA is performed fast enough to adapt to the fast fading, it is also referred to as adaptive resource allocation in literature.

In [LL03], a semi-distributed adaptive resource allocation scheme, which splits the resource allocation between RNC and BSs, is proposed for multi-cell OFDMA systems. The RNC makes the decision which resource unit, e.g. chunk, is assigned to which BS as well as the used transmit power on a long-term basis, e.g. at super-frame level. The BSs then make the decision which resource unit is assigned to which user on a short-term basis, e.g. at frame level. The RNC centrally controls a set of BSs and targets at the maximization of the system throughput and captures the interference avoidance gain by exploiting long-term channel knowledge which tracks the slow fading. Locally, the BS tracks the fast fading and makes the actual allocation for users according to the instantaneous SINR, so capturing the multi-user diversity gain. As RNC only requires the information on channel slow fading and makes the decision at a super-frame level, the rate of information exchange between RNC and BSs is significantly reduced.

Note that once the RNC has made the decision which resource is assigned to which BS

with a given transmit power, the interference from that BS to users served by other BSs is pre-determined in an OFDMA system. Thus, the BS can accurately predict the instantaneous SINRs without knowing the resource allocation of other BSs. However, when the BSs are equipped with multiple antennas, the co-channel interference changes with the used transmit beamforming vectors of the interfering BSs even under a stable channel and fixed transmit power. As the transmit beamforming vectors chosen for different users are generally different, the BS cannot estimate the instantaneous SINR without knowing the resource allocation of other BSs: some users can be "hit" by the beams of the neighboring BSs, whereas other users can be in a very favorable situation, depending on their channel conditions and the directions of the interferers [VTZZ06].

Just like in the cellular system, individual APs in the REC, including the BS and the RNs, cannot estimate the instantaneous SINR without knowing the resource allocation of other APs when the APs are equipped with multiple antennas.

By observing that the channel spatial structure is highly correlated over time and frequency, the AP can dynamically re-select on a long-term basis the proper transmit beamforming vectors for each served user while still achieving high beamforming gain, cf. Section 4.4.2. However, if the BS makes the actual allocation of individual users so as to determine also the transmit beamforming vectors used by the APs and only let the APs perform AMC according to accurate instantaneous SINRs, the multi-user diversity gain cannot be exploited since the APs are left with no choices on resource allocation among users.

In order to let the BS pre-determine the transmit beamforming vectors used by APs, while still enabling individual AP to exploit multi-user diversity, the author of this thesis proposes a two-level resource allocation approach in [CFRZ07, FRC$^+$07]. The two-level resource allocation reads as follows:

- On a long-term basis, each AP, including both BS and RN, groups users according to spatial correlation in such a way that users in the same group are highly correlated and then the BS makes the decision which resource unit, i.e. chunk, is used by which user group. Moreover, user groups may share the same chunk if the co-channel interference among them is sufficiently low.

- On a short-term basis, individual AP makes the adaptive allocation among users in the same group according to their instantaneous SINRs.

Due to the high spatial correlation, users in the same group can be efficiently served by the same transmit beamforming vector, and so such a user group is referred to as a logical beam hereafter. Indeed, the optimal transmit beamforming vectors of highly spatially correlated users are highly correlated, and it follows that the interference generated by the AP on users in other groups while serving users in the same logical beam with individual optimal beamforming vectors is also highly correlated. Hence, once the BS has made the

decision which resource is used by which logical beam, the co-channel interference and thus the instantaneous SINR can be accurately estimated at the APs even when individual optimal beamforming vectors are used for users in the same logical beam.

In other words, on a long-term basis, each AP, including both BS and RN, independently partition users into logical beams, and the BS performs the resource allocation among all logical beams and allows the logical beams with sufficiently low co-channel interference to share the same resource in spatial domain; on a short-term basis, each AP performs the adaptive resource allocation in time and frequency domains among users within each logical beam with respect to the independently measured instantaneous SINR values.

In Section 5.2, the algorithm used to construct the logical beams by grouping spatially correlated users is illustrated. In Section 5.3, the resource allocation approach for BS to assign logical beams is investigated. Finally, the performance of the proposed two-level resource allocation approach is assessed with numerical simulations in Section 5.4.

5.2 Construction of logical beams

5.2.1 Dynamic logical beams

In this section, algorithms of dynamically constructing logical beams with respect to the channel conditions are studied.

Intuitively, a simple greedy algorithm can be applied to construct logical beams. The AP randomly selects one user or selects the one with the best channel condition, and makes the selected user the first user of a logical beam. All remaining users whose spatial correlation with respect to the first user is above a certain threshold are then assigned to that logical beam. New logical beams are successively constructed and the construction ends when all users have been assigned.

This algorithm is very simple and computationally not intensive, but the result might be not very good. It can be imagined that, when most of the users have already been assigned, new beams will have to be generated in the remaining users, and the beam size is consequently too small. Furthermore, the starting point is chosen by considering a single user only, and it is not generally optimized with respect to the others.

For this reason, following the philosophy of sorting by merging [Knu98], a more advance bottom-up algorithm is proposed and described as follows:

1. Initially, each user is supposed to form an individual logical beam,

2. Iteratively, two logical beams with the highest spatial correlation are merged,

3. Finally, the algorithm ends when no more logical beams shall be merged, i.e. the correlation between any pair of logical beams is lower than a certain threshold.

The spatial structure of the channel is reflected by the channel correlation matrix [GSsS$^+$03]. Thus, the spatial correlation of user i and user j can be for instance defined as the amplitude of the inner product of the corresponding dominant eigenvectors $\mathbf{v}_{i,1}$ and $\mathbf{v}_{j,1}$ of the channel correlation matrix of user i and user j, respectively, i.e.

$$\rho_{i,j} = |\mathbf{v}_{i,1}^{\mathrm{H}} \mathbf{v}_{j,1}| \tag{5.1}$$

[GV96]. The spatial correlation between two logical beams is then further defined as the minimum of the spatial correlations between each pairs of users. As a result, the spatial correlation between logical beams A and B is expressed as

$$\rho_{A,B} = \min_{i \in A, j \in B} \rho_{i,j}. \tag{5.2}$$

For example, consider a five-user case. If the correlation matrix $\mathcal{R}$ with elements $[\mathcal{R}]_{i,j} = \rho_{i,j}$ is

$$\mathcal{R} = \begin{bmatrix} 1 & 0.9 & 0.2 & 0.7 & 0.1 \\ 0.9 & 1 & 0.3 & 0.8 & 0.2 \\ 0.2 & 0.3 & 1 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.5 & 1 & 0.4 \\ 0.1 & 0.2 & 0.6 & 0.4 & 1 \end{bmatrix} \tag{5.3}$$

and the correlation threshold is 0.4, then two logical beams, $\{1, 2, 4\}$ and $\{3, 5\}$, are constructed after three iterations by the proposed algorithm, as depicted in Figure 5.1.

5.2.2 Fixed logical beams

In this section, an algorithm of constructing fixed logical beams is presented.

Instead of the proposed dynamic construction of logical beams, logical beams can be also built in a fixed way. Indeed, dividing a cell into a number of the sectors can be viewed as a specific case of logical beam construction and each logical beam, i.e. sector, is then served by an individual sectorized antenna. Moreover, if the AP is equipped with ULA instead of sectorized antennas, fixed logical beams can be also built based on GoB. In case of GoB, a fixed grid of transmit beam that will be used by APs is pre-determined and users select the best beams independently, cf. Section 4.4.3. Thus, users selecting the same beam are inherently grouped together, forming one logical beam.

Compared to dynamic logical beams, fixed logical beams are not adapted to the actual users' propagation conditions and so the resulting performance is expected to be worse
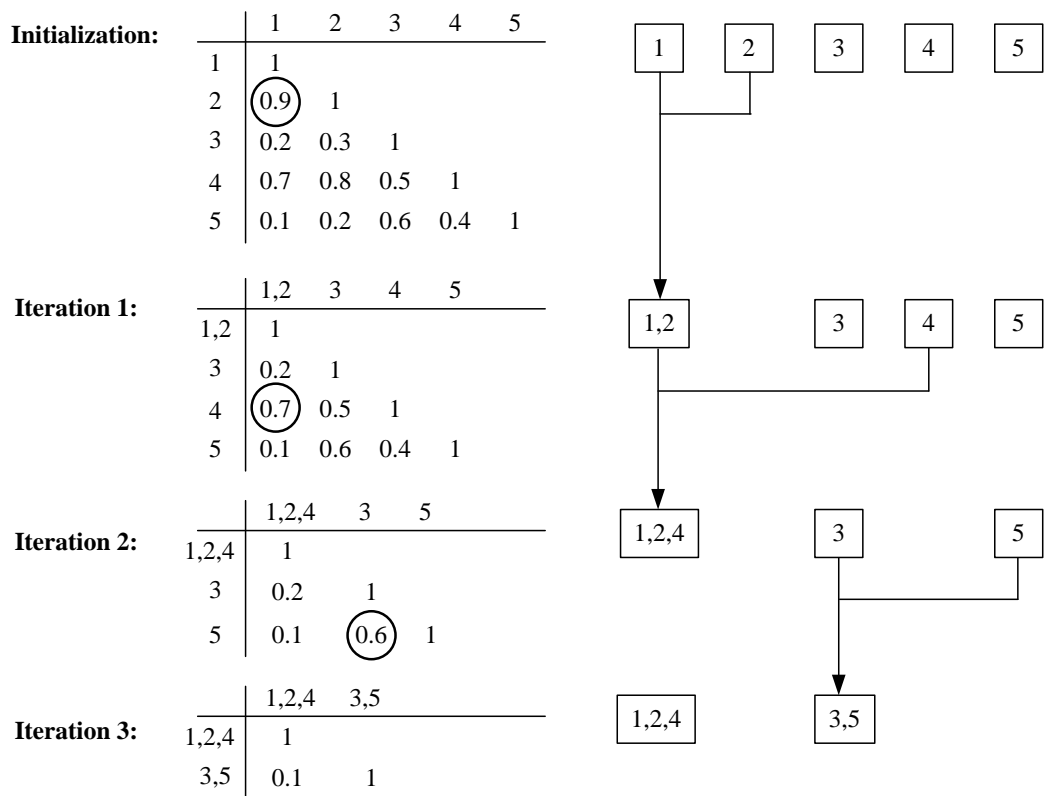
Figure 5.1: An example of logical beam construction in a five-user case.

than that of dynamic logical beam construction. For an example, supposing ten users are distributed in a way as depicted in Figure 5.2, the constructed logical beam by following pre-defined six sectors is obviously not so reasonable as the five dynamically constructed logical beams.

Fixed Logic Beams   Dynamic Logic Beams
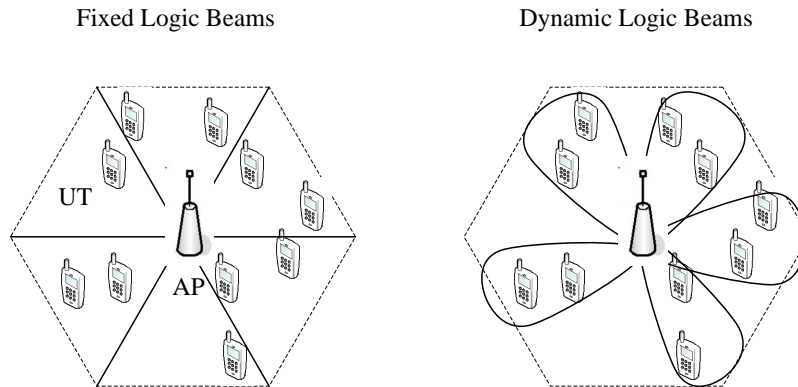
Figure 5.2: Comparison between the fixed and the dynamic logical beam constructions.

5.3 Adaptive Resource Allocation of logical beams

5.3.1 Introduction

Once the logical beams are independently constructed by each AP, the BS optimizes the resource allocation of logical beams by which the resources are assigned to the logical beams instead of individual users. The resource allocation is supposed to adapt to current interference scenario. Interference avoidance gain is expected to be achieved by letting logical beams with low co-channel interference share the resources, and by avoiding logical beams with high co-channel interference to transmit simultaneously.

In a REC, users either directly communicate with the BS by means of single-hop transmission or communicate with the BS via certain RN by means of two-hop transmission, as depicted in Figure 5.3. For the sake of clarity, the BS-to-RN link is referred to as *relay link*, while both BS-to-UT and RN-to-UT links are called *access links* hereafter.

As the relay link also passes data through the wireless interface, resources have to be allocated for both relay links and access links. For a two-hop communication, the end-to-end throughput of a certain user is equal to the lower one between the throughput of the relay link and the access link. Hence, on the one hand, the resources assigned to the relay link must be enough to forward all data scheduled for the access link; on the other hand, if
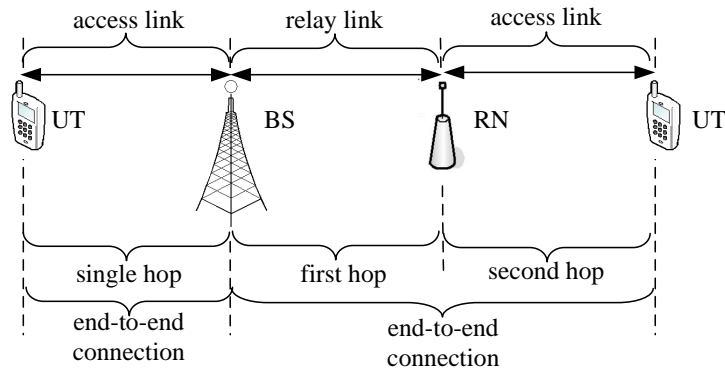
Figure 5.3: Illustration of access link and relay link in a REC.

too many resources are assigned to the relay link, some of them will not be used, resulting in a waste. Therefore, the resources adaptively allocated for relay links and access links shall be somehow balanced such that the end-to-end throughput is maximized and no waste occurs.

In summary, the proposed adaptive resource allocation for a REC has the following two objectives:

- to enhance the spectrum efficiency by letting multiple logical beams simultaneously transmit on the same resources if the co-channel interference is sufficiently low;

- to maximize the end-to-end throughput by balancing the resources assigned to the first and second hops, i.e. relay links and access links.

The rest of this section is organized as follows. Section 5.3.2 derives an algorithm for adaptively building groups of the logical beams. By assuming spatial-reuse among beams in the same group, the proposed algorithm groups the logical beams in such as way that the spectrum efficiency, i.e. total throughput of the group, is maximized. In Section 5.3.3 and Section 5.3.4, two adaptive resource allocation algorithms are proposed. Both algorithms are able to keep the resource balance between the first and second hops. Moreover, the first one focuses on achieving the maximum cell throughput, while the second one also maximizes the cell throughput but additionally aims at a fair resource allocation among users.

5.3.2 Resource Sharing among logical beams

In this section, the resource sharing among logical beams, thus achieving high spectrum efficiency, is discussed.

It is obvious that, when certain logical beams have low co-channel interference with each other, resource sharing among them should be allowed in the resource allocation in order to achieve high spectrum efficiency. In other words, the optimization of resource sharing can be regarded as the creation of groups of spatially compatible logical beams which share the same resources. In the following, in order to qualify the performance of the resource sharing, an estimate of achievable rate is derived for each logical beam and is proposed to be used as the performance metric.

In order to estimate the achievable rate of a logical beam, the SINR of the individual users in the logical beam has to be firstly derived.

It is assumed that the long-term channel information in terms of the correlation matrix of the channel between any user and the AP is available at the BS. Thus, the transmit beamforming vectors for individual users are also known by the BS as they are supposed to be determined based on the channel correlation matrix. For instance, by using eigen-beamforming, the transmit beamforming vector is equal to the dominant eigenvector of the channel correlation matrix of the user [Hay96].

Let P_i and $\mathbf{w}_i$ denote the transmit power and the transmit beamforming vector for user i, respectively. By letting $\mathbf{R}_i$ denote the channel correlation matrix of user i, the signal strength received by user i, denoted with S_i, can be written as

$$S_i = P_i \cdot \mathbf{w}_i^{\mathrm{H}} \mathbf{R}_i \mathbf{w}_i. \tag{5.4}$$

Moreover, by letting $\mathbf{R}_i^{(j)}$ denote the correlation matrix of the channel between user i and the AP serving user j, the amount of interference that is injected by user j on user i, denoted with $I_{i/j}$, can be expressed as

$$I_{i/j} = P_j \cdot \mathbf{w}_j^{\mathrm{H}} \mathbf{R}_i^{(j)} \mathbf{w}_j. \tag{5.5}$$

Suppose that logical beam A shares the resources with logical beam B. As the BS has no information about the resource assignment for individual users in the logical beam, it is suggested to let the BS estimate the interference received by user i in logical beam A from users in logical beam B, $I_{i/B}$, by the average of the interference injected in user i by all users in logical beam B, i.e.

$$I_{i/B} = \frac{1}{|B|} \sum_{j \in B} I_{i/j}. \tag{5.6}$$

From (5.4) and (5.6), the SINR of user i in logical beam A when logical beam A and B share the same resource, denoted with $\mathrm{SINR}_{i/B}$, is given by

$$\mathrm{SINR}_{i/B} = \frac{S_i}{I_{i/B} + \sigma_z^2} \tag{5.7}$$

with σ_z^2 being the noise power.

Under the assumption of independent data streams, the total interference from more than one interfering logical beam is simply the sum of all the interference. For example, by supposing logical beam A, B and C to share the same resource, the SINR of user i in logical beam A, denoted with $\mathrm{SINR}_{i/(B,C)}$, is equal to

$$\mathrm{SINR}_{i/(B,C)} = \frac{S_i}{I_{i/B} + I_{i/C} + \sigma_z^2}. \tag{5.8}$$

Once the SINR value $\mathrm{SINR}_{i/(B,C)}$ is calculated, the achievable data rate by user i when sharing resources with users in logical beams B and C, denoted with $r_{i/(B,C)}$, can be determined according to the assumed AMC scheme, cf. Section 2.7.

By letting R_i indicate the data rate request of user i, the amount of resources required by user i with a rate of $r_{i/(B,C)}$ to meet its data rate request is $\frac{R_i}{r_{i/(B,C)}}$. By defining the data rate request of logical beam A as the sum of the data rate request of all users in that beam, i.e. $\sum_{i \in A} R_i$, the rate of logical beam A, defined as the ratio between the data rate request and the required resources, is given by

$$r_{A/(B,C)} = \frac{\sum_{i \in A} R_i}{\sum_{i \in A} \frac{R_i}{r_{i/(B,C)}}}. \tag{5.9}$$

Finally, the total rate achieved by the group of logical beams which share the same resource is simply equal to the sum of the rate achieved by each logical beams in the group. For example, by supposing resource sharing among the group consisting of logical beams A, B and C, the total rate of that group $r_{\{A,B,C\}}$ is given by

$$r_{\{A,B,C\}} = r_{A/(B,C)} + r_{B/(C,A)} + r_{C/(A,B)}. \tag{5.10}$$

Based on the derived performance metric in terms of total rate for any group of the logical beams, a very simple yet efficient method for creating a group achieving highest performance can be derived following the best-fit principle. The algorithm starts with an empty group. At each iteration, the best logical beam that gives the highest increase of the total rate is added into the group. The iteration ends when the performance in terms of total rate decreases by adding one more logical beam. Let $\mathcal{A} = \{A_i\}_{i=1}^N$ be the candidate set initially consisting of all N logical beams and $\mathcal{G}$ be the created group of logical beams. The algorithm can be described as follows:

1. Initially, set $\mathcal{G} = \emptyset$

2. While $\mathcal{A} \neq \emptyset$

a) For all logical beam A_i in $\mathcal{A}$

    i. Calculate the rate achieved if adding A_i to the current group $\mathcal{G}$, $r_{\mathcal{G} \cup \{A_i\}}$

    ii. If the achievable rate decreases, i.e. $r_{\mathcal{G} \cup \{A_{i*}\}} \leq r_{\mathcal{G}}$, remove the logical beam A_i from the candidate set, i.e. let $\mathcal{A} = \mathcal{A} - \{A_{i*}\}$

b) Find logical beam A_j satisfying $r_{\mathcal{G} \cup \{A_j\}} \geq r_{\mathcal{G} \cup \{A_i\}}$ for all $A_i \in \mathcal{A}$

c) Add the found logical beam A_j into the group, i.e. $\mathcal{G} = \mathcal{G} \cup \{A_j\}$

5.3.3 Chunk-by-Chunk Balancing (CCB)

In this section, an algorithm is proposed for the resource allocation at the BS. The proposed algorithm aims at maximizing the cell throughput by successively allocating chunks to the group of logical beams having the highest total rate as far as none of the logical beam meets its rate request. It further keeps the balance of the resource allocation between the first and second hops by making sure that enough resources are assigned for the relay links after each chunk assignment on the logical beams, i.e. the access links, and so referred to as Chunk-by-Chunk Balancing (CCB).

In particular, the proposed algorithm consists of having a stepwise alternating allocation between the access links and the relay links. The pseudo-code description of the CCB algorithm is reported in Figure 5.4.

Before starting the actual allocation, the relay links are partitioned into spatially compatible groups. The links in the same group generate low co-channel interference to each other and so they can transmit simultaneously and are separated in spatial domain. The groups are built in such a way that the spectrum efficiency is maximized and these groups remain unchanged during the rest of the algorithm, cf. Step A in Figure 5.4. Moreover, the logical beams consisting of all access links are gathered into a list. In the initial stage, no chunk is assigned to either access links or relay links.

In the allocation procedure, the BS firstly uses the logical beam grouping algorithm derived in Section 5.3.2 to identify the group of logical beams that achieves the highest total data rate from all available logical beams in the list, cf. Step B in Figure 5.4.

The BS then starts a chunk-by-chunk assignment of all available chunks to the identified group of logical beams, i.e. the access links, and their corresponding relay links until either the rate request of any logical beam in the group has been satisfied or there are no more chunks available. During this phase, the group of logical beams firstly receives an allocation of a single chunk, cf. Step C in Figure 5.4. It is then checked whether the resources allocated to the corresponding relay link are enough to forward the data. If the condition is not satisfied, the required additional chunks are given to that relay link, cf. Step D in Figure 5.4. In such a way, the chunks are successively assigned to the identified group of logical beams as far as none of the logical beams in that group meets its rate

CCB-Allocation(*NumChunks, Beams, FirstHopLinks*)

A:

    FirstHopGroups ← GroupFirsthopLinks(*FirstHopLinks*)
    NumFreeChunks ← *NumChunks*
    BeamsList ← *Beams*
    AssignedChunks[*i*] ← 0 for all *i* in *FirstHopGroups*

    while *NumFreeChunks* > 0 **and** *BeamsList* ≠ **Null**

B:

        AllocationGroup ← GetBestGroup(*BeamsList*)
        AssignedChunks[*AllocationGroup*] ← 0
        BeamFinished ← **false**
        while *NumFreeChunks* > 0 **and** *BeamFinished* ≠ **true**

C:

            AssignedChunks[*AllocationGroup*] ←*AssignedChunks*[*AllocationGroup*] +1
            NumFreeChunks ← *NumFreeChunks* - 1
            for *L* **in** *FirstHopGroups*

D:

                do if RequiredChunks(*L*)>*AssignedChunks*[*FirstHopGroups*]
                  then *AssignedChunks*[*FirstHopGroups*] ← RequiredChunks(*L*)
                    NumFreeChunks ← *NumFreeChunks* - RequiredChunks(*L*)

            for *C* **in** *AllocationGroup*

E:

                do if RequiredChunks(*C*) <= *AssignedChunks*[*AllocationGroup*]
                  then RemoveFromList(*BeamsList, C*)
                    BeamFinished ← **true**

Figure 5.4: Pesudo-code description of the CCB algorithm.

request. Once the rate request of any of the logical beams is satisfied, the logical beam is removed from the list, cf. Step E in Figure 5.4.

The allocation procedure, first logical beam grouping and then the chunk-by-chunk resource assignment to the group, will continue till no logical beams is remained in the list.

By always allocating chunks to the group of logical beams that achieves the highest data rate, the CCB algorithm maximizes the cell throughput and the mutual interference diversity has been fully exploited. Furthermore, at each step, each allocation on the access links corresponds to an equivalent allocation on the relay links, which balances the allocation between the first and second hops and thus guarantees the efficient usage of resources for the end-to-end connection.

It is noticed that the static groups of the relay links at the very beginning makes the CCB algorithm suboptimal. In fact, it is likely that at each step some of the relay links might receive more resources than they actually need, since all relay links belonging to the same group receive the same amount of resources as the most demanding one. This effect, however, becomes relevant only in the case of highly unbalanced allocations, i.e. non uniform traffic distribution. Furthermore, adaptively grouping the relay links according to the beams allocation would also lead to suboptimal solutions, since it would force the creation of groups according to the instantaneous needs, which could be later changed by a forthcoming allocation.

5.3.4 Iterative Independent Balancing (IIB)

The CCB algorithm proposed in the previous section maximizes the cell throughput by always allocating resources to logical beams with highest data rate but tends to lead to unfairness among users. In this section, an algorithm is proposed to maintain the fairness among all users in addition to the maximization of cell throughput by optimized spatial reuse.

In the proposed algorithm, the dynamic allocation of the access and relay links are independently performed using an iterative procedure, and therefore it is referred to as iterative independent balancing (IIB), The pseudo-code description of the IIB algorithm is reported in Figure 5.5.

The same procedure is used to independently allocate resources to the relay links and the access links, i.e. the logical beams. The actual allocation function is similar to the one used for assigning resources to the logical beams in the CCB, but in this case the allocations for the logical beams and for the relay links are separately performed, cf. Step A and Step B in Figure 5.5. This allows the dynamic grouping of both the access links and the relay links according to their actual requests and mutual interference conditions, thus improving the efficiency in the use of resources.

IIB-Allocation(*NumChunks, Beams, FirstHopLinks*)

---

        finished ← **false**
        while (*finished* == **false**)
            ClearAllocation()
A:

            finished ← ResourceAllocation(*NumChunks, Beams*)
            NumFreeChunks ← GetNumFreeChunks()
B:

            finished ← ResourceAllocation(*NumFreeChunks, FirstHopLinks*)
C:

            if (*finished* == **false**)
            then
                for k **in** all users
                        do *RateRequest*[k] ← *RateRequest*[k] x *ScalingFactor*

Figure 5.5: Pesudo-code description of the IIB algorithm.

In the initial stage, it is assumed that the data rate requests of all users needs to be satisfied, and the rate requests on the corresponding relay links are also calculated accordingly. The allocation of the access links and the relay links can then be carried out using the allocation function *ResourceAllocation*. The allocation is considered to be failed when the number of the required chunks exceed the available one, and it follows that the rate requests of all users are proportionally scaled down and another iteration of the algorithm is executed, cf. Step C in Figure 5.5.

The details of the allocation function *ResourceAllocation* can be explained by the pseudo-code description reported in Figure 5.6. For the allocation of the access links and the relay links, the links in the list are referred to as all logical beams and all relay links, respectively. The allocation procedure is similar to the one used in the CCB in the sense that it always allocates chunks to the group of links with the highest total rate, and removes the links from the list when the its rate request is satisfied. The difference is that, instead of receiving one chunk after another, the group receives in one step all chunks required to satisfy the rate request of at least one link in that group. The allocation is considered to be successful and the function returns a value of true in the case that the amount of chunks required to satisfy the rate request of all links does not exceed the available chunks. Otherwise, the allocation is considered to be failed and the function returns a value of false.

The IIB algorithm, as opposed to the CCB, provides a fair resource assignment to the logical beams and thus to all the users. Moreover, the IIB algorithm also avoids unneces-

ResourceAllocation(*NumChunks, List*)

> *RemainingChunks* ← *NumChunks*
> **while** *RemainingChunks* >= 0
> > **if** List == **Null**
> > **then return true**
> > *AllocationGroup* ← GetBestGroup(*List*)
> > *AssignedChunks*[*AllocationGroup*] ← RequiredChunks(*AllocationGroup*)
> > *RemainingChunks* ← *RemainingChunks* - RequiredChunks(*AllocationGroup*)
> > **for** *L* **in** *AllocationGroup*
> > > **do if** RequiredChunks(L) <= *AssignedChunks*[*AllocationGroup*]
> > > **then** RemoveFromList(*List, L*)
> > **return false**

Figure 5.6: Pesudo-code description of the allocation procedure in the IIB algorithm.

sary allocation to the relay links, thus leading to a higher cell throughput.

5.4 Performance Assessment

5.4.1 Simulation Setup

In this section, the setup of the numerical simulation for the evaluation of the proposed adaptive resource allocation in a REC is presented.

The downlink transmission of a relay-enhanced hexagonal cell with a radius of 500 meters is simulated. Six RNs are uniformly located on a circle centered at the BS, whose radius is $2/3$ of the cell radius. 800 users are distributed in the whole cell area. For each snapshot of the simulation, the user positions are randomly generated according to uniform distribution. Each AP, including both BS and RN, is equipped with a 4-element ULA, which has an antenna spacing of half-wavelength and an elevation antenna gain of 14 dB. The UT is equipped with one directional antenna with an elevation gain of 0 dB. The values of the maximum transmit power of the BS and the RNs are 46 dBm and 40 dBm, respectively. By letting $P_{\text{tot},m}$ and N_c denote the maximum transmit power of AP m and the number of chunks in the frequency direction, the transmit power of AP m for each user is supposed to be constant and equal to

$$P_m = \frac{P_{\text{tot},m}}{N_c}.$$ (5.11)

In the channel model, path-loss and shadow-fading are generated according to (2.1) by considering an urban macro scenario. Multi-path propagation is not modeled.

By knowing the transmit power, the path-loss and shadow-fading, the average receive power of each user can be determined. It is further assumed that each user communicates with the AP that provides the highest receive signal power.

The spatial structure of the channel between user k and AP m is characterized by the LOS angle between them, denoted with $\theta_{k,m}$. The channel between user k and AP m is denoted with $\mathbf{h}(\theta_{k,m})$. Furthermore, the spatial correlation between user i and user j users served by the same AP m is given by

$$\rho_{i,j} = \mathbf{h}(\theta_{i,m})^{\mathrm{H}} \cdot \mathbf{h}(\theta_{j,m}). \tag{5.12}$$

By assuming the spatial correlation $\rho_{i,j}$ only depends on the difference in the LOS angle $\Delta\theta_{i,j} = \theta_{i,m} - \theta_{j,m}$ and the number M_t of transmit antennas at the AP, it can be written as

$$\rho_{i,j} = \begin{cases} \frac{1}{M_t} \left| \frac{1-\exp(j\pi M_t \sin \Delta\theta_{i,j})}{1-\exp(j\pi \Delta\theta_{i,j})} \right|, & 0 < |\Delta\theta_{i,j}| \leq \pi/2 \\ 0, & |\Delta\theta_{i,j}| > \pi/2 \\ 1, & |\Delta\theta_{i,j}| = 0 \end{cases}, \tag{5.13}$$

which is depicted in Figure 5.7.
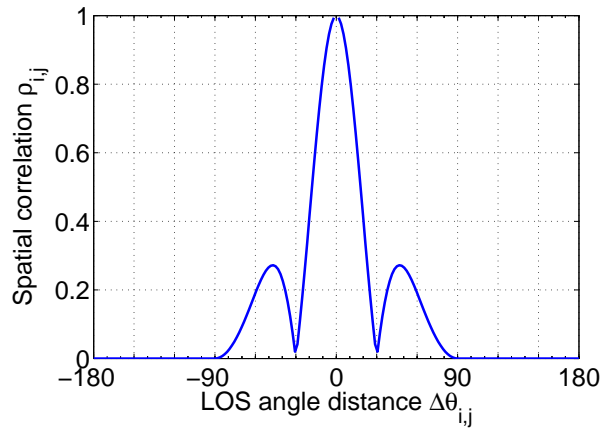
Figure 5.7: User spatial correlation as a function of LOS angle distance.

In case of dynamic logical beam construction, by knowing the spatial correlation $\rho_{i,j}$ between any pair of users, each AP constructs logical beams using the algorithm proposed in Section 5.2. The threshold of the spatial correlation is set to 0.4.

In case of fixed logical beam construction, by assuming $N = 6$ logical beams per AP, each AP divides its sub-cell into N sectors with angular size of $2\pi/N$ and users locating in the same sector form a logical beam.

It is further assumed that each AP always transmits user data by means of eigenbeam-forming, i.e. using a transmit beamforming vector directing towards the user [Hay96], as depicted in Figure 5.8 .
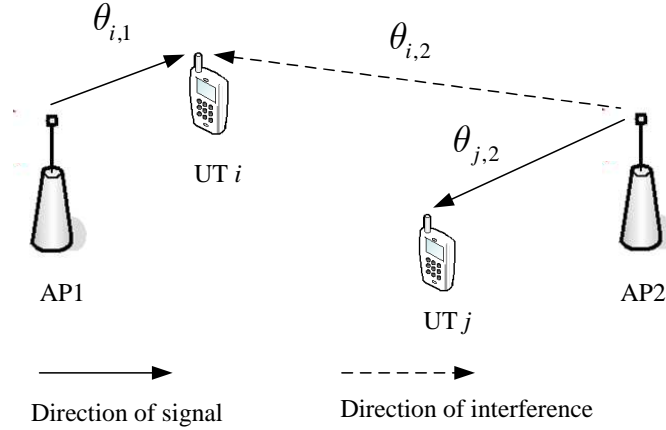
Figure 5.8: Illustration of the co-channel interference.

From Figure 5.8, it can be seen that the data symbols of user i are transmitted from AP 1 using beamforming vector $\mathbf{h}(\theta_{i,1})/|\mathbf{h}(\theta_{i,1})|$, and so the signal power received by user i, S_i as defined in (5.4), is given by

$$S_i = P_m \cdot \|\mathbf{h}(\theta_{i,1})\|^2 = P_m \cdot M_t. \tag{5.14}$$

It can be further derived from Figure 5.8 that, when the data symbol of user j is trans-mitted over the beamforming vector $\mathbf{h}(\theta_{j,2})/|\mathbf{h}(\theta_{j,2})|$, the amount of interference received by user i, $I_{i/j}$ as defined in (5.5), is equal to

$$I_{i/j} = P_m \cdot \left| \frac{\mathbf{h}^{\mathrm{H}}(\theta_{j,2})}{|\mathbf{h}(\theta_{j,2})|} \cdot \mathbf{h}(\theta_{i,2}) \right|^2 = P_m \cdot M_t \cdot \rho_{j,i}^2. \tag{5.15}$$

From (5.14) and (5.15), the estimated receive SINRs of individual users and, in turn, the estimated rate for each logical beam can be derived at the BS, cf. Section 5.3.2. The adaptive resource allocation algorithms, CCB and IIB, can then be carried out accordingly.

Since the multi-path propagation, i.e. fast fading, is not modeled, the frame-basis re-source allocation at each AP is not implemented in the simulation. Indeed, multi-user diversity gain is achieved by means of adaptive OFDMA among the users belonging to the same logical beam. Moreover, the achieved multi-user diversity gain also depends on the size of the logical beam. The larger the size of the logical beam, the higher the multi-user diversity gain can be exploited. In order to have a more meaningful performance evalu-ation, it is proposed to introduce the multi-user diversity gain when evaluating the final

end-to-end throughput. Define the multi-user diversity gain G_{MUD} as the ratio between the channel gain achieved by adaptive OFDMA G_{adapt} and the average channel gain achieved by fixed or random allocation G_{fix},

$$G_{\mathrm{MUD}} = \frac{G_{\mathrm{adapt}}}{G_{\mathrm{fix}}}. \tag{5.16}$$

Thus, by taking adaptive OFDMA into account, the SINR of user k in logical beam A is higher than the one used in the adaptive resource allocation algorithm SINR_i as defined in (5.7), and is given by

$$\mathrm{SINR}_i^{\mathrm{MUD}} = \mathrm{SINR}_i \cdot G_{\mathrm{MUD}}(|A|), \tag{5.17}$$

with $|A|$ being the number of users in logical beam A. The SINR value $\mathrm{SINR}_i^{\mathrm{MUD}}$ in (5.17) is used in evaluating the achieved end-to-end throughput.

From Appendix A.3, the multi-user diversity gain G_{MUD} is a function of the number K of user, i.e.

$$G_{\mathrm{MUD}}(K) = 1 + \frac{1}{2} + \cdots + \frac{1}{K}, \tag{5.18}$$

as seen in (A.30) and depicted in Figure 5.9.
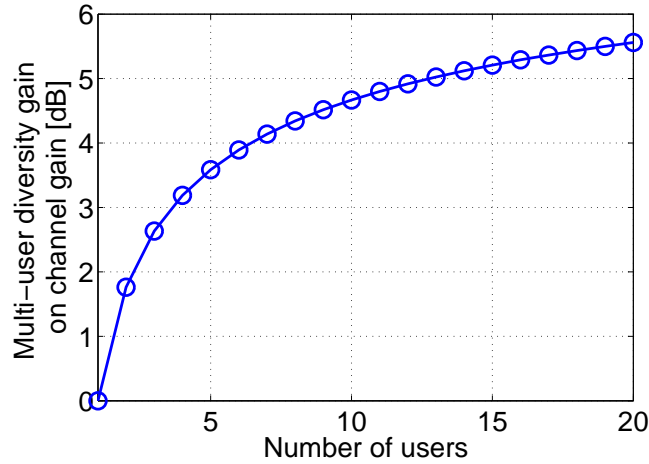
Figure 5.9: Multi-user diversity gain achieved by adaptive resource allocation.

5.4.2 Numerical Results

In this section, the performance of the CCB and IIB algorithms in combination with the fixed and the dynamic logical beam construction methods are presented.

In Figure 5.10, the performance comparison between CCB and IIB algorithms is shown for both fixed and dynamic logical beam construction, in terms of end-to-end cell throughput as well as user fairness metrics Jain's Index as defined in (3.24).
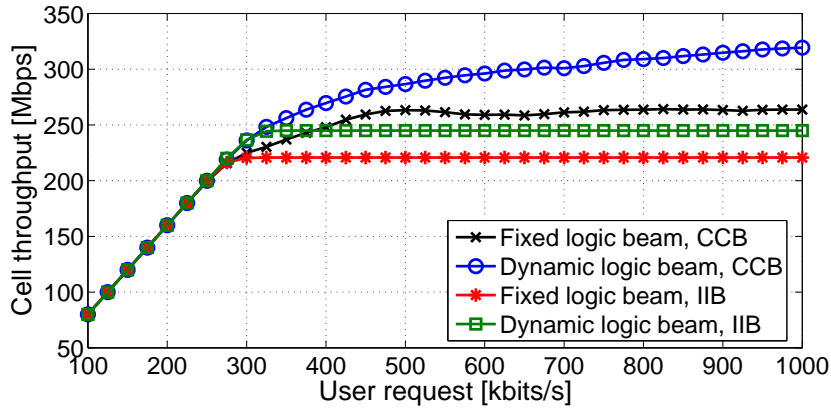
(a) Cell throughput
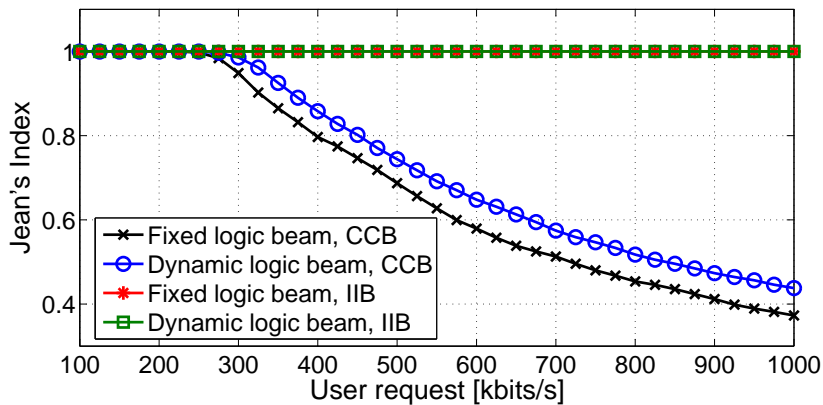
(b) Fairness, Jain's index

Figure 5.10: Comparison of the algorithms for adaptive resource allocation in a REC in terms of cell throughput and Jain's index.

It can be observed in Figure 5.10(a) that the cell throughput achieved adopting the CCB algorithm can be significantly higher than that of the IIB. The higher throughput of the CCB is, however, obtained at the expense of fairness. It can be seen in Figure 5.10(b) that the the Jain's index is always close to 1 for the IIB algorithm, which indicate a complete fair scenario, while CCB is unable to guarantee fairness when the offered traffic is above the saturation point of the system.

The highest cell throughput, together with the fact that it does not take user fairness into account, makes the CCB algorithm most appropriate for best-effort traffic. The IIB

algorithm, on the other hand, is more appropriate delay-sensitive traffic.

For the comparison between the fixed and dynamic logical beam construction, first focus on the performance achieved by the CCB algorithm. From Figure 5.10(a), it can be seen that, while the throughput of the fixed logical beam construction (curves with crosses) almost saturates just above 260 Mbps, using the dynamic logical beam construction (curves with cycles) it keeps growing approximately linearly, although with a lower rate. Due to the better exploitation of the spatial reuse, the dynamic logical beam construction outperforms the fixed logical beam construction. Same conclusion can be made by comparing the performance of the fixed and dynamic logical beam construction achieved by the IIB algorithm: the IIB algorithm with dynamic logical beam (curves with squares) saturates at a higher cell throughput than that with fixed logical beam construction (curves with stars).

In summary, the following conclusions can be derived from the presented numerical results:

- Compared to fixed logical beam construction, the proposed dynamic approach allows significant performance improvement by adapting the construction of beams to the propagation condition of all users.

- Both CCB and IIB algorithms optimize the spatial reuse among logical beams and balance the resource allocation for the first and second hops.

- Compared to the CCB algorithm, the IIB algorithm provides complete fairness in resource allocation among all users at the cost of relatively lower cell throughput.

6 Conclusions

Beyond third generation (B3G) mobile communication systems are expected to provide a variety of services such as voice, image and data transmission with different quality of service (QoS) and rate requirements for "anytime-anywhere". OFDM to exploit the frequency selective channel property, MIMO to attain high spectral efficiency and relaying to combat path-loss and shadowing at high frequency are considered as the key technologies for B3G systems. Moreover, adaptive resource allocation is becoming more important, as it can significantly improve the system spectral efficiency by adapting the radio resource allocation to varying channel fading, interference scenario and traffic load. This thesis has studied the downlink (DL) adaptive resource allocation for multi-user MIMO OFDMA system with fixed relays.

Adaptive resource allocation in a single cell without inter-cell interference has been investigated in Chapter 3. By assuming ideal transmit side channel state information (CSI), concurrent transmission of multiple users on the same time-frequency resources, i.e. spatial division multiple access (SDMA), is enabled by means of the zero-forcing beamforming (ZFBF) technique. Two kinds of optimization problems, namely power minimization and rate maximization, have been formulated and studied. Joint approach that performs joint optimization of resource allocation in both frequency and spatial domains has been promoted. In particular, low computational complexity sub-optimal algorithms have been proposed due to the need for fast adaptation in practical systems. In case of power minimization problem, the proposed successive bit insertion (SBI) algorithm iteratively raises the user data rate while maximizing the power efficiency and guaranteeing the user priority. The performance achieved by the proposed SBI algorithm has been analyzed and the results have shown that the performance only slightly degrades from the optimal solution obtained by means of exhaustive searching. In case of rate maximization problem, in order to take into account the user fairness, the objective functions for the two commonly used user fairness strategies, i.e. proportional fairness (PF) and max-min fairness (MMF), have been derived, and variants of the SBI algorithm have been then proposed so as to maximize the relative objective functions. Additionally, while the SBI algorithm allows adaptive power allocation in the spatial domain using successive rate increase, another sub-optimal algorithm, named successive user insertion (SUI) algorithm, has been proposed for the scenario where equal power sharing in the spatial domain is assumed. The SUI algorithm iteratively adds users to resources while keeping equal power sharing on each resource. From numerical results, it has been concluded that, (1) the proposed joint approach outperforms the

existing disjoint approach, (2) the SBI algorithm achieves higher spectrum efficiency than the SUI algorithm by allowing adaptive power allocation in the spatial domain at a cost of more iterations, and (3) compared to the PF strategy, the MMF strategy sacrifices cell throughput for better user fairness.

Compared to fixed resource allocation, adaptive resource allocation requires additional signaling for the acquisition of channel and traffic knowledge as well as for the delivery of allocation results from the base station (BS) to all users. As a result, the overhead caused by the additional signaling mitigates the adaptation gain, i.e. the gain obtained by adaptive resource allocation, and so with the purpose of maximizing the system performance in terms of effective throughput, the signaling required for the adaptive resource allocation has been analyzed and optimized in Chapter 4, which are briefly summarized in the following.

A chunk consisting of adjacent sub-carriers and OFDM symbols has been considered as the basic resource unit in the adaptive resource allocation. While the signaling overhead caused by the delivery of allocation results decreases linearly with the chunk dimension, the adaptation gain is expected to decrease with the chunk dimension but not significantly thanks to the existing high correlation of channel fading among adjacent sub-carriers and OFDM symbols. The choice of chunk dimension requires a trade-off between the reduction of the overhead and the loss of the adaptation gain. An analytical study has been presented in Section 4.2, from which it can be seen that, the adaptation gain depends on the chunk dimension as well as the channel correlation functions. By defining the relative chunk dimension as the chunk dimension normalized by the coherence time and the coherence bandwidth, the adaptation gain is derived as a decreasing function of the relative chunk dimension. By using the effective throughput as evaluation metric, the optimum chunk dimension can be identified for different system parameters and propagation scenarios. Compared to indoor environment, the coherence bandwidth in outdoor environment is smaller due to larger delay spread, resulting in a higher loss in adaptation gain for the same chunk dimension, and so the optimum chunk dimension is smaller. Similarly, the optimum chunk dimension is smaller at higher velocities because of the smaller coherence time.

To cope with the time-variant property of the channel fading, the users' CSI needs to be periodically updated at the BS so as to adapt the resource allocation to the currant channel status. In Section 4.3, the optimum update interval has been derived by evaluating the effective throughput, which takes into account both performance loss due to outdated CSI and overhead reduction. By defining the relative update interval as the ratio between the update interval in unit of frames and the coherence time, the throughput has been approximated by a linear decreasing function of the relative update interval, which is valid for any velocity, based on the numerical results. The effective throughput has then been expressed as a function of the velocity and the overhead caused by one CSI update, based on which the optimum channel update interval that maximizes the effective throughput has been analytically derived. The optimum channel update interval has shown to be inversely proportional to the square root of the velocity, and so at a lower velocity the optimum up-

date interval is larger, i.e. less frequent CSI update is required. Furthermore, the optimum relative channel update interval has been shown to be proportional to the square root of the velocity, and so the optimum relative channel update interval is smaller at a lower velocity, resulting in less CSI mismatch and thus lower loss of adaptation gain. Indeed, at low velocity, the channel does not change rapidly, and so the CSI does not need to be updated very often. It follows that, at low velocities, the overhead is not so relevant and the loss of adaptation gain becomes relatively more critical. Hence, at a lower velocity a lower loss in the adaptation gain should be ensured by decreasing the relative update interval.

The aforementioned ZFBF requires users' instantaneous channel knowledge in terms of full channel matrices to be available at the transmitter, which causes a huge amount of overhead. As presented in the previous discussion, more frequent update of channel knowledge is required at higher velocities. When the velocity increases to some extent, the adaptation gain may not be able to compensate the loss due to the increased signaling overhead any more. It has been shown in literature that there are other beamforming techniques requiring less channel knowledge compared to ZFBF. Two of them, namely Generalized eigenbeamforming (GenEigBF) and fixed grid-of-beamforming (GoB), have been addressed in Section 4.4. GenEigBF enables adaptive SDMA based on the knowledge of the users' channel correlation matrices, and fixed GoB enables adaptive SDMA based on the knowledge of the index of the users' best beam. When instantaneous CQI is available at the BS, resource allocation can be optimized in both frequency and spatial domains using the proposed joint approach. In case of fixed GoB, the instantaneous CQI, in terms of users' receive SINR values of the best beams on each chunk, can be measured and then reported to the BS by all users. However, in case of GenEigBF, users cannot measure the receive SINR values, because the transmit beamforming vectors on each chunk, which are optimized according to the users served on that chunk, are not known until the AP performs the allocation. Therefore, two methods have been proposed to estimate the receive SINRs from the channel correlation matrices and instantaneous CQI in terms of the channel norms. While the first method generally yields a conservative estimate of the receive SINRs and consequently results in a conservative allocation by ignoring the multiple receive antennas, the second method generally produces an aggressive estimate of the receive SINRs and consequently results in an aggressive allocation by including the gain from multiple receive antennas based on the approximated channel matrices. The results have shown that GenEigBF based on conservative SINR estimate achieves higher throughput than GenEigBF based on aggressive SINR estimate, thus indicating that a conservative resource allocation is beneficial. This is because although more bits are transmitted following the aggressive allocation, the code word error rate is too high due to the overestimated SINR, leading to lower throughput. As GenEigBF always adapts the transmit beamforming vectors to the channel correlation matrices of individual user group, it is expected to provide better performance than fixed GoB, which has been observed in low and moderate SNR region from the simulation. However, it has been shown that in high SNR region, e.g. a SNR being greater than 20 dB, GenEigBF does not exhibited any gain over fixed GoB.

This is because that GenEigBF suffers from non-ideal adaptive selection of modulation and coding schemes due to the estimation error in receive SINR, especially in high SNR region, while fixed GoB does not. Since the signaling overhead caused by CQI acquisition is much less than that caused by CSI acquisition, GenEigBF and fixed GoB are suggested at high velocities when ZFBF is not beneficial due to too high signaling overhead.

While the change of traffic load of DL transmission is known at the BS, the change of traffic load of uplink transmission can only be known by the BS from a certain control message such as bandwidth request (BW-REQ) transmitted by users. BW-REQ can be transmitted by using random access. Moreover, the transmission on a random access channel is normally based on slotted ALOHA protocol in conjunction with the truncated binary exponential back-off algorithm. In Section 4.5, the performance of BW-REQ transmission, in terms of the delay between the arrival and the successful transmission of the BW-REQ, has been addressed. By means of Markov chain, the average delay has been analytically derived as a function of the number of users, the BW-REQ arrival rate, the number of transmission opportunities, and the two parameters in the back-off algorithm, namely the initial back-off window size and the maximum back-off stage. It can then be concluded that, in order to minimize the delay, the initial back-off window size should be always selected as small as possible, e.g. one frame, while the maximum back-off stage should be optimized with respect to the number of users, the number of transmission opportunities and the BW-REQ arrival rate. Additionally, a novel grouping mechanism has been proposed, which enables more efficient usage of the resources by letting different users choose different transmission mode, i.e. modulation and coding scheme, with respect to their own channel conditions to transmit BW-REQ. In particular, all available transmission opportunities are organized into groups, and each group is assigned with a different transmission mode. Transmission opportunities assigned with higher order of transmission mode occupy less resources. Thus, for a fixed amount of resources, more opportunities can be accommodated, leading to lower delay. Alternatively, with the proposed grouping mechanism, less resources are required to provide the same number of transmission opportunities, and so the signaling overhead is reduced while still guaranteeing the same delay. It is worthy to mention that with the proposal, users independently select the transmission mode and then transmit over the corresponding transmission opportunities without coordination from the BS, and so no additional signaling overhead is introduced.

Fixed relay nodes (RNs) have been shown to extend the coverage of the BS or enhance the cell-edge capacity by forwarding data between BS and users. In a relay enhanced cell, the multiple access points (APs), including one base station (BS) and multiple relay nodes (RNs), interfere with each other when transmitting on the same resource. Thus, the adaptive resource allocation in a REC needs to take into account the interference among multiple APs, which has been investigated in Chapter 5. When complete centralized approach is infeasible in practical systems due to extremely high computational complexity and huge signaling for the exchange of instantaneous channel information, a two-step approach has been proposed for the adaptive resource allocation in the REC. Two kinds of

links can be identified in the REC, namely the access links, i.e. the links between APs and users, and the relay links, i.e. the links between BS and RNs. On a long-term basis, all APs independently group together the user access links with high spatially correlations, forming the so-called logical beams, and then the BS decides for each chunk which logical beams are served on it according to the long-term average interference as well as the traffic load. On a short-term basis, each AP decides for each chunk which user of the logical beam transmits. Since users belonging to the same logical beam exhibit high spatial correlation, the interference to other users generated by serving them is also highly correlated, which allows individual APs to accurately estimate the inter-cell interference on a short-term basis. By following the proposed two-step approach, mutual interference diversity gain is exploited by means of dynamic logical beam construction at each AP and dynamic logical beam allocation at the BS, and multi-user diversity gain is exploited by means of adaptive user allocation at the AP based on instantaneous CSI. The proposed algorithm for logical beam construction dynamically creates logical beams with respect to the channel spatial structure of all users, and it is shown to outperform a fixed construction approach. Furthermore, two algorithms, chunk-by-chunk balancing (CCB) algorithm and iterative independent balancing (IIB) algorithm, have been proposed to the allocation of logical beams. Both of them target at optimizing the resource sharing among logical beams and guaranteeing the balancing of resource assignment between the access links and the relay links as well. The proposed CCB algorithm maximizes the cell throughput, while the proposed IIB algorithm provides fairness among users.

Indeed, users in the cellular system can be classified as inner cell users that are near to the serving AP, and outer cell users located at the border of the coverage area of serving AP and adjacent APs. Inner cell users generally experience very low and negligible inter-cell interference, and thus the joint approach proposed for the single AP scenario can be applied for the adaptive resource allocation of inner cell users. On the other hand, outer cell users experience strong inter-cell interference, and so the two-step approach proposed for multiple APs scenario should be applied in order to cope with the inter-cell interference and gain from mutual interference diversity.

A Appendix

A.1 Derivation of the Local Variance of the channel coefficient within Chunk

In this appendix, the derivation of the local variance of the channel coefficient $h_{n,t}^{(k)}$ is presented.

Suppose $x(n)$ is a stationary random process with mean value m_x and auto-covariance $C(\Delta n)$. Define the block averaging value s as the mean value over $[0, P-1]$,

$$s = \frac{1}{P} \sum_{p=0}^{P-1} x(p). \tag{A.1}$$

The mean value of s is

$$m_s = E\left\{ \frac{1}{P} \sum_{p=0}^{P-1} x(p) \right\} = \frac{1}{P} \sum_{p=0}^{P-1} E\{x(p)\} = m_x. \tag{A.2}$$

And the variance of s can be written as

$$
\begin{aligned}
\sigma_s^2 &= E\left\{ \left(\frac{1}{P} \sum_{p_1=0}^{P-1} x(p) \right)^2 \right\} - m_s^2 \\
&= \frac{1}{P^2} \sum_{p_1=0}^{P-1} \sum_{p_2=0}^{P-1} \left(x^*(p_1) x(p_2) - m_x^2 \right) \\
&= \frac{1}{P^2} \sum_{p_1=0}^{P-1} \sum_{p_2=0}^{P-1} C(p_2 - p_1).
\end{aligned}
\tag{A.3}
$$

Since for the value of $(p_2 - p_1)$ there are P '0's, $(P-1)$ '1's and '-1's, ..., 1 '$P-1$' and '$-(P-1)$', it follows

$$\sigma_s^2 = \frac{1}{P} \sum_{\Delta n = -(P-1)}^{P-1} \left(1 - \frac{|\Delta n|}{P} \right) C(\Delta n). \tag{A.4}$$

The channel fading, $h_{n,t}^{(k)}$, is a stationary two-dimensional random process. Let $R_f(\cdot)$ and $R_t(\cdot)$ denote the auto-correlation function of channel fading $h_{n,t}^{(k)}$ in time and frequency direction, respectively. Note that the mean value of $h_{n,t}$ is zero and so its auto-covariance is the same as auto-correlation.

Ω in (4.11) is the variance of the block averaging $\bar{h}_m^{(k)}$ defined in (4.9). Following the similar procedure described above for one-dimensional case, Ω can be expressed as

$$
\begin{aligned}
\Omega \quad = \quad & \frac{1}{n_{\text{sub}}} \sum_{p=-(n_{\text{sub}}-1)}^{n_{\text{sub}}-1} \left(1 - \frac{|p|}{n_{\text{sub}}} \right) R_f(nf_s) \cdot \\
& \cdot \frac{1}{n_{\text{symb}}} \sum_{q=-(n_{\text{symb}}-1)}^{n_{\text{symb}}-1} \left(1 - \frac{|q|}{n_{\text{symb}}} \right) R_t(qT_s),
\end{aligned}
\tag{A.5}
$$

with f_s and T_s being the sub-carrier spacing and the OFDM symbol duration, respectively.

If the continuous auto-correlation function is known instead of discrete one, since channel fading is assumed to be flat within sub-carrier and constant over one OFDM symbol, Ω can be calculated by integral over continuous auto-correlation function as

$$
\begin{aligned}
\Omega \quad = \quad & \frac{1}{B_{\text{chunk}}} \int_{-B_{\text{chunk}}}^{B_{\text{chunk}}} \left(1 - \frac{|f|}{B_{\text{chunk}}} \right) R_f(f) df \cdots \\
& \cdot \frac{1}{T_{\text{chunk}}} \int_{-T_{\text{chunk}}}^{T_{\text{chunk}}} \left(1 - \frac{|t|}{T_{\text{chunk}}} \right) R_t(t) dt
\end{aligned}
\tag{A.6}
$$

with $B_{\text{chunk}} = n_{\text{sub}} f_s$ and $T_{\text{chunk}} = n_{\text{symb}} T_s$ representing the bandwidth and the time duration of the chunk, respectively.

From (A.6), Ω is known as a function of the chunk dimension and the auto-correlation function of the channel coefficient $h_{n,t}$. By assuming exponential power delay profile and Jakes Doppler spectrum, the auto-correlation functions of the channel fading are given by

$$
R(f,t) = R(f) \cdot R(t) = \frac{1}{1 + j2\pi\sigma_\tau f} \cdot J_0 \left(2\pi f_{\text{D,max}} t \right)
\tag{A.7}
$$

[Ste92], where σ_τ is the RMS delay spread, $f_{\text{D,max}}$ the maximum Doppler spread and $J_0(\cdot)$ is the zero-order Bessel function of the first kind.

As introduced in Section 2.3.3, the coherence bandwidth and the coherence time are defined as the reciprocal of the delay spread and the maximum Doppler spread, respectively. By substituting (2.17) and (2.21), the auto-correlation function of the channel coefficient in (A.7) becomes

$$
R(f,t) = \frac{1}{1 + j2\pi\frac{f}{5B_{\text{coh}}}} \cdot J_0 \left(2\pi \frac{t}{T_{\text{coh}}} \right).
\tag{A.8}
$$

By substituting the auto-correlation function described by (A.8) into (A.6) and further substituting the normalized chunk dimension $\tilde{B}_{\mathrm{chunk}}$ defined in (4.22) and the normalized chunk duration $\tilde{T}_{\mathrm{chunk}}$ defined in (4.23), Ω can be calculated by

$$
\begin{aligned}
\Omega \;=\; & \frac{1}{\tilde{B}_{\mathrm{chunk}}} \int_{-\tilde{B}_{\mathrm{chunk}}}^{\tilde{B}_{\mathrm{chunk}}} \left(1 - \frac{|f|}{n_{\mathrm{sub}}}\right) \frac{1}{1 + \frac{j2\pi}{5}f} df \cdots \\
& \cdot \frac{1}{\tilde{T}_{\mathrm{chunk}}} \int_{-\tilde{T}_{\mathrm{chunk}}}^{\tilde{T}_{\mathrm{chunk}}} \left(1 - \frac{|t|}{\tilde{T}_{\mathrm{chunk}}}\right) J_0(2\pi t) dt.
\end{aligned}
\tag{A.9}
$$

A.2 Derivation of the Optimal Update Interval

In the appendix, the proof of Theorem 1 stated in Section 4.3.3 is presented .

As both the throughput $\bar{\rho}_0$ and the effective throughput ρ should be positive, from

$$\bar{\rho}_0 = -\xi_1 \cdot \frac{T_{\mathrm{up}}T_{\mathrm{f}}}{T_{\mathrm{coh}}(v)} + \xi_0 \;\; > \;\; 0,$$

$$1 - \frac{\beta}{T_{\mathrm{up}}} \;\; > \;\; 0,$$

the reasonable update interval T_{up} is in the range of

$$\max(1, \beta) < T_{\mathrm{up}} < \frac{\xi_0 T_{\mathrm{coh}}}{\xi_1 T_{\mathrm{f}}}. \tag{A.10}$$

Since the effective throughput as a function of the update interval T_{up} in (4.32) is twice differentiable, its maximum can be determined by the second derivative test [AP72] as follows. From (4.32), the first and second derivative of the effective throughput ρ with respect to the update interval T_{up} are equal to

$$\frac{d\rho}{dT_{\mathrm{up}}} = -\frac{\xi_1 T_{\mathrm{f}}}{T_{\mathrm{coh}}} + \frac{\beta \xi_0}{T_{\mathrm{up}}^2}. \tag{A.11}$$

and

$$\frac{d^2\rho}{d^2 T_{\mathrm{up}}} = -\frac{2\beta \xi_0}{T_{\mathrm{up}}^3} \tag{A.12}$$

respectively. By letting the first derivative be equal to zero, the stable point in the range given by (A.10) is

$$T_{\mathrm{up}}^* = \sqrt{\frac{\xi_0 \beta T_{\mathrm{coh}}}{\xi_1 T_{\mathrm{f}}}}. \tag{A.13}$$

Moreover, it can be seen from (A.12) that the second derivative at the stable point T_{up}^* is negative, therefore the effective throughput ρ has a maximum at T_{up}^* according to the second derivative, which is equal to test [AP72].

Since the coherence time T_{coh} is a function of the speed of light c, the carrier frequency f_c and the velocity v, cf. (2.21), the optimal update interval that maximizes the effective throughput is equal to

$$T_{\mathrm{up,opt}} = T_{\mathrm{up}}^* = \sqrt{\frac{\xi_0 c}{\xi_1 T_{\mathrm{f}} f_c}} \cdot \sqrt{\frac{\beta}{v}}. \tag{A.14}$$

By substituting the optimal update interval $T_{\mathrm{up,opt}}$ in (A.14) into (4.32), the achieved maximum effective throughput is given by

$$\rho_{\mathrm{max}} = \rho(T_{\mathrm{up,opt}}) = \xi_1 T_{\mathrm{f}} \frac{f_c}{c} (\sqrt{\beta v})^2 - 2\sqrt{\xi_1 \xi_0 T_{\mathrm{f}} \frac{f_c}{c}} \sqrt{\beta v} + \xi_0. \tag{A.15}$$

In order to analyze the behavior of the maximum effective throughput $\rho_{\max}$ with respect to the overhead β and the velocity v, let's calculate the derivative of with $\rho_{\max}$ given by (A.15) with respect to $\sqrt{\beta v}$

$$\frac{d\rho_{\max}}{d(\sqrt{\beta v})} = 2\xi_1 T_{\mathrm{f}} \frac{f_c}{c} \sqrt{\beta v} - 2\sqrt{\xi_1 \xi_0 T_{\mathrm{f}} \frac{f_c}{c}}. \tag{A.16}$$

If $\beta \leq 1$, from (A.10) it follows

$$1 < \sqrt{\frac{\xi_0 c}{\xi_1 T_{\mathrm{f}} f_c}} \cdot \sqrt{\frac{\beta}{v}}, \tag{A.17}$$

and so

$$v < \frac{\xi_0 c \beta}{\xi_1 T_{\mathrm{f}} f_c}. \tag{A.18}$$

By substituting (A.18) into (A.16), the following inequality can be obtained

$$
\begin{aligned}
\frac{d\rho_{\max}}{d(\sqrt{\beta v})} &< 2\xi_1 T_{\mathrm{f}} \frac{f_c}{c} \sqrt{\beta \frac{\xi_0 c \beta}{\xi_1 T_{\mathrm{f}} f_c} \xi_0 c} - 2\sqrt{\xi_1 \xi_0 T_{\mathrm{f}} \frac{f_c}{c}} \\
&= 2\sqrt{\xi_1 \xi_0 T_{\mathrm{f}} \frac{f_c}{c}} (\beta - 1) \\
&\leq 0.
\end{aligned}
\tag{A.19}
$$

If $\beta > 1$, from (A.10) it follows

$$\beta < \sqrt{\frac{\xi_0 c}{\xi_1 T_{\mathrm{f}} f_c}} \cdot \sqrt{\frac{\beta}{v}}, \tag{A.20}$$

and so

$$v < \frac{\xi_0 c}{\xi_1 T_{\mathrm{f}} f_c \beta}. \tag{A.21}$$

By substituting (A.21) into (A.16), the following inequality can be obtained

$$\frac{d\rho_{\max}}{d(\sqrt{\beta v})} < 0. \tag{A.22}$$

Thus, the first derivative of the maximum effective throughput $\rho_{\max}$ with respect to $\sqrt{\beta v}$ is always negative for any value of $\sqrt{\beta v}$, from which it can be concluded that the maximum effective throughput $\rho_{\max}$ monotonically decreases with $\sqrt{\beta v}$.

A.3 Derivation of Multi-user Diversity Gain in Adaptive OFDMA

In this appendix, the multi-user diversity gain achieved by adaptive OFDMA, i.e. by assigning the chunk to the user with the highest channel gain among all K users, is derived as a function of the number of users.

Let $h_{n,t}^{(k)}$ denote the channel coefficient experienced by user k on chunk n of OFDM symbol t and suppose $h_{n,t}^{(k)}$ is a stationary two dimensional zero-mean Gaussian random process. Moreover, the channel coefficient of all users is assume to be independent and identically distributed (i.i.d).

The channel gain $|h_{n,t}^{(k)}|^2$ follows Rayleigh distribution [Pap65] and so the pdf and the cdf of the channel gain as the square value of the channel coefficient are given by

$$p_{|h_{n,t}^{(k)}|^2}(x) = e^{-x} \tag{A.23}$$

and

$$F_{|h_{n,t}^{(k)}|^2}(x) = 1 - e^{-x}, \tag{A.24}$$

respectively. By performing adaptive OFDMA, the resulting channel gain on chunk n of OFDM symbol t is the highest one among all K users,

$$|h_{n,t}^{\mathrm{adapt}}|^2 = \max\{|h_{n,t}^{(1)}|^2, \dots, |h_{n,t}^{(K)}|^2\}, \tag{A.25}$$

which is the K-th order statistic [Wei] and its pdf is given by

$$p_{|h_{n,t}^{\mathrm{adapt}}|^2}(x) = K \left(F_{|h_{n,t}^{(k)}|^2}(x) \right)^{K-1} \cdot p_{|h_{n,t}^{(k)}|^2}(x) = K e^{-x} \left(1 - e^{-x} \right)^{K-1}. \tag{A.26}$$

The mean value of the resulting channel gain from adaptive OFDMA $|h_{n,t}^{\mathrm{adapt}}|^2$ is then equal to

$$\begin{aligned} E\{|h_{n,t}^{\mathrm{adapt}}|^2\} &= \int_0^\infty K e^{-x} \left(1 - e^{-x} \right)^{K-1} dx \\ &= 1 + \tfrac{1}{2} + \cdots + \tfrac{1}{K} \end{aligned} \tag{A.27}$$

In case of fixed allocation, as the user assignment is independent of the channel conditions, the channel gain $|h_{n,t}^{\mathrm{fix}}|^2$ obtained with fixed allocation has the same distribution as the channel gain of any user $|h_{n,t}^{(k)}|^2$. Hence, its mean value is given by

$$E\{|h_{n,t}^{\mathrm{fix}}|^2\} = \int_0^\infty e^{-x} dx = 1. \tag{A.28}$$

Define the multi-user diversity gain G_{MUD} as the ratio between the channel gain achieved by adaptive OFDMA and the one achieved by fixed OFDMA, i.e.

$$G_{\mathrm{MUD}} = \frac{E\{|h_{n,t}^{\mathrm{adapt}}|^2\}}{E\{|h_{n,t}^{\mathrm{fix}}|^2\}}. \tag{A.29}$$

From (A.27) and (A.28), the multi-user diversity gain G_{MUD} can be expressed as a function of the number K of users ,

$$G_{\mathrm{MUD}}(K) = 1 + \frac{1}{2} + \cdots + \frac{1}{K}. \tag{A.30}$$

Nomenclature

Abbreviations

AMC   Adaptive Modulation and Coding, page 37

AoA   Angle of Arrival, page 19

AoD   Angle of Departure, page 19

AP   Access Point, page 4

AS   Angle Spread, page 24

AWGN   Additive White Gaussian Noise, page 27

B3G   Beyond third Generation, page 1

BC   Broadcast Channel, page 29

BCH   Broadcast Control Channel, page 34

BF   Beamforming, page 29

BLDPCC   Punctured Block Low-density Parity Check Code, page 38

BS   Base Station, page 2

BW-REW   Bandwidth Request, page 89

CCB   Chunk by Chunk Balancing, page 115

CCDF   Complimentary Cumulative Distribute Function, page 51

CDMA   Code Division Multiple Access, page 3

CIR   Channel Impulse Response, page 20

CP   Cyclic Prefix, page 26

CQI   Channel Quality Indication, page 9

CSI	Channel State Information, page 9
CTF	Channel Transfer Function, page 22
CWER	Code Word Error Rate, page 39
DCA	Dynamic Channel Assignment, page 11
DFT	Discrete Fourier Transform, page 25
DJ	Disjoint, page 60
DL	Downlink, page 2
DS	Delay Spread, page 22
FCA	Fixed Channel Assignment, page 11
FDD	Frequency Division Duplex, page 16
FDMA	Frequency Division Multiple Access, page 3
FEC	Forward Error Correction, page 37
FP	First Priority, page 50
GoB	Grid-of-Beamforming, page 14
ICI	Inter-carrier Interference, page 25
IDFT	Inverse Discrete Fourier Transform, page 25
IFFT	Inverse Fast Fourier Transform, page 26
IIB	Iterative Independent Balancing, page 117
ISI	Inter-symbol Interference, page 1
LOS	Line of Sight, page 120
MAC	Medium Access Control, page 37
MC	Multi-carrier, page 1
MCS	Modulation and Coding Scheme, page 8
MIB	Mutual Information per Bit, page 40
MIMO	Multi-input Multiple-output, page 2

MISO	Multiple Input Multiple Output, page 32
MMF	Max-Min fairness, page 5
MMSE	Minimum Mean Square Error, page 33
NLOS	Non Line of Sight, page 2
OFDM	Orthogonal Frequency Division Multiplexing, page 2
OFDMA	Orthogonal Frequency Division Multiple Access, page 3
Ovh	Overhead, page 72
P2S	parallel-to-serial, page 26
PAS	Power Azimuth Spectrum, page 24
PCPC	Per Chunk Power Constraint, page 45
PDF	Probability Density Function, page 22
PF	Proportional Fairness, page 5
PUPC	Per User Power Constraint, page 45
QoS	Quality of Service, page 3
RAC	Random Access Channel, page 34
REC	Relay-enhanced Cell, page 2
RMS	Root Mean Square, page 22
RN	Relay Node, page 2
RNC	Radio Network Controller, page 11
S2P	serial-to-parallel, page 26
SBI	Successive Bit Insertion, page 13
SDMA	Spatial Division Multiple Access, page 3
SINR	Signal-to-Interference and Noise Ratio, page 9
SNR	Signal to Noise Ratio, page 29
SUI	Successive User Insertion, page 13

SVD	Singular Value Decomposition, page 33
TBEB	Truncated Binary Exponential Back-off, page 10
TDD	Time Division Duplex, page 9
TDMA	Time Division Multiple Access, page 3
TO	Transmission Opportunity, page 90
UL	Uplink, page 8
ULA	Uniform Linear Array, page 16
UT	User Terminal, page 2
WP	Weighted Priority, page 50
ZFBF	Zero-forcing Beamforming, page 6

Operators

$(\cdot)^*$	complex conjugate, page 15
$(\cdot)^{*T}$	transpose, page 15
$(\cdot)^{-1}$	matrix inverse, page 15
$(\cdot)^{\mathrm{H}}$	complex Hermitian, page 15
$[\cdot]_{i,j}$	element in the i-th row and j-th column of a matrix in bracket, page 15
$\circledast$	circular convolution, page 27
$\odot$	element-wise multiplication, page 27
$\|\cdot\|$	Frobenius norm, page 15
$\|\cdot\|$	cardinality of a set, page 54
$\mathrm{diag}[\cdot]$	diagonal matrix containing elements in the argument, page 15
$E[\cdot]$	expectation, page 15
$\mathcal{F}[\cdot]$	Fourier transform, page 23
$\tilde{\mathcal{F}}[\cdot]$	discrete Fourier transform, page 27
$\tilde{\mathcal{F}}^{-1}[\cdot]$	inverse discrete Fourier transform, page 26

$\mathrm{tr}[\cdot]$ trace, page 15

Symbols

β fraction of the resources available in one frame required by one CSI update, page 75

$\mathcal{B}_n$ set of beams in the fixed grid of beams on chunk n, page 85

$\tilde{B}_{\mathrm{chunk}}$ chunk bandwidth normalized by channel coherence bandwidth, page 71

B_{chunk} chunk bandwidth, page 71

B_{coh} coherence bandwidth, page 23

B_k relative data rate of user k with respect to its rate request, page 49

$\bar{C}_{\mathrm{tot,eff}}^{\mathrm{adapt}}$ effective average capacity achieved by adaptive resource allocation, page 72

$\bar{C}_{\mathrm{tot}}^{\mathrm{adapt}}$ average of total capacity achieved by adaptive resource allocation, page 71

$\bar{C}_{\mathrm{tot}}^{\mathrm{fix}}$ average of total capacity achieved by fixed resource allocation scheme C, page 68

$\bar{C}_{\mathrm{tot}}^{\mathrm{S}}$ average of total capacity achieved by allocation scheme C, page 68

c speed of light, page 20

$C_m^{(k)}$ average capacity of user k on chunk m, page 69

$C_{n,t}^{(k)}$ Shannon capacity of user k on sub-carrier n of OFDM symbol t, page 68

$\bar{d}$ average delay of BW-REQ transmission using random access scheme, page 92

$\bar{d}^{(\mathrm{Poll})}$ average delay of BW-REQ transmission using polling scheme, page 90

$\delta_{k,j,n,q}$ binary variable describing the resource allocation, page 47

$\hat{\mathbf{d}}_k$ estimate of the transmit signal intended for user k $\mathbf{d}_k$, page 30

$\mathbf{d}$ vector of transmit signals of all users, page 31

$\mathbf{d}_k$ vector of transmit signal intended for user k, page 30

D distance between two adjacent base stations, page 16

d distance between transmitter and receiver, page 18

$\Delta P_{k,n}$	power increase of user k on chunk n, page 49
$\overline{\Delta r}$	rate increase, page 49
$\mathbf{F}_k$	receiver filter or receive beamforming matrix of user k, page 30
$\mathbf{f}_{k,n}$	receive beamforming vector of user k on chunk n, page 81
$f(c)$	required SINR for the reliable transmission at a rate of c, page 38
$f_{\mathrm{D,max}}$	maximum Doppler frequency, page 20
f_c	carrier frequency, page 20
$f_{\mathrm{D},m,n}$	Doppler frequency for the m-th sub-path of the n-th path, page 20
γ	path-loss exponent, page 18
Γ_r	set of supported data rates, page 38
$\mathbf{g}$	discrete channel impulse response, page 27
G_{Rx}	antenna gain of receiver, page 21
G_{Tx}	antenna gain of transmitter, page 21
G_l	effective channel gain of l-th data symbol in case of ZFBF, page 31
$G_{k,j,n}$	effective channel gain of user k on chunk n given that user j is also served on chunk n, page 47
$G_{k,n}$	channel gain on chunk n of user k, page 38
$g_{k,n}$	channel norm of user k on chunk n, page 83
$g_{u,s,n}$	amplitude of the CIR on the n-th path between antenna pair (u,s), page 21
$g_{u,s}(\tau, t)$	CIR between antenna pair (u,s) as function of delay τ and time t, page 21
$\bar{h}_m^{(k)}$	local mean of the channel coefficient of user k on chunk m, page 69
$\mathbf{H}$	channel matrix of all users, page 31
$\mathbf{h}$	discrete channel transfer function, page 27
$\mathbf{h}_{k,i}^{\mathrm{MISO}}$	the i-th equivalent MISO channel of channel matrix $\mathbf{H}_k$, page 33
$\mathbf{H}_k$	channel matrix of user k, page 30

$h(f,t)$	channel frequency response at frequency f and time t , page 22
h_n	channel transfer function on sub-carrier n, page 27
x_n	transmit signal on sub-carrier n, page 27
y_n	receive signal on sub-carrier n, page 27
z_n	noise on sub-carrier n, page 27
$\mathbf{I}$	unitary matrix, page 32
$I_{i/B}$	interference received by user i in logical beam A from the interfering logical beam B, page 113
$I_{i/j}$	interference injected by user j on user i , page 113
$I_{k,n}$	power of interference on chunk n of user k, page 38
$J(x)$	Jain's index, page 54
K	number of users, page 29
k^*	index of the selected user, page 49
n^*	index of the selected chunk, page 49
λ	probability of BW-REQ arrival, page 90
$\lambda_{k,i}$	the i-th eigenvalue of channel correlation matrix $\mathbf{R}_k$, page 84
L	initial back-off window in units of frames, page 91
L	total number of data symbols, page 29
L_k	number of data symbols transmitted by user k, page 29
m	maximum back-off stage, page 91
M_r	total number of receive antennas, page 29
M_t	number of transmit antennas, page 29
M_{r_k}	number of receive antennas at user k, page 29
N	number of sub-carriers, page 25
N	number of transmission opportunities, page 90
n_{sub}	number of sub-carriers per chunk, page 67

n_{symb}	number of OFDM symbols per chunk, page 67
N_c	number of chunks, page 38
$N_{k,n}$	power of noise on chunk n of user k, page 38
$\phi_{m,n}$	AoD of the m-th sub-path in the n-th path, page 19
$\Psi_{k,n}$	cost function in optimization problem, page 49
$\psi_{m,n}$	initial phase for the m-th sub-path in the n-th path, page 21
$\varphi_{m,n}$	AoA of the m-th sub-path in the n-th path, page 19
p	conditional collision probability, page 92
$p(\theta)$	power azimuth specturm, page 24
P_{chunk}	transmit power available for each chunk, page 54
P_{tot}	total available transmit power, page 54
p_{tx}	probability of a user to transmit in a frame for random access, page 94
P_k	transmit power allocated to user k , page 113
P_n	power of the n-th path, page 19
$P_{k,n}$	transmit power on chunk n of user k, page 38
Q	number of supported data rates, page 38
$\mathbf{r}$	time-domain receive signal in MC systems, page 27
$\mathbf{R}_k$	channel correlation matrix of user k, page 82
$\mathbf{R}_i^{(j)}$	correlation matrix of the channel between user i and the AP serving user j, page 113
ρ	effective data throughput, page 75
ρ_0	data throughput, page 75
ρ_{max}	maximum data throughput achieved with the optimum interval of CSI update, page 79
$\rho_{A,B}$	spatial correlation between logical beam A and B, page 109
$\rho_{i,j}$	spatial correlation between user i and j, page 109

r_{ctl}	rate of control data transmission, page 72
$R_g(\Delta d)$	correlation function of the CIRs at two antenna elements spaced by Δd, page 24
$R_g(\tau)$	auto-correlation function of the CIR $g(\tau, t)$ as a function of delay difference $\Delta \tau$ at $\Delta t = 0$, page 22
$R_g(\tau; \Delta_t)$	auto-correlation function of the time-variant CIR $g(\tau, t)$ as a function of delay difference $\Delta \tau$ and time difference Δt, page 22
$R_g(\tau_1, \tau_2; t_1, t_2)$	auto-correlation function of the time-variant CIR $g(\tau, t)$, page 22
$R_h(\Delta f; \Delta t)$	auto-correlation function of the CTF $h(f, t)$ as a function of frequency difference Δf and time difference Δt, page 22
$R_h(f_1, f_2; t_1, t_2)$	auto-correlation function of the CTF $h(f, t)$, page 22
R_k	the data rate request of user k, page 46
$r_k(t)$	data rate granted to user k in frame t, page 53
$r_{\{A,B,C\}}$	total achievable data rate of the group of logical beams $\{A, B, C\}$ that share the same resource, page 114
$r_{A/(B,C)}$	achievable rate of logical beam A when sharing resource with logical beam B and C, page 114
$r_{k,n}$	rate on chunk n of user k, page 38
$\mathbf{s}$	time-domain transmit signal in MC system, page 26
σ_τ	delay spread, page 22
σ_z^2	noise power, page 38
$\sigma_{k,i}$	the i-th singular value of channel matrix $\mathbf{H}_k$, page 33
S_k	useful signal power of user k , page 113
τ_n	delay of the n-th path, page 19
θ_v	direction of the relative motion of UT, page 19
$\tilde{T}_{\text{chunk}}$	chunk duration normalized by channel coherence time, page 71
T	symbol duration of the baseband signal, page 25
$T_k'(i)$	average data rate of user k being updated after iteration i, page 57

T'_{up}	relative interval of CSI update, page 76
T_{chunk}	chunk duration, page 71
T_{coh}	coherence time, page 23
t_c	size of the historical time window for average rate computation, page 53
T_f	frame duration, page 76
T_g	duration of guard interval, page 27
$T_k(t)$	average data rate of user k till frame t, page 53
T_s	useful OFDM symbol duration, page 25
T'_s	overall OFDM symbol duration including guard interval, page 28
$T_{\mathrm{up,opt}}$	optimum interval of CSI update, page 79
T_{up}	interval of CSI update, page 75
$\mathcal{U}_n$	the set of users served on chunk n, page 54
$\mathbf{v}_{k,i}$	the i-th eigenvector of channel correlation matrix $\mathbf{R}_k$, page 84
$\mathbf{v}_{k,i}$	the i-th right singular vector of channel matrix $\mathbf{H}_k$, page 33
$\mathrm{var}_m^{(k)}$	local variance of the channel coefficient of user k on chunk m, page 69
$\mathbf{W}$	transmit beamforming matrix of all users , page 31
$\mathbf{W}_k$	transmit filter or transmit beamforming matrix of user k, page 30
$\mathbf{w}_{k,n}$	transmit beamforming vector of user k on chunk n, page 81
W_0	initial back-off window in units of transmission opportunities, page 91
W_i	back-off window after i collisions, page 91
$\mathbf{x}$	frequency-domain transmit signal in MC system, page 25
X_σ	shadowing , page 18
$\mathbf{y}$	vector of receive signals of all users, page 31
$\mathbf{y}_k$	vector of receive signal of user k, page 30
y	frequency-domain receive signal in MC systems, page 27

z   frequency-domain additive white Gaussian noise, page 27

$\mathbf{z}_k$ vector of noise of user k, page 30

ž   time-domain additive white Gaussian noise, page 27

Bibliography

[ALS⁺03] Wang Anchun, Xiao Liang, Zhou Shidong, Xu Xibin, and Yao Yan. Dynamic resource management in the fourth generation wireless systems. In *Proc. IEEE International Conference on Communication Technology (ICCT'03)*, volume 2, pages 1095–1098, April 2003.

[AP72] Milton Abramowitz and Irene A. Stegun Powers, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1972.

[Asz95] D. Asztely. On Antenna Arrays in Mobile Communication Systems, Fast Fading and GSM Base Station Receiver Algorithms. Master's thesis, Royal Institute of Technology, February 1995.

[BaRT99] R. Berezdivin and R. Breinig adn R. Topp. Next-generation wireless communications concepts and technologies . In *Proc. Fifth International Symposium on Signal Processing and its Applications (ISSPA'99)*, Brisbane, Australia, August 1999.

[BaRT02] R. Berezdivin and R. Breinig adn R. Topp. Next-generation wireless communications concepts and technologies . *IEEE Communications Magazine*, 40(3), March 2002.

[BBG⁺00] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi. Cdma/hdr: a bandwidth efficient high speed wireless data service for nomadic users. *IEEE Communications Magazine*, 38(7), July 2000.

[BC02] N. Benvenuto and G. Cherubini. *Algorithm for Communication Systems and their Applications*. John Wiley & Sons, October 2002.

[BDL⁺02] L. Becchetti, S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal, and A. Vitaletti. Parallel scheduling problems in next generation wireless networks. In *Proc. the fourteenth annual ACM symposium on Parallel algorithms and architectures (SPAA'02)*, pages 238–247, New York, NY, USA, 2002. ACM Press.

[Bia00]   G. Bianchi. Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–548, March 2000.

[BKA05]   Karsten Brninghaus, Stephan Karger, and David Astely. Link performance models for system level simulations of broadband radio access systems. In *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'05)*, 2005.

[Bla85]   R. E. Blahut. *Fast Algorithms for Digital Signal Processing*. Addison-Wesley, 1985.

[BO99]   M. Bengtsson and B. Ottersten. Optimal downlink beamforming using semi-definite optimisation. In *Proc. 37th Annual Allerton Conference on Communication,Control, and Computing*, September 1999.

[Bon04]   Thomas Bonald. A score-based opportunistic scheduler for fading radio channels. In *Proc. European Wireless Conference (EW'04)*, February 2004.

[BPS98]   E. Biglieri, J. Proakis, and S. Shamai. Fading chanenls: Information-theoritic and communications aspects. *IEEE Transactions on Information Theory*, 44:2619–2692, October 1998.

[Car94]   C. Carneheim. Frequency Hopping GSM. In *Proc. 44th IEEE Vehicular Technology Conference (VTC'94)*, Sockholm, 1994.

[Cav72]   J. K. Cavers. Variable-rate transmission for Rayleigh fading channels. *IEEE Transactions on Communications*, 20, February 1972.

[CEO02]   G. Cherubini, E. Eleftheriou, and S. Oelcer. Filtered Multi Tone Modulation for very High-Speed Digital Subscriber Lines. *IEEE Journal on Selected Areas in Communications*, 20(5), June 2002.

[CFRZ07] E. Costa, A. Frediani, S. Redana, and Y. Zhang. Dynamic Spatial Resource Sharing in Relay Enhanced Cells. In *Proc. 13th European Wireless Conference*, Paris, France, April 2007.

[CH05]   Jihoon Choi and R. W. Heath. Interpolation based transmit beamforming for MIMO-OFDM with limited feedback. *IEEE Transactions on Signal Processing*, 53(11), 2005.

[CM04]   L. U. Choi and R. D. Murch. A transmit preprocessing technique for multi-user MIMO systems using a decomposition approach. *IEEE Transactions on Wireless Communications*, 3(1), January 2004.

[COE05]   C.Wengerter, J. Ohlhorst, and A. G. E. Elbwart. Fairness and throughput analysis for generalized proportional fair frequency scheduling in ofdma. In *Proc. IEEE Vehicular Technology Conference (VTC 2005-Spring)*, volume 3, pages 1903– 1907, May 2005.

[Cor01]   Thomas H. Cormen. *Introduction to Algorithms*. MIT Press, 2001.

[Cos83]   B. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29:439–414, May 1983.

[CS00]   Justin Chuang and Nelson Sollenberger. Beyond 3g: Wideband wireless data access based on ofdm and dynamic packet assignment. *IEEE Communications Magazine*, July 2000.

[CS03a]   G. Caire and S. Shamai. On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Transactions on Information Theory*, 49(7), November 2003.

[CS03b]   Guiseppe Caire and Shlomo Shamai. On the achievable throughput of a multi-antenna gaussian broadcast channel. *IEEE Transactions on Information Theory*, 49(7), July 2003.

[CTB96]   G. Caire, G. Taricco, and E. Biglieri. Capacity of bit-interleaved channels. *Electronics Letters*, 32(12):1060–1061, June 1996.

[CZT07]   Elena Costa, Ying Zhang, and Alberto Totaro. Joint Adaptive FDMA/SDMA in Chunk-based OFDM Systems. In *Proc. IEEE International Conference on Communications*, Glasgow, Scotland, 2007.

[DJT03]   Love D.J., Heath R.W. Jr., and Strohmer T. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Transactions on Information Theory*, 49(6):2735–2747, October 2003.

[Dol47]   C. L. Dolph. A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level. In *Proceedings of the IRE*, volume 35, pages 489–492, 1947.

[DS05]   Goran Dimic and D. Sidiropoulos. On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm. *IEEE Transactions on Signal Processing*, 53(10), October 2005.

[DTSA05]   Le Hai Doan, SeeHo Ting, K. Sakaguchi, and K. Araki. Hierarchical sub-group power and modulation coding adaptation - a new frequency-space link adaptation scheme in MIMO-OFDM eigenmode adaptive transmission system. In *Proc. IEEE 61st Vehicular Technology Conference (VTC 2005-Spring)*, Stockholm, Sweden, May 2005.

[FGH05]   Martin Fuchs, Giovanni Del Galdo, and Martin Haardt. A novel tree-based scheduling algorithm for the downlink of multi-user MIMO systems with ZF beamforming. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, volume III, March 2005.

[FH96]   R. F. H. Fischer and J. B. Huber. A new loading algorithm for discrete multitone transmission. In *Proc. IEEE Global Communications Conference (GLOBECOM'96)*, London, UK, November 1996.

[FN94]   Christof Farsakh and Josef A. Nossek. Application of Space Division Multiple Access to Mobile Radio. In *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'04)*, September 1994.

[FN95]   C. Farsakh and J. A. Nossek. Channel allocation and downlink beamforming in an SDMA mobile radio system. In *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'05)*, pages 687–691, September 1995.

[FN96]   C. Farsakh and J. A. Nossek. A real time downlink channel allocation scheme for an SDMA mobile radio system. In *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'96)*, October 1996.

[Fos96]   G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multiple antennas. *Bell Labs Technical Journal*, 1(2), Autumn 1996.

[FRC$^+$07] A. Frediani, S. Redana, E. Costa, A. Capone, and Y. Zhang. Dynamic Resource Allocation in Relay Enhanced Cells based on WINNER System. In *Proc. IST Summit*, Budapest, Hungary, July 2007.

[Gal68]   R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[GJJV03]   A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE Journal on Selected Areas in Communications*, 21(5):684–702, June 2003.

[GSsS$^+$03] D. Gesbert, M. Shafi, Da shan Shiu, P.J. Smith, and A. Naguib. From theory to practice: an overview of MIMO space-time coded wireless systems. *IEEE Journal on Selected Areas in Communications*, 21(3), April 2003.

[GV96]   Gene H. Golub and Charles F. VanLoan. *Matrix computations*. Johns Hopkins Univ. Press, 1996.

[Hay96]   Simon Haykin. *Adaptive Filter Theory*. Prentice-Hall, third edition, 1996.

[Hoe92]   P. Hoeher. A statistical discrete-time model for the WSSUS multipath channel. *IEEE Transactions on Vehicular Technology*, 41:461–468, November 1992.

[Hot01]   J.M. Hotlzman. Asymptotic analysis of proportional fair algorithm. In *Proc. PIMRC*, 2001.

[HT04]   Harri Holma and Antti Toskala, editors. *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. Wiley, sept. 2004.

[IEE04]   *IEEE Std. 802.16-2004, IEEE standard for local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems.* 2004.

[IST05a]   IST-2003-507581 WINNER. D2.10 Final report on identified RI key technologies, system concept, and their assessment, December 2005.

[IST05b]   IST-2003-507581 WINNER. D2.4 assessment of adaptive transmission technologies, February 2005.

[IST05c]   IST-2003-507581 WINNER. D5.4 final report on link level and system level channel models, September 2005.

[IST06]   IST-4-027756 WINNER II. D6.13.7 v1.00 Test Scenarios and Calibration Cases Issue 2, December 2006.

[ITU03]   ITU Recommendation M.1645: Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000. Available: http://www.itu.int/rec/R-REC-M.1645/en, 2003.

[JCH84]   R. Jain, D. Chiu, and W. Hawe. Quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Technical report, DEC Research Report TR-301, September 1984.

[JG04]   S.A. Jafar and A. Goldsmith. Transmitter Optimization and Optimality of Beamforming for Multiple Antenna Systems. *IEEE Transactions on Wireless Communications*, 3(4):1165–1175, July 2004.

[JKG⁺02]   M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek. Transmit matched filter and transmit wiener filter for the downlink of fdd ds-cdma systems. In *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'02)*, 2002.

[JL03]   Jiho Jang and Kwang Bok Lee. Transmit power adaptation for multiuser ofdm systems. *IEEE Journal on Selected Areas in Communications*, 21(2), February 2003.

[KH05]   Hoon Kim and Youngnam Han. A proportional fair scheduling for multicarrier transmission systems. *IEEE Communications Letters*, 9(3), March 2005.

[KJC+03]   Yungsoo Kim, Byung Jang Jeong, Jaehak Chung, Chan-Soo Hwang, J. S. Ryu, Ki-Ho Kim, and Young Kyun Kim. Beyond 3G: vision, requirements, and enabling technologies. *IEEE Communications Magazine*, 41(3), March 2003.

[KLKL01]   I. Kim, H. L. Lee, B. Kim, and Y. H. Lee. On the use of linear programming for dynamic subchannel and bit allocation in multiuser ofdm. In *Proc. IEEE Global Communications Conference (GLOBECOM'96)*, November 2001.

[KLL03]   D. Kivanc, G. Li, and H. Liu. Computationally e¡Àcient bandwidth allocation and power control for OFDMA. *IEEE Transactions on Wireless Communications*, 2(6), November 2003.

[KMT98]   F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Jounrnal of the Operation Research Society*, 49, 1998.

[KN96]   I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey. *IEEE Transactions on Wireless Communications*, 3(3):10–31, 1996.

[Knu98]   Donald Knuth. *The Art of Computer Programming*. Addison-Wesley, 1998.

[KPD06]   M. Kaneko, P. Popovski, and J. Dahl. Proportional fairness in multi-carrier system: upper bound and approximation algorithms. *IEEE Communications Letters*, 10(6), June 2006.

[KR00]   Brian S. Krongold and Kannan Ramchandran. Computationally Efficient Optimal Power Allocation Algorithms for Multicarirer Communication Systems. *IEEE Transactions on Communications*, 48(1), 2000.

[KRT03]   Iordanis Koutsopoulos, Tianmin Ren, and Leandros Tassiulas. The impact of space division multiplexing on resource allocation: A unified approach. In *IEEE INFOCOM*, 2003.

[KST04]   K. Kar, S. Sarkar, and L. Tassiulas. Achieving proportional fairness using local information in aloha networks. *IEEE Transactions on Automatic Control*, 49(10), October 2004.

[Lee93] W. C. Lee. *Mobile Communication Fundamentals*. New York: Wiley, 1993.

[LL03] Guoqing Li and Hui Liu. Downlink dynamic resource allocation for multi-cell OFDMA system. In *Proc. IEEE Vehicular Technology Conference (VTC 2003-Fall)*, volume 3, pages 1698 – 1702, October 2003.

[LL04] Guoqing Li and Hui Liu. On the optimality of the ofdma network. In *The 38th Annual Asilomar Conferences on Signals, Systems and Computers*, Asilomar, CA, November 2004.

[LN04] Lei Li and Zhisheng Niu. An integrated subchannel scheduling algorithm for adaptive modulation and coding (AMC) MIMO-OFDM wireless systems . In *Proc. Joint Conference of the 10th Asia-Pacific Conference on Communications and 5th International Symposium on Multi-Dimensional Mobile Communications*, Melbourne, Australia, August 2004.

[MA06] P.D. Morris and C.R.N. Athaudage. Fairness Based Resource Allocation for Multi-User MIMO-OFDM Systems . In *Proc. IEEE 63rd Vehicular Technology Conference (VTC 2006-Spring)*, Melbourne, Australia, May 2006.

[Mak] Andrew Makhorin. GNU Linear Programming Kit. Free Software Foundation.

[MI02] Sklavos Alexandros Liu Yin Maniatis Ioannis, Weber Tobias. Pilots for joint channel estimation in multi-user ofdm mobile radio systems. *IEEE 7th Internationa Symposium on Spread Sprectrum Techniques & Applications*, 1:44–48, Sep 2002.

[MK07a] T. F. Maciel and A. Klein. A convex quadratic SDMA grouping algorithm based on spatial correlation. In *Proc. IEEE International Conference on Communications (ICC 2007)*, Glasgow, Scotland, 2007.

[MK07b] T. F. Maciel and A. Klein. A resource allocation strategy for SDMA/OFDMA systems. In *Proc. IST Mobile and Wireless Communications Summit*, Budapest, Hungary, 2007.

[MP92] M. Mouly and M. B. Pautet. *The GSM System for Mobile Communications*. Telecom Publishing, 1992.

[MSEA03] K.K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beamforming with finite rate feedback in multiple-antenna systems. *IEEE Transactions on Information Theory*, 49(10), October 2003.

[NLTW98] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. 16(8), October 1998.

[NP00]   Richard Van Nee and Ramjee Prasad. *OFDM Wireless Multimedia Communication*. Artech House, Bosten,London, 2000.

[OWN96]   Alan V. Oppenheim, Alan S. Willsky, and S. Hamid Nawab. *Signals and Systems*. Prentice Hall, second edition, 1996.

[Pap65]   A. Papoulis. *Probability, random variables and stochastic process*. McGraw-Hill, 1965.

[PHS05]   C. Peel, B. Hochwald, and A. Swindlehurst. A vector-perturbation technique for near-capacity multi-antenna multi-user communication. *IEEE Transactions on Communications*, 53(1):195–202, January 2005.

[PJKL03]   In-Soon Park, Young-Ho Jung, In-Hyoung Kim, and Yong-Hoon Lee. Proportional fair scheduling for wireless communication with multiple transmit and receive antennas. In *Proc. IEEE Vehicular Technology Conference (VTC 2003-Fall)*, volume 3, pages 1573–1577, October 2003.

[PJKL04]   In-Soon Park, Young-Ho Jung, In-Hyoung Kim, and Yong-Hoon Lee. Dynamic subchannel and bit allocation in multiuser MIMO/OFDMA systems. In *Proc. IEEE Vehicular Technology Conference (VTC 2004-Spring)*, volume 2, May 2004.

[Pot95]   G. J. Pottie. System design choices in personal communications. *IEEE Personal Communications Magazine*, 2(5):50–67, October 1995.

[Pro01]   John G. Proakis. *Digital Communication*. Publishing House of Electronics Industry, third edition, 2001.

[PSSH04]   Christian B. Peel, Quentin H. Spencer, A. Lee Swindlehurst, and Martin Haardt. An introduction to the multi-user MIMO downlink. *IEEE Communications Magazine*, pages 60–67, October 2004.

[PWSea04] R. Pabst, B. Walke, D. Schultz, and et al. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Communications Magazine*, pages 80–89, Sep 2004.

[Rap02]   Theodore S. Rappaport. *Wireless Communicaitons:Principles and Practice*. Prentice-Hall PTR, Upper Saddle River, NJ, 2002.

[RC00]   W. Rhee and J. Cioffi. Increase in capacity of multiuser OFDM system using dynamic subchannel allocation. In *Proc. IEEE Vehicular Technology Conference (VTC 2000-Spring)*, pages 1085–1089, Barcelona, Spain, 2000.

158

[RFLT98]   F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8), October 1998.

[RPS⁺06]   N. Riato, G. Primolevo, U. Spagnolini, T. Baudone, and L.Moretti L. Coletti. A cross-layer architecture for SDMA. *IST Mobile Summit*, June 2006.

[SAE05]   Z. Shen, J. G. Andrews, and B. L. Evans. Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints. *IEEE Transactions on Wireless Communications*, 4(6), November 2005.

[Sal67]   B. Saltzberg. Performance of an Efficient Parallel Data Transmission System. *IEEE Transactions on Communications*, 15, December 1967.

[SB04a]   Martin Schubert and Holger Boche. Solution of the Multi-User Downlink Beamforming Problem with Individual SINR Constraints. *IEEE Trans. on Vehicular Techn.*, 53(1):18–28, January 2004.

[SB04b]   Martin Schubert and Holger Boche. Solution of the Multi-User Downlink Beamforming Problem with Individual SINR Constraints. *IEEE Trans. on Vehicular Techn.*, 53(1):18–28, January 2004.

[SBC07]   Stephan Stiglmayr, Martin Bossert, and Elena Costa. Adaptive coding and modulation in ofdm systems using bicm and rate-compatible punctured codes. In *Proc. IEEE European Wireless (EW'07)*, 2007.

[SBO06]   R. Stridh, M. Bengtsson, and B. Ottersten. System evaluation of optimal downlink beamforming with congestion control in wireless communication. *IEEE Transactions on Wireless Communications*, 5(4), April 2006.

[SCA⁺06]   Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans. Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization. *IEEE Transactions on Signal Processing*, 54(9), September 2006.

[SH05]   M. Sharif and B Hassibi. On the capacity of MIMO broadcast channels with partial side information. *IEEE Transactions on Information Theory*, 51(2), February 2005.

[SH07]   Masoud Sharif and Babak Hassibi. A Comparison of Time-Sharing, DPC, and Beamforming for MIMO Broadcast Channels With Many Users. *IEEE Transactions on Communications*, 55(1), January 2007.

[SS04]   Q. Spencer and A. L. Swindlehurst. Channel allocation in multi-user MIMO wireless communications systems. In *Proc. IEEE International Conference on Communications (ICC 2004)*, 2004.

[SSH04]   Quentin H. Spencer, A. Lee Swindlehurst, and Martin Haardt. Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels . *IEEE Transactions on Signal Processing*, 52(2), February 2004.

[Sta93]   TIA/EIA Interim Standard-95. *Mobile station-base station compatibiity standard for dual-mode wideband spread spectrum cellular system.* Telecommunications Industry Association, 1993.

[Ste92]   Raymond Steele. *Mobile Radio Communications*. Pentech Pess, 1992.

[STT+02]   H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj. A fourth-generation MIMO-OFDM broadband wireless system: Design, performance, and field trial results. *IEEE Communications Magazine*, 40(9):143–149, September 2002.

[SV98]   D. Stiliadis and A. Varma. Effcient fair queueing algorithms for packet-switched networks. *IEEE/ACM Transactions on Networking*, 6(2), April 1998.

[SW+07]   P. Svedman, S.K. Wilson, , L. J. Cimini, and B. Ottersten. Opportunistic beamforming and scheduling for ofdma systems. *IEEE Transactions on Communications*, 55(5), May 2007.

[Tan96]   Andrew S. Tanenbaum. *Computer Networks*. Prentice-Hall, 1996.

[TCS04]   Peter Trifonov, Elena Costa, and Egon Schulz. Adaptive user allocation, bit and power loading in multi-carrier systems. In *Proc. 9th International OFDM-Workshop*, Dresden, Germany, September 2004.

[Tel99]   I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6), Nov./Dec. 1999.

[TH05]   Taiwen Tang and R. W. Heath. Opportunistic feedback for downlink multiuser diversity. *IEEE Communications Letters*, 9(10), October 2005.

[Vai92]   P.P. Vaidjanathan. *Multirate Systems and Filter Banks*. NJ: Prentice-Hall, 1992.

[VJG03]   S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Transactions on Information Theory*, 49(10), October 2003.

[VTL02]   Pramod Viswanath, David N. C. Tse, and Rajiv Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6), June 2002.

[VTZZ06]   R. Veronesi, V. Tralli, J. Zander, and M. Zorzi. Distributed dynamic resource allocation for multicell SDMA packet access networks. *IEEE Transactions on Wireless Communications*, 5(10):2772–2783, October 2006.

[Vuc91]   B. Vucentic. An adaptive coding scheme for time-varing channels. *IEEE Transactions on Communications*, 39, May 1991.

[VZLT05]   A. Vineal, Y. Zhang, M. Lott, and A. Tiurlikov. Performance Analysis of the Random Access in IEEE 802.16. In *Proc. PIMRC*, Berlin, Germany, September 2005.

[WCLM99]   Cheong Yui Wong, Roger S. Cheng, Khaled Ben Letaief, and Ross D. Murch. Multiuser OFDM with adaptive sub-carrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications*, 17:1747–1758, October 1999.

[WE71]   S. Weinstein and P. Ebert. Data Transmission by Frequency Division Multiplexing Using the Discrete Fourier Transform. *IEEE Transactions on Communications*, 19(5), October 1971.

[Wei]   Eric W. Weisstein. Order statistic. http:// mathworld.wolfram.com/ OrderStatistic.html.

[WG00]   Z. Wang and G.B. Giannakis. Wireless Multicarrier Communications. *IEEE Signal Processing Magazine*, May 200.

[Wil06]   Thorsten Wild. Successive user insertion for long-term adaptive beams with SDMA using short-term CQI. In *Proc. International OFDM Workshop*, Hamburg, Germany, August 2006.

[Win]   Wireless World Initiative New Radio. www.ist-winner.org.

[Wol98]   Laurence A. Wolsey. *Integer Programming*. A Wiley-Interscience Publication, 1998.

[WSS04]   H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the gaussian mimo broadcast channel. In *Proc. IEEE International Symposium Information Theory*, June 2004.

[YG06]   Taesang Yoo and Andrea Goldsmith. On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming. *IEEE Journal on Selected Areas in Communications*, 24(3), March 2006.

[ZC06]   Y. Zhang and E. Costa. An Efficient Usage of Contention Period for Bandwidth Request. In *Proc. 12th European Wireless Conference*, Athens, Greece, April 2006.

[ZCE07]   Y. Zhang, E. Costa, and E.Schulz. Adaptive FDMA/SDMA in OFDM Systems with Reduced Channel Feedback. In *Proc. 12th International OFDM-Workshop*, Hamburg, Germany, August 2007.

[ZCL05]   Ying Zhang, Elena Costa, and Matthias Lott. Signaling overhead in an mc-cdma system. In *European Wireless*, Nicosia, Cyprus, April 2005.

[ZCL06]   Ying Zhang, Elena Costa, and Matthias Lott. Signaling Overhead for ASBA in an MC-CDMA System. *European Transactions on Telecommunications*, 17:1–11, April 2006.

[Zet95]   P. Zetterberg. A comparison of two systems for down link communication with base station antenna arrays. *IEEE Transactions on Vehicular Technology*, 48(5), November 1995.

[Zha02]   Ruifeng Zhang. Scheduling for maximum capacity in SDMA/TDMA Networks. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02)*, Orlando, Florida, USA, May 2002.

[ZKA01]   J. Zander, S-L Kim, and M. Almgren. *Radio Resource Management for Wireless Networks*. Artech House Publishers, 2001.

[ZL05]   Ying Jun Zhang and Khaled Ben Letaief. An efficient resource-allocation scheme for spatial multi-user access in MIMO/OFDM systems. *IEEE Transactions on Communications*, 53(1), January 2005.

Lebenslauf

Name:	Ying Zhang
Anschrift:	Ottobrunner Str. 15a, 81737 München
Gebursdatum:	09. Februar 1980
Gebursort:	Suzhou, V. R. China
Fammilienstand:	verheiratet

Schulausbildung

1986-1992	Grundschule in Suzhou, V. R. China
1992-1998	Gymnasium in Suzhou, V. R. China

Studium

1998-2002	Studium der Elektrotechnik und Informationstechnik an der Südosten Universität, Nanjing, V. R. China Studienabschluß: Bachelor of Engineering
2002-2004	Master of Science in Communications Engineering Programm an der Technische Universität München, Deutschland Studienabschluß: Master of Science

Berufstätigkeit

seit 2004	Forschungstätigkeit als Mitarbeiter bei Nokia Siemens Networks GmbH & Co. KG, München, Deutschland

Curriculum Vitae

Name:	Ying Zhang
Address:	Ottobrunner Str. 15a, 81737 München
Date of Birth:	09. Februar 1980
Place of Birth:	Suzhou, P. R. China
Family status:	Married

School

Sept 1986 - July 1992	Elementary School of Jinchang District, Suzhou, P. R. China
Sept 1992 - July 1995	The 21st Middle School, Suzhou, P. R. China
Sept 1992 - July 1998	Suzhou High School, Suzhou, P. R. China

Study

Sept 1998 - July 2002	Department of Electronic Information Engineering at Southeast University, Nanjing, P. R. China Graduation: Bachelor of Engineering
Oct 2002 - Oct 2004	Master of Science in Communications Engineering Program at Technical University Munich, Munich, Germany Graduation: Master of Science

Work experience

since Dec 2004	Freelance coworker at Nokia Siemens Networks GmbH & Co. KG, Munich, Germany