# A Novel Algorithm for Distributed Dynamic Interference Coordination in Cellular OFDMA Networks

Von der Fakultät für Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde
eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

vorgelegt von

Marc C. Necker

geb. in Böblingen

| | |
|---|---|
| Hauptberichter: | Prof. Dr.-Ing. Dr. h. c. mult. Paul J. Kühn |
| 1. Mitberichter: | Prof. Dr.-Ing. Phuoc Tran-Gia, Universität Würzburg |
| 2. Mitberichter: | Prof. Dr.-Ing. Andreas Kirstädter |
| Tag der Einreichung: | 26. September 2008 |
| Tag der mündlichen Prüfung: | 18. Februar 2009 |

Institut für Kommunikationsnetze und Rechnersysteme
der Universität Stuttgart

2009

# Abstract

Several currently emerging standards for broadband cellular communication are based on Orthogonal Frequency Division Multiple Access (OFDMA). Examples are the IEEE 802.16e system, a subset of which is promoted by the WiMAX Forum as *Mobile WiMAX*, and the future 3GPP Long Term Evolution (LTE). In OFDMA, different data streams are multiplexed in time and frequency based on the underlying Orthogonal Frequency Division Multiplexing (OFDM) transmission system. This basically corresponds to a combination of Frequency and Time Division Multiple Access (FDMA and TDMA). Resource blocks that form the resource assignment granularity for data streams are defined in the time/frequency plane. OFDM is a form of spread-spectrum communications and offers many advantages, such as robustness in multipath fading environments, low-cost implementations by means of Discrete Fourier Transform (DFT), or the possibility to exploit multi-user diversity.

A major problem in FDMA/TDMA systems and therefore also in OFDMA systems is the inter-cell interference that occurs when transmissions take place on the same resource blocks in neighboring cells. In the downlink direction, neighboring base stations cause inter-cell interference when transmitting on the same resource blocks, and in the uplink direction the interference comes from the transmissions of the mobile terminals. This interference may lead to severe performance degradation or connection loss especially in the border areas of cells. Classical FDMA/TDMA systems such as the Global System for Mobile communications (GSM) solve this problem by avoiding the reuse of the same frequency bands in neighboring cells. This comes at the drawback of wasting precious frequency resources in every cell, and hence it drastically reduces the utilization of the scarce frequency spectrum. Instead, it is desirable to reduce the interference by other means.

A promising approach to solve the problem of inter-cell interference is inter-cell Interference Coordination (IFCO), where neighboring base stations coordinate their transmissions in order to minimize interference. This can be particularly effective when combined with beamforming antennas, which additionally allow the exploitation of spatial multiplexing and thus the transmission to spatially separated terminals in neighboring cells on the same frequency/time resources. Interference coordination goes well along with the observation of an increasing node cooperation for emerging networks, which gives rise to large improvements of the network performance.

Interference coordination can be done on different time scales. Static schemes are applied during the network planning process. Semi-static schemes react on cell-load or user-load variations and operate on time-scales in the order of seconds or several hundred milli-seconds. Finally,

dynamic schemes operate on the frame level or slightly above. With respect to the degree of distribution, global schemes operate with full system knowledge and may instantly distribute scheduling decisions to any base station. Distributed schemes feature a central coordinator, which communicates with all base stations under realistic signaling constraints. Decentralized schemes operate without a central entity. The base stations directly communicate with each other, for example via a fixed network. Finally, local schemes get along without direct communication among the base stations and operate only on local state information.

In this monograph, a novel distributed and dynamic interference coordination scheme for the downlink of an OFDMA network is proposed and evaluated in detail. The presented architecture comprises a central coordinator which collects state information from all base stations, and which distributes coordination information after processing this information. Additionally, every base station performs a dynamic local coordination among its cell sectors taking into account the information delivered by the central coordinator. While the local coordination can take place on the frame level, the communication intervals with the central coordinator may be in the order of half a second or even seconds. The overall signaling delay with the central coordinator may be in the same order of magnitude.

In order to derive the new distributed algorithm, the notion of an interference graph is first introduced as a powerful tool for interference coordination in Chapter 4. In an interference graph, the mobile terminals appear as the vertices, and the edges represent critical interference relations among the mobile terminals. Such a graph may be constructed based on different criteria, where a creation based on the worst case Signal-to-Interference-Ratio (SIR) for every mobile terminal is developed and used here. An analysis of the fundamental properties of the interference graph is conducted, and it is shown how interference coordination can be achieved by a vertex coloring of the interference graph. The colors thereby correspond to disjoint sets of resources in an OFDMA MAC frame. Different heuristics for the NP-complete graph coloring problem are applied and their performance is compared. Furthermore, it will be shown how scheduling constraints such as QoS requirements can be incorporated in this process. A resource assignment strategy based on graph coloring constitutes a global and fully dynamic interference coordination algorithm. It serves as a reference system and forms the basis of the proposed distributed algorithm.

In the proposed distributed dynamic interference coordination algorithm, the central coordinator builds a global interference graph based on state information collected from all base stations. A variation of a weighted graph coloring problem is then solved by the central coordinator, attempting to assign a number of different colors to every mobile terminal in the graph. An efficient heuristic for this coloring process is proposed that takes into account the specific properties of the interference graph. The coloring information of the mobile terminals is then distributed to the corresponding base stations in intervals in the order of half a second or even several seconds.

In every base station, a local interference graph is constructed that comprises the interference relations among all terminals served by the base station in all its cell sectors. Resources are assigned to the mobile terminals based on this local interference graph and taking into account the coloring constraints delivered by the central coordinator. Furthermore, QoS constraints can be considered. This resource assignment problem is formulated as an optimization problem, namely as a Binary Integer Linear Program (BILP). A genetic algorithm is devised in order to efficiently solve this problem. For this, a suitable genetic representation is developed, to-

gether with suitable genetic operators. The presentation of the algorithm is rounded off with a discussion of architectural and signaling aspects.

The performance of the proposed interference coordination algorithm is evaluated by means of event-driven simulations, which are complemented by Monte-Carlo simulations for the parametrization of the algorithm. For this purpose, a comprehensive model of an OFDMA-based IEEE 802.16e network was developed. The model comprises 19 base station sites with three-sectorized cells and a linear phased array beamforming antenna serving each cell sector. A wrap-around technique was applied to avoid border effects in the scenario. Path loss and shadowing were considered using empirical models with full interference among the cells. The obtained results can therefore be applied to the Partial Use of Subchannels (PUSC) zone when all base stations operate with the same permutation base or to the band Adaptive Modulation and Coding (band AMC) zone if fast fading is neglected. For all event-driven simulations, full terminal mobility with $v = 30$ km/h was assumed. The data path of all connections was modeled in detail ranging from an abstract physical layer model based on an average value interface up to the data sources on the IP layer. For the latter, both greedy as well as on-off traffic sources were considered. A detailed description of this model is given in Chapter 6 of this monograph. Furthermore, the basis for this is laid in Chapter 3 with an in-depth discussion of fundamental modeling concepts in cellular networks.

The results of the proposed distributed IFCO approach are compared to uncoordinated reference systems with a frequency reuse of 1 and 3, i.e., to networks where the full available bandwidth and only one third of the available bandwidth is used in every cell sector, respectively. Furthermore, the results are compared to networks with global and local interference coordination. A Random scheduler is applied in order to obtain a fair comparison of the different systems. The performance evaluation is based on direct and indirect metrics. Indirect metrics comprise the resource utilization and the Signal-to-Interference-and-Noise-Ratio (SINR). Direct metrics include the aggregate cell sector throughput, which relates to the overall spectral efficiency of the system. Furthermore, fairness is taken into account by measuring the 5% throughput quantile of all mobile terminals, which is a good indicator for the throughput achieved at the cell edge. Besides average values of these metrics, SINR and throughput results are also considered depending on the geographic position of the mobile terminal.

The performance evaluation in Chapter 7 first compares the uncoordinated frequency reuse 1 and 3 reference systems for the case of greedy traffic sources. The reuse 1 system achieves a higher aggregate throughput, while the reuse 3 system shows the better cell edge performance. Compared to the uncoordinated reuse 3 system, a network with a global interference coordination based on the proposed interference graph achieves a significant gain of more than 70% in the aggregate sector throughput. At the same time, the cell edge throughput is increased by more than 30%. The parametrization of the global IFCO algorithm thereby allows to trade off the aggregate sector throughput and the cell edge throughput. It will further be shown that even a local coordination in every base station increases the aggregate sector throughput by about 40% while sacrificing only about 25% in cell edge performance. Nevertheless, the cell edge throughput is still larger than in an uncoordinated frequency reuse 1 system. Different possibilities to enhance the cell edge performance by means of Fractional Frequency Reuse (FFR) are also investigated. It will be seen that FFR effectively allows to trade off the aggregate sector throughput and the cell edge throughput against each other.

The proposed distributed dynamic interference coordination algorithm achieves a significant performance gain compared to the uncoordinated reference systems. Again, its parametrization allows to trade off the aggregate throughput and the cell edge throughput. Depending on the parameter choice, the aggregate throughput can be increased by about 30% while increasing the cell edge throughput by more than 130% compared to the uncoordinated frequency reuse 3 system. Alternatively, for a different parameter choice, the aggregate throughput is increased by about 50% with a cell edge throughput gain of about 100%. It will further be shown that the signaling delay for the communication with the central coordinator and the update period of the information provided from there may be in the order of half a second while maintaining about the same performance gain.

It will be demonstrated that for the case of uneven load distributions among the cell sectors the proposed algorithm can effectively balance the load between the cells by dynamically re-assigning radio resources from one cell sector to other cell sectors. The performance gain with respect to the aggregate and the cell edge throughput is thereby in the same order of magnitude as with a uniform load distribution. Finally, it will be shown that significant performance gains can be maintained in a scenario with on-off traffic sources without any major modifications to the proposed algorithm.

In conclusion, it can be said that the proposed distributed and dynamic interference coordination algorithm efficiently solves the problem of inter-cell interference in OFDMA networks and significantly increases their aggregate and cell edge throughput. At the same time, the algorithm features a medium complexity, which is an important prerequisite for the implementation in real networks.

# Kurzfassung

Mehrere aufkommende Standards für mobile zelluläre Breitbandkommunikationssysteme basieren auf Orthogonal Frequency Division Multiple Access (OFDMA). Beispiele sind das IEEE 802.16e-System, von welchem eine Untermenge vom WiMAX-Forum als *Mobile WiMAX* vorangetrieben wird, und die sich derzeit in der Standardisierung befindliche Long Term Evolution (LTE) der 3GPP. Basierend auf dem zugrunde liegenden Orthogonal Frequency Division Multiplexing-Verfahren (OFDM) werden in einem OFDMA-System die verschiedenen Datenströme in Zeit und Frequenz gemultiplext. Dies entspricht auf einer bestimmten Abstraktionsebene einer Kombination aus Frequenzmultiplex und Zeitmultiplex (FDM und TDM). Hierbei werden in der Zeit/Frequenzebene Ressourcenblöcke definiert, welche die Zuweisungsgranularität für einen Datenstrom darstellen. OFDM ist ein Frequenzspreizungsverfahren und bietet viele Vorteile, wie beispielsweise robustes Verhalten bei Mehrwegeausbreitung der Radiosignale, preisgünstige Hardware basierend auf Diskreter Fouriertransformation oder die Möglichkeit, Mehrbenutzerdiversität auszunutzen.

Ein großes Problem in Zeit- und Frequenzmultiplexsystemen und damit auch in OFDMA-Systemen stellt die interzelluläre Interferenz dar, welche entsteht, wenn Datenübertragungen in benachbarten Zellen auf derselben Ressource erfolgen. In der Vorwärtsrichtung erzeugen benachbarte Basisstationen diese Interferenz. In der Rückwärtsrichtung kommt die Interferenz entsprechend von den mobilen Endgeräten. Diese Interferenz führt häufig zu einer Beeinträchtigung der Systemleistung bis hin zum Verbindungsabbruch, insbesondere an den Zellrändern. Klassische Zeit- und Frequenzmultiplexsysteme wie das Global System for Mobile communications (GSM) lösen dieses Problem, indem sie die Verwendung der gleichen Frequenzbänder in benachbarten Zellen vermeiden. Dies geht jedoch einher mit einer deutlich verringerten Ausnutzung des knappen Frequenzspektrums. Es ist daher wünschenswert, Interferenz mit anderen Mitteln zu reduzieren.

Ein vielversprechender Ansatz zur Reduzierung interzellulärer Interferenz ist interzelluläre Interferenzkoordination (Interference Coordination, IFCO). Bei dieser Art der Interferenzreduzierung koordinieren benachbarte Basisstation ihre Übertragungen mit dem Ziel, die aufeinander erzeugte Interferenz zu minimieren. Dies kann besonders effizient mit Strahlformung bei Mehrantennensystemen kombiniert werden, welche zusätzlich ein räumliches Multiplexing und somit ein gleichzeitiges Bedienen von räumlich separierten mobilen Endgeräten in benachbarten Zellen auf der gleichen Ressource ermöglicht. Das Konzept der Interferenzkoordination fügt sich nahtlos in den in den letzten Jahren beobachtbaren Trend von zunehmender Kooperation zwischen Netzknoten ein, welche ein Schlüssel zur Erreichung hoher Netzkapazitäten ist.

Interferenzkoordination kann auf unterschiedlichen Zeitskalen betrieben werden. Statische Verfahren werden während der Netzplanung angewendet. Semi-statische Verfahren reagieren auf Variationen bei der aggregierten Zelllast oder der Last einzelner Nutzer. Sie operieren daher auf einer Zeitskala von Sekunden oder mehreren hundert Millisekunden. Schließlich operieren dynamische Verfahren auf der Ebene der Rahmendauer oder wenig darüber. Im Hinblick auf den Verteilungsgrad operieren globale Verfahren mit vollem Systemwissen und der Möglichkeit, Schedulingentscheidungen sofort an alle Basisstationen weiterzugeben. Verteilte Verfahren arbeiten mit einem zentralen Koordinator, welcher über ein Signalisiernetz mit den Basisstationen kommuniziert. Dezentrale Verfahren arbeiten ohne eine zentrale Komponente, wobei die Basisstationen beispielsweise über ein kabelgebundenes Netz direkt miteinander kommunizieren. Letztendlich arbeiten lokale Verfahren ausschließlich auf lokal in einer Basisstation verfügbarer Zustandsinformation und benötigen daher keine Kommunikation mit anderen Netzkomponenten.

In dieser Arbeit wird ein neues verteiltes und dynamisches Verfahren zur Interferenzkoordination in der Vorwärtsrichtung eines OFDMA-Netzes vorgeschlagen und detailliert bewertet. Die vorgestellte Architektur umfasst einen zentralen Koordinator, welcher Zustandsinformationen von allen Basisstationen sammelt, und welcher nach einer Bearbeitung dieser Zustandsinformationen Steuerinformationen zur Koordinierung an alle Basisstationen verteilt. Zusätzlich werden die Sektoren einer jeden Basisstation durch ein lokales dynamisches Verfahren koordiniert, welches die Informationen des zentralen Koordinators berücksichtigt. Während die lokale Koordination auf der Ebene der Rahmendauer stattfindet, kann die Kommunikationsperiode mit dem zentralen Koordinator in der Größenordnung von einer halben Sekunde bis hin zu mehreren Sekunden liegen. Die gesamte Verzögerung zwischen Basisstation und zentralem Koordinator kann dabei in derselben Größenordnung liegen.

Um das neue Verfahren abzuleiten, wird in Kapitel 4 zunächst das Konzept eines Interferenzgraphen eingeführt, welcher ein mächtiges Werkzeug zur Realisierung von Interferenzkoordination darstellt. In einem Interferenzgraphen stellen die mobilen Endgeräte die Knoten dar. Die Kanten repräsentieren kritische Interferenzbeziehungen zwischen den mobilen Endgeräten. Ein solcher Graph kann basierend auf unterschiedlichen Kriterien aufgebaut werden, wobei in dieser Arbeit ein Ansatz basierend auf einer “worst case” Analyse des Signal-zu-Interferenz-Verhältnisses (Signal-to-Interference-Ratio, SIR) für jedes mobile Endgerät entwickelt und verwendet wird. Zunächst werden die grundlegenden Eigenschaften des Interferenzgraphen untersucht, und es wird gezeigt, dass eine Ressourcenzuweisung mit Hilfe einer Knotenfärbung des Interferenzgraphen zur Interferenzkoordination herangezogen werden kann. Die Farben entsprechen hierbei disjunkten Kombinationen von Ressourcenblöcken in einem OFDMA-Rahmen. Zur Lösung des NP-vollständigen Färbeproblems werden verschiedene Heuristiken untersucht und hinsichtlich ihrer Leistung bewertet. Darüberhinaus wird eine Färbeheuristik entwickelt, welche die zusätzliche Berücksichtigung von QoS-Anforderungen und Schedulingprioritäten ermöglicht. Eine solche Vorgehensweise entspricht einem globalen dynamischen Interferenzkoordinierungsverfahren. Es dient als Referenzsystem und bildet die Basis für den im Folgenden entwickelten verteilten Algorithmus.

Im vorgeschlagenen verteilten und dynamischen Interferenzkoordinierungsverfahren baut der zentrale Koordinator einen globalen Interferenzgraphen basierend auf den Zustandsinformationen der Basisstationen auf. Anschließend löst der zentrale Koordinator eine Variation eines

gewichteten Graphenfärbeproblems, wobei versucht wird, einem mobilen Endgerät eine möglichst große Anzahl verschiedener Farben zuzuweisen. Hierfür wird eine effiziente Heuristik vorgeschlagen, welche die Eigenschaften des Interferenzgraphen während des Färbevorgangs berücksichtigt. Diese Färbeinformation wird den Basisstationen dann in Intervallen von einer halben Sekunde oder sogar einigen weniger Sekunden signalisiert.

In jeder Basisstation wird zusätzlich ein lokaler Interferenzgraph erzeugt, welcher alle mobilen Endgeräte in den Zellsektoren der Basisstation umfasst. Eine Basisstation weist den einzelnen Übertragungen zu den mobilen Endgeräten dann basierend auf diesem lokalen Interferenzgraphen Ressourcen zu, wobei die durch die Farben vorgegebenen Randbedingungen des zentralen Koordinators berücksichtigt werden. Des Weiteren können QoS-Anforderungen mit einbezogen werden. Dieses Ressourcenzuweisungsproblem wird als Optimierungsproblem formuliert, genauer als ein Binäres lineares Programm (Binary Integer Linear Program, BILP). Zur Lösung dieses Problems wird ein genetischer Algorithmus abgeleitet. Dies beinhaltet die Entwicklung einer geeigneten genetischen Repräsentation sowie der entsprechenden genetischen Operatoren. Die Präsentation des verteilten Algorithmus wird abgerundet durch eine Diskussion von architekturellen Aspekten und von Signalisieranforderungen.

Die Leistung des vorgeschlagenen Verfahrens zur Interferenzkoordination wird mit Hilfe ereignisgesteuerter Simulationen bewertet, welche durch Monte-Carlo-Simulationen zur Parametrisierung des Algorithmus ergänzt werden. Zu diesem Zweck wurde ein umfassendes Modell eines mehrzelligen OFDMA-basierten IEEE 802.16e-Netzes entwickelt. Das Modell umfasst 19 Basisstationen mit jeweils drei Zellsektoren und einem linear angeordneten Mehrantennensystem pro Zellsektor. Zur Vermeidung von Randeffekten wurde das Szenario mit “wrap-around” simuliert. Pfadverlust und Abschattungen wurden mit Hilfe empirischer Modelle berücksichtigt, wobei alle Zellen volle Interferenz aufeinander ausüben. Die Ergebnisse sind daher gültig für die Permutationszone Partial Use of Subchannels (PUSC) für den Fall, dass alle Basisstation mit der selben Permutationsbasis arbeiten, oder für die Permutationszone band Adaptive Modulation and Coding (band AMC) für den Fall, dass schneller Signalschwund vernachlässigt wird. Für alle ereignisgesteuerten Simulationen wurde eine volle Mobilität der Teilnehmer mit einer Geschwindigkeit von $v = 30$ km/h angenommen. Der Datenpfad aller Verbindungen wurde von einem Abstraktionsmodell der Bitübertragungsschicht bis hin zu den Verkehrsgeneratoren auf der IP-Ebene detailliert modelliert. Für letztere wurden sowohl Verkehrsquellen angenommen, welche die volle verfügbare Bandbreite ausnutzen können, als auch Quellen basierend auf einem An-Aus-Modell. Eine Beschreibung dieses Modells findet sich in Kapitel 6 dieser Arbeit. Die Basis hierfür wird in Kapitel 3 mit einer ausführlichen Diskussion fundamentaler Modellierungskonzepte in zellulären Netzen gelegt.

Die Leistungsdaten des vorgeschlagenen verteilten und dynamischen Interferenzkoordinierungsverfahren werden verglichen mit unkoordinierten Referenzsystemen mit Frequenzwiederholfaktoren 1 und 3, d.h. mit Systemen, in denen die gesamte verfügbare Bandbreite bzw. ein drittel davon in jedem Zellsektor verwendet wird. Darüberhinaus werden die Leistungsdaten mit global und lokal interferenzkoordinierten Netzen verglichen. Um einen fairen Vergleich der Systeme zu gewährleisten, wurde ein Zufallsscheduler zur Priorisierung der Datenströme eingesetzt. Die Leistungsbewertung basiert auf direkten und indirekten Metriken. Indirekte Metriken umfassen die Ressourcenausnutzung und das Signal-zu-Interferenz-Rausch-Verhältnis (Signal-to-Interference-Noise-Ratio, SINR). Direkte Metriken schließen den aggregierten Sek-

tordurchsatz mit ein, welcher direkt mit der durchschnittlichen spektralen Effizienz korreliert. Des Weiteren wird Fairness durch die Messung des 5% Durchsatzquantils aller mobilen Endgeräte berücksichtigt. Dieses ist ein guter Indikator für den am Zellrand erzielten Durchsatz. Neben den Durchschnittswerten dieser Metriken werden das SINR und der Durchsatz auch in Abhängigkeit der geographischen Position eines mobilen Endgerätes betrachtet.

Die Leistungsbewertung in Kapitel 7 vergleicht zunächst die unkoordinierten Referenzsysteme mit Frequenzwiederholfaktoren 1 und 3 für den Fall einer Verkehrsquelle, welche die gesamte zur Verfügung stehende Bandbreite ausnutzen kann. Das System mit Wiederholfaktor 1 erzielt einen besseren Gesamtdurchsatz, wohingegen das System mit Wiederholfaktor 3 eine bessere Leistung am Zellrand aufweist. Im Vergleich zu einem System mit Wiederholfaktor 3 kann ein global koordiniertes System basierend auf einem Interferenzgraphen eine signifikante Leistungssteigerung von mehr als 70% im Gesamtdurchsatz erzielen. Gleichzeitig kann der Zellranddurchsatz um mehr als 30% gesteigert werden. Die Parametrisierung des Algorithmus erlaubt es dabei, den Gesamtdurchsatz und den Zellranddurchsatz gegeneinander abzuwägen. Es wird außerdem gezeigt, dass eine lokale Koordination in jeder Basisstation den Gesamtdurchsatz um ca. 40% steigern kann, wobei lediglich 25% der Zellrandleistung eingebüßt werden. Dennoch ist der Zellranddurchsatz immer noch größer als in einem unkoordinierten System mit Frequenzwiederholfaktor 1. Darüberhinaus werden verschiedene Möglichkeiten untersucht, die Zellrandleistung mit Hilfe eines Fractional Frequency Reuse (FFR) zu verbessern. Es wird gezeigt, dass durch den FFR der Gesamtdurchsatz und der Zellranddurchsatz gegeneinander abgewogen werden können.

Das vorgeschlagene verteilte und dynamische Interferenzkoordinierungsverfahren erzielt einen hohen Leistungsgewinn im Vergleich zu den unkoordinierten Referenzsystemen. Auch hier erlaubt die Parametrisierung eine Abwägung von Gesamt- und Zellranddurchsatz. Abhängig von der Parameterwahl kann der Gesamtdurchsatz um etwa 30% gegenüber einem unkoordinierten System mit Wiederholfaktor 3 gesteigert werden während der Zellranddurchsatz sich sogar um mehr als 130% verbessert. Alternativ ergibt sich für eine andere Parameterwahl ein Gewinn beim Gesamtdurchsatz von ca. 50% bei einem immer noch hohen Gewinn beim Zellranddurchsatz von ungefähr 100%. Es wird weiterhin gezeigt, dass die Signalisierverzögerung mit dem zentralen Koordinator und die Kommunikationsperiode in der Größenordnung einer halben Sekunde liegen können wobei fast derselbe hohe Leistungsgewinn gewahrt wird.

Es wird weiterhin gezeigt, dass der Algorithmus auch für den Fall einer ungleichen Lastverteilung auf die Zellen die Last effizient zwischen den Zellen ausgleichen kann. Hierfür werden dynamisch Ressourcen von einem Sektor zum anderen umverteilt. Der Gewinn beim Gesamtdurchsatz und dem Zellranddurchsatz ist dabei in derselben Größenordnung wie bei einer gleichförmigen Lastverteilung. Letztendlich wird gezeigt, dass ohne große Modifikationen am Algorithmus auch im Falle einer An-Aus-Verkehrsquelle ein signifikanter Leistungsgewinn erzielt werden kann.

Zusammenfassend kann gesagt werden, dass das vorgeschlagene verteilte und dynamische Verfahren zur Interferenzkoordination das Problem der interzellulären Interferenz in OFDMA-Netzen effizient löst und deren Gesamt- und Zellranddurchsatz erheblich steigert. Der Algorithmus hat dabei lediglich eine mittlere Komplexität, was eine wesentliche Voraussetzung für die Anwendung in realen Netzen ist.

# Contents

# List of Figures

# List of Tables

# Abbreviations and Symbols

### Abbreviations

| | |
|---|---|
| 3GPP | 3rd Generation Partnership Project |
| AAA | Authentication, Authorization, and Accounting |
| AAS | Advanced Antenna System |
| AMC | Adaptive Modulation and Coding |
| AMPS | Advanced Mobile Phone Service |
| AQM | Active Queue Management |
| ARP | Autonomous Reuse Partitioning |
| ARQ | Automatic Repeat Request |
| ASIC | Application Specific Integrated Circuit |
| ASN | Access Service Network |
| ASN-GW | Access Service Network Gateway |
| ASP | Application Service Provider |
| ATM | Asynchronous Transfer Mode |
| AWGN | Additive White Gaussian Noise |
| BCA | Borrowing Channel Assignment |
| BEC | Backward Error Correction |
| BER | Bit Error Ratio |
| BILP | Binary Integer Linear Program |
| BLER | Block Error Ratio |
| BPSK | Binary Phase Shift Keying |

| | |
|---|---|
| BRM | Basic Reference Model |
| CC | Convolutional Coding |
| ccdf | complementary cumulative distribution function |
| cdf | cumulative distribution function |
| CDMA | Code Division Multiple Access |
| CEPT | Conférence Européenne des administrations des Postes et des Télécommunications (European Conference of Postal and Telecommunications Administrations) |
| CID | Connection Identifier |
| CP | Cyclic Prefix |
| CPS | Common Part Sublayer |
| CPS | Control Plane Server |
| CQI | Channel Quality Indicator |
| CRC | Cyclic Redundancy Check |
| CRN | Common Random Numbers |
| CS | Convergence Sublayer |
| CSD | Circuit Switched Data |
| CSN | Connectivity Service Network |
| CTC | Convolutional Turbo Coding |
| CXZ | Co-existence Zone |
| DCA | Dynamic Channel Assignment |
| DFT | Discrete Fourier Transform |
| DL | Downlink |
| DoA | Direction of Arrival |
| DSL | Digital Subscriber Line |
| ECR | Effective Code Rate |
| EDGE | Enhanced Data rates for GSM Evolution |
| eNB | eNode B |
| EV-DO | Evolution Data Only |

| | |
|---|---|
| FCH | Frame Control Header |
| FDD | Frequency Division Duplex |
| FDMA | Frequency Division Multiple Access |
| FEC | Forward Error Correction |
| FFR | Fractional Frequency Reuse |
| FFT | Fast Fourier Transform |
| FPGA | Field Programmable Gate Array |
| FUSC | Full Usage of Subchannels |
| GA | Genetic Algorithm |
| GPRS | General Packet Radio Service |
| GPS | Global Positioning System |
| GRE | Generic Routing Encapsulation |
| GSM | Global System for Mobile communications |
| GSM | Groupe Spécial Mobile |
| GW | Gateway |
| HARQ | Hybrid Automatic Repeat Request |
| HSDPA | High Speed Downlink Packet Access |
| HSPA | High Speed Packet Access |
| HSUPA | High Speed Uplink Packet Access |
| I-WLAN | Interworking Wireless Local Area Network |
| i.i.d. | independent and identically distributed |
| ICI | Intercarrier Interference |
| ICIC | Inter-Cell Interference Coordination |
| IDFT | Inverse Discrete Fourier Transform |
| IDMA | Interleave Division Multiple Access |
| IEEE | Institute of Electrical and Electronics Engineers |
| IETF | Internet Engineering Task Force |
| IFCO | Interference Coordination |

| | |
|---|---|
| ILP | Integer Linear Program |
| IMT | International Mobile Telecommunications |
| IP | Internet Protocol |
| ISI | Intersymbol Interference |
| ITU | International Telecommunication Union |
| ITU-T | International Telecommunication Union – Telecommunication Standardization Sector |
| LAN | Local Area Network |
| LoS | Line-of-Sight |
| LP | Linear Program |
| LTE | Long Term Evolution |
| MAC | Media Access Control |
| MAI | Multiple Access Interference |
| MBMS | Multimedia Broadcast Multicast Service |
| MCS | Modulation and Coding Scheme |
| MIMO | Multiple Input Multiple Output |
| MIP | Mixed Integer Program |
| MME | Mobility Management Entity |
| MTU | Maximum Transmission Unit |
| MUD | Multi-User Detection |
| NAP | Network Access Provider |
| NGMN | Next Generation Mobile Networks |
| NLoS | Non-Line-of-Sight |
| NMT | Nordic Mobile Telephone |
| NP | Nondeterministic Polynomial-time |
| NSP | Network Service Provider |
| NTT | Nippon Telephone and Telegraph |
| NWG | Network Working Group |
| OFDM | Orthogonal Frequency Division Multiplexing |

| | |
|---|---|
| OFDMA | Orthogonal Frequency Division Multiple Access |
| PAPR | Peak to Average Power Ratio |
| pdf | probability density function |
| PDU | Protocol Data Unit |
| PHS | Payload Header Suppression |
| PMS | Power Matched Scheduling |
| PMX | Partially Matched Crossover |
| PN | Pseudo Noise |
| PSK | Phase Shift Keying |
| PUSC | Partial Usage of Subchannels |
| QAM | Quadrature Amplitude Modulation |
| QoS | Quality of Service |
| QPSK | Quadrature Phase Shift Keying |
| RAN | Radio Access Network |
| RNC | Radio Network Controller |
| RNG | Random Number Generator |
| RRM | Radio Resource Management |
| RTT | Round Trip Time |
| S-GW | Serving Gateway |
| SAE | Service Architecture Evolution |
| SAP | Service Access Point |
| SC | Subcarrier |
| SC | Single Carrier (10 – 66 GHz LoS operation in IEEE 802.16) |
| SCa | Single Carrier (below 11 GHz NLoS operation in IEEE 802.16) |
| SCTP | Stream Control Transmission Protocol |
| SDMA | Space Division Multiple Access |
| SDU | Service Data Unit |
| SINR | Signal-to-Interference-and-Noise Ratio |

| | |
|---|---|
| SIR | Signal-to-Interference Ratio |
| SMS | Short Message Service |
| SNR | Signal-to-Noise Ratio |
| SOFDMA | Scalable Orthogonal Frequency Division Multiple Access |
| SS | Subscriber Station |
| SWO | Squeaky Wheel Optimization |
| TCP | Transmission Control Protocol |
| TDD | Time Division Duplex |
| TDMA | Time Division Multiple Access |
| TG | Task Group |
| TR | Technical Report |
| TS | Technical Specification |
| UL | Uplink |
| UMTS | Universal Mobile Telecommunication System |
| VRT | Variance-Reduction Technique |
| WAN | Wide Area Network |
| WiMAX | Worldwide Interoperability for Microwave Access |
| WLAN | Wireless Local Area Network |

## Symbols

| | |
|---|---|
| $a$ | Averaging factor of exponential average filter for burst profile management |
| $A_F$ | Activity factor for on-off traffic model |
| $B$ | Total available bandwidth |
| $b$ | Number of colors assigned to one vertex with fractional graph coloring |
| $B_{cell}$ | Bandwidth available in a cell |
| $B_{reuse1}$ | Bandwidth available in the reuse 1 band of FFR |
| $B_{reuse3}$ | Bandwidth available in the reuse 3 band of FFR |
| $\bar{C}$ | Average clustering coefficient of whole graph |
| $\boldsymbol{C}$ | Set of colors used for the graph coloring in the outer optimization problem |
| $\boldsymbol{C}_i$ | Set of colors assigned to mobile terminal $m_i$ by the central coordinator |
| $c$ | Speed of light |
| $C_k$ | Clustering coefficient $C_k$ of vertex $m_k$ |
| $\vec{c}_i$ | Georgraphic coordinates of transceiver serving cell sector $i$ |
| $d$ | Distance between mobile terminal and base station antenna |
| $d$ | Distance of antenna elements in a linear phased array |
| $d_{13}$ | Distance ratio for system with distance-based FFR |
| $D_{aggr}$ | Steady-state aggregate throughput |
| $d_{BS}$ | Distance between base stations (inter-site distance) |
| $d_{corr}$ | Decorrelation length |
| $D_{edge}$ | Steady-state 5% throughput quantile |
| $\bar{D}_{edge}$ | 5% throughput quantile including transient effects |
| $deg(m_k)$ | Degree of vertex $m_k$ |
| $d_{ic}$ | Coordination diameter |
| $d_{ic,i}$ | Coordination diameter of the inner interference graph |
| $d_{ic,o}$ | Coordination diameter of the outer interference graph |
| $d_{MT}$ | Distance of mobile terminal to its serving base station |
| $D_S$ | Desired minimum SIR |

| | |
|---|---|
| $D_{S,i}$ | Desired minimum SIR of the inner interference graph |
| $D_{S,o}$ | Desired minimum SIR of the outer interference graph |
| $\boldsymbol{E}$ | Adjacency matrix |
| $e_{jk}$ | Elements of adjacency matrix $\boldsymbol{E}$ |
| $E_s$ | Energy per data symbol |
| $E[\bullet]$ | Expected value of $\bullet$ |
| $F$ | Fitness of a solution in a genetic algorithm |
| $f_C$ | Carrier frequency |
| $\Delta f$ | OFDM subcarrier spacing |
| $f_D$ | Doppler shift |
| $f_{D,max}$ | Maximum doppler shift in the scenario |
| $f_i$ | Frequency of a subcarrier |
| $G$ | Geometry factor |
| $G_{array}$ | Power gain of linear phased array |
| $G_{min}$ | Geometry threshold in Adaptive Fractional Frequency Reuse |
| $G_{MT}$ | Mobile terminal antenna gain |
| $G_S$ | Antenna gain in boresight direction |
| $h_b$ | Height of base station antenna |
| $H_k$ | Sampled channel transfer function on subcarrier $k$ |
| $h_k$ | Taps of the discrete-time channel impulse response $h(t)$ |
| $h_m$ | Height of mobile terminal's antenna |
| $h(t)$ | time-variant channel impulse response |
| $\boldsymbol{I}$ | Interference matrix with elements $I_{kl}$ describing interference to mobile $m_k$ when transmitting to mobile $m_l$ |
| $I$ | Sum power of all interfering signals |
| $I_{kl}$ | Interference to mobile $m_k$ when transmitting to mobile $m_l$ |
| $k$ | Constant of proportionality for distance-dependent path loss |
| $k_{OLF}$ | Overlength factor for random genomes |
| $k_s$ | Number of cell sectors served by a base station |

| | |
|---|---|
| $L$ | Length of genome |
| $\mathrm{ld}\, n$ | Logarithm of $n$ to the basis two |
| $L_{p\ (\mathrm{dB})}$ | Path loss in logarithmic domain |
| $L_{p\ (\mathrm{lin})}$ | Path loss in linear domain |
| $L_{VF}$ | Length of a virtual frame |
| $M$ | Number of required colors for the coloring of a graph |
| $\boldsymbol{M}_b$ | Set of all mobile terminals served by base station $b$ |
| $n_k$ | Mobile terminal with index $i$ |
| $N$ | Noise power |
| $N$ | Number of mobile terminals per cell sector |
| $N_0$ | Noise power per data symbol |
| $N_{ant}$ | Number of antenna elements in a linear phased array |
| $N_{cells}$ | Overall number of cell sectors in the scenario |
| $N_{drops}$ | Number of drops in a Monte-Carlo simulation run |
| $N_F$ | Number of resource partitions a permutation zone is divided into |
| $N_{gen}$ | Number of generations of a genetic algorithm run |
| $\boldsymbol{N}_k$ | Neighborhood of vertex $m_k$ |
| $n_k$ | Sampled noise |
| $N_{max}$ | Maximum number of mobile terminals below the geometry threshold $G_{min}$ in Adaptive Fractional Frequency Reuse |
| $n_o$ | Number of occupied resource partitions in a virtual frame |
| $N_{sc}$ | Number of of subcarriers in an OFDM transmission system |
| $N_{sched}$ | Number of served terminals based on the scheduler's decision |
| $N_{slots,f}$ | Number of slots in frequency direction of a permutation zone |
| $N_{slots,t}$ | Number of slots in time direction of a permutation zone |
| $N_{srvd}$ | Number of terminals served within one permutation zone of a MAC frame |
| $n(t)$ | AWGN noise signal |
| $N_{ts}$ | Tournament size if tournament selection is used |
| $n_u$ | Number of unserved mobile terminals in a virtual frame |

| | |
|---|---|
| $\|\boldsymbol{P}\|$ | Population size of a generation in a genetic algorithm |
| $p$ | Fraction of total available bandwidth assigned to reuse 1 band in FFR |
| $p_{jk}$ | Path loss from transceiver $tr_j$ to mobile $m_k$ |
| $P_{L,target}$ | Target BLER |
| $\boldsymbol{P}_N$ | Population set in generation $N$ |
| $P_{Tr}$ | Transmit power |
| $P_{Tr,low}$ | Transmit power in soft-reuse frequency bands |
| $r$ | Frequency reuse factor (sometimes also $1/r$) |
| $R_C$ | Crossover rate in a genetic algorithm |
| $r_c$ | Code rate of an FEC code |
| $r_{eff}$ | Effective frequency reuse factor |
| $\boldsymbol{R}_i$ | Set of colors assigned to mobile terminal $m_i$ by base station during inner optimization |
| $R_M$ | Mutation rate in a genetic algorithm |
| $R_R$ | Replacement rate in a steady state genetic algorithm |
| $R_{SI}$ | Swap-in probability |
| $\boldsymbol{S}$ | Set of feasible solutions |
| $S$ | Received signal power |
| $SINR_{avg}$ | Average SINR based on an exponential average filter |
| $SINR_{inst}$ | Instantaneous SINR of the last received data burst |
| $S_k$ | Signal power received by mobile $m_k$ |
| $T_B$ | Real-time duration of a batch in an event-driven simulation run |
| $t_{C,delay}$ | Signaling delay with central coordinator |
| $t_{C,innerup}$ | Update period for inner interference graph |
| $t_{C,up}$ | Update period for information from central coordinator |
| $T_G$ | Length of guard interval |
| $th_{low}$ | Lower SINR threshold for system with SINR-based FFR |
| $th_{up}$ | Upper SINR threshold for system with SINR-based FFR |
| $T_{off}$ | Random variable for the duration of the off-period in the on-off traffic model |

| | |
|---|---|
| $T_{on}$ | Random variable for the duration of the on-period in the on-off traffic model |
| $T_P$ | Random variable for the period duration of the on-off traffic model |
| $T_R$ | Real-time duration of an event-driven simulation run |
| $T_S$ | OFDM symbol duration |
| $T_{STP}$ | Short term time period for measurement of throughput quantile |
| $T_U$ | Time duration of usable portion of OFDM symbol |
| $u_i$ | Utility for serving mobile terminal $m_i$ during inner optimization |
| $\boldsymbol{V}_k$ | Set of mobile terminal $m_k$ containing all other mobile terminals that must not be assigned the same resources |
| $v$ | Terminal velocity |
| $v_i$ | Number of resource partitions that need to be assigned to mobile terminal $m_i$ |
| $v_k$ | Phase shift of signal for phased array antenna element $k$ |
| $\boldsymbol{W}_k$ | Set containing all interference relations of mobile terminal $m_k$ |
| $\bar{\boldsymbol{W}}_k$ | Set containing all non-critical interference relations of mobile terminal $m_k$ |
| $w_k$ | Complex weighting factor of phased array antenna element $k$ |
| $w_k$ | Samples of the time continuous transmitted signal $w(t)$ |
| $w(t)$ | Time continuous transmitted signal |
| $w_{taper,k}$ | Tapering weight of phased array antenna element $k$ |
| $x_{ikl}$ | Matrix elements decribing the resource allocation during inner optimization |
| $x_k$ | Transmitted data or pilot symbol on subcarrier $k$ |
| $\tilde{x}_k$ | Transmitted data or pilot symbols after IDFT |
| $\tilde{y}_k$ | Received data or pilot symbols before DFT |
| $y(t)$ | Time continuous received signal |
| | |
| $\alpha$ | Steering angle of linear phased array |
| $\beta$ | Path loss exponent |
| $\chi(G)$ | Chromatic number of graph $G$ |
| $\delta_i$ | Discrete delta function: $\delta_0 = 1$, $\delta_k = 0 \ \forall k \neq 0$ |

| | |
|---|---|
| $\varepsilon$ | Chase combining efficiency |
| $\eta$ | Additional gain achieved by incremental redundancy when HARQ is used |
| $\theta$ | Phase shift of a radio echo |
| $\kappa$ | Scaling factor in fitness function of genetic algorithm for the solution of the inner optimization problem |
| $\lambda$ | Wavelength |
| $\xi$ | Constant offset in fitness function of genetic algorithm for the solution of the inner optimization problem |
| $\rho$ | Resource utilization |
| $\rho_{BS}$ | Correlation coefficient for links of one mobile terminal towards different base stations |
| $\rho_{up,down}$ | Correlation coefficient for links in the uplink and downlink direction of one mobile terminal |
| $\sigma$ | Standard deviation |
| $\tau$ | Delay of a radio echo |
| $\varphi$ | Angle of interest of linear phased array |

# 1 Introduction

In the past decade, mobile communication has penetrated everyday's life. The main use cases have been and still are voice and the low-rate Short Message Service (SMS). However, within only a few years, high rate wireless data services have become affordable. This includes direct Internet connectivity, but also dedicated services such as video or map services. This goes along with a transition from circuit switched systems to packet switched radio networks. Recent developments and standardization efforts are further pushing towards high data rate packet switched networks and high-quality services for end users. As the available radio spectrum is limited, this brings up the need for advanced algorithms and concepts to increase the spectral efficiency of wireless networks.

## 1.1 Overview of Wireless Networks

Wireless networks can be classified into Wireless Local Area Networks (Wireless LANs, WLANs) with limited geographical coverage and cellular networks which cover a large geographical area. For Wireless LANs, the Institute of Electrical and Electronics Engineers (IEEE) is the dominating standardization body, and the IEEE 802.11 standard family the prevailing technology [Ste05]. The most prominent active standard currently is IEEE 802.11g, which provides net data rates of up to 54 MBit/s with a coverage of a few ten meters. Recent developments will push this to up to 248 MBit/s in the upcoming IEEE 802.11n standard with the help of Multiple Input Multiple Output (MIMO) techniques. Wireless LAN networks are designed to cover small areas also known as *hotspots*, such as local networks on private properties, on company premises, or in public places such as airports or cafes. This is further supported by a mobility mechanism on layer 2 that allows to set up a network consisting of multiple access points.

In contrast, cellular networks cover a large geographic area, such as a country. They divide the covered area into cells, each served by a base station transceiver. The size of a cell depends on the underlying technology, its surrounding terrain, and the expected traffic demand. In rural areas with a low traffic density, the diameter of a cell may be in the order of tens of kilometers. In urban areas with a high traffic density, the diameter may be as small as a few hundred meters. Small cells allow for higher data rates within the cells. Moreover, a large number of cells balances the offered traffic over many transceivers allowing for a higher network capacity compared to large cells. The base stations are part of the Radio Access Network (RAN), and they are connected to a wide area backbone network referred to as *core network*. Cellular net-

works usually offer a sophisticated mobility management, making them suitable for covering large areas.

Many technologies co-exist in the area of cellular networks. The 2nd generation Global System for Mobile communications (GSM) is certainly the most successful system to date. In September 2007, there were over 2.5 billion GSM subscribers, which made up 85% of the world's mobile market [GSMA08]. The GSM standards are currently maintained by the 3rd Generation Partnership Project (3GPP). Besides the 3GPP, other standardization bodies have been active in specifying cellular networks, such as the IEEE with the IEEE 802.16e standard. Both the 3GPP and the IEEE are currently working on new standards based on Orthogonal Frequency Division Multiple Access (OFDMA), namely the 3GPP Long Term Evolution (LTE), and the IEEE 802.16m. The latter is the successor of IEEE 802.16e, another system that was already based on OFDMA. In OFDMA, the transmissions are multiplexed in time and frequency based on the underlying Orthogonal Frequency Division Multiplexing (OFDM) system. It can therefore be regarded as a combination of Time and Frequency Division Multiple Access (TDMA/FDMA). OFDM is inherently robust against multipath fading, and it allows for an inexpensive implementation by means of a Discrete Fourier Transform (DFT).

Recently, hotspot and cellular techniques have increased their competition. Several efforts have been made to cover larger areas such as city centers by means of WLAN technology. On the other hand, the concept of femto cells has been developed within the 3GPP, which cover small regions comparable to the coverage area of WLAN access points. The corresponding hardware is also referred to as Home Node B. They are meant as a replacement of WLAN access points at the customer premises and can thus easily increase the capacity of the cellular network.

## 1.2 Interference and Node Cooperation

Small cells in urban areas and Home Node Bs increase the problem of inter-cell interference, which occurs when close-by transceivers use the same radio resources for transmission and/or reception. This interference may lead to severe performance degradation or connection loss especially in the border areas of cells. GSM and WLAN networks solve this problem by avoiding the reuse of the same frequency bands in neighboring cells. This comes at the drawback of wasting precious frequency resources in every cell, and hence it drastically reduces the utilization of the scarce frequency spectrum. Instead, it is desirable to reduce the interference by other means. A promising approach to solve the problem of inter-cell interference is inter-cell Interference Coordination (IFCO), where neighboring base stations or access points coordinate their transmissions in order to minimize interference. This can be particularly effective when combined with beamforming antennas, which additionally allow the exploitation of spatial multiplexing and thus the transmission to spatially separated terminals on the same frequency/time resources.

Interference coordination goes well along with the observation of an increasing node cooperation for emerging networks, which gives rise to large improvements of the network performance. Research in wireless networks focused on increasing the data rate on a single link by means of advanced techniques for a long time. Recently, more research has been targeting the cooperation

of network nodes, where IFCO is one prominent example. Other examples include cooperative MIMO (see Section 2.6.3) and other network MIMO techniques (see Section 2.6.5).

In this monograph, a novel distributed and dynamic interference coordination scheme for the downlink of a cellular OFDMA network is presented. The algorithm is evaluated at the example of an IEEE 802.16e network, but the basic idea is applicable to any kind of OFDMA or FDMA/TDMA network. It can also be applied to other network types, such as wireless LANs or mesh networks, possibly with technology specific adaptations. The proposed architecture comprises a central coordinator that collects state information from all base stations of a large area, and which distributes coordination information after processing this information. Additionally, every base station performs a dynamic local coordination among its cell sectors taking into account the information delivered by the central coordinator. While the local coordination can take place on the frame level, the communication intervals with the central coordinator may be in the order of half a second or even seconds. The overall signaling delay with the central coordinator may be in the same order of magnitude. The performance of the algorithm is thoroughly evaluated by means of fully dynamic event-driven simulations, which are complemented by Monte-Carlo simulations to parametrize the algorithm and study its convergence behavior.

## 1.3 Outline

The remainder of this monograph is divided into three parts. Chapters 2 and 3 deal with the thematic and methodological basics. Chapters 4 and 5 deal with interference coordination in particular. Finally, Chapters 6 and 7 focus on the simulation and performance evaluation of the considered scenarios.

Chapter 2 first introduces the basic concepts of wireless networks. It thereby focuses on cellular networks, albeit many explanations are generally applicable to any kind of wireless network. After a brief overview of the history of cellular networks, the Chapter covers aspects of the physical and Media Access Control (MAC) layer. This includes the wireless channel, the impact of noise and interference, and principles of data transmission in wireless networks on the physical and the MAC layer. The Chapter is concluded by an introduction to the IEEE 802.16e technology, which is often referred to as *Mobile WiMAX*. Appendix A amends this description by an overview of the 3GPP LTE network architecture.

Chapter 3 is the second basic Chapter in this monograph, dealing with the methodological basics. It is divided into three parts. The first part deals with the modeling of wireless networks in general and cellular networks in particular. After a brief introduction into the topic of simulation, empirical models of wireless channels are described. This is followed by a discussion of cellular network models, including terminal mobility, noise, and interference models. Finally, physical layer abstractions are discussed. The second part introduces important metrics that are commonly used for the evaluation of cellular networks. The third part of Chapter 3 gives a brief introduction to the broad field of optimization. Evolutionary algorithms and in particular genetic algorithms are detailed, which are a powerful tool for solving combinatorial optimization problems. In addition, Appendix B gives a glossary of important terms in the area of genetic algorithms.

The state-of-the-art in interference coordination is elaborated in Chapter 4. This goes along with a classification of the discussed algorithms with respect to their time scale of operation and their degree of distribution. Furthermore, the concept of an interference graph as a powerful tool for interference coordination in cellular networks is proposed in the same Chapter. The properties of this graph are studied. Different strategies for resource assignment based on this graph are introduced, thereby applying well-known graph coloring algorithms and variations thereof. The Chapter is concluded by an overview of related work in the area of ad hoc and mesh networks.

Based on the previously introduced interference graph, Chapter 5 proposes a novel distributed and dynamic interference coordination algorithm. After an overview of the basic principle, the involved optimization problems are formulated. Subsequently, solution approaches to these problems based on graph-coloring and genetic algorithms are presented, and their implementation complexity is discussed. Furthermore, the integration of fine-granular Quality of Service (QoS) requirements into the algorithms is considered. The Chapter is concluded by a discussion of signaling and measurement procedures required for the implementation of the algorithm in real networks. The flow charts of the here developed crossover operators can be found in Appendix C.

The scenario and system model for the performance evaluation is described in Chapter 6. This includes the configuration of previously discussed models as well as the object-oriented model of the underlying IEEE 802.16e system. Furthermore, basic algorithms of the simulated network are outlined, such as the algorithms for burst profile management and scheduling. Finally, the parametrization of the simulation environment for Monte-Carlo simulations and event-driven simulations is given.

Chapter 7 contains an in-depth performance evaluation of the considered systems based on fully dynamic event-driven simulations. First, a comparison of uncoordinated reference systems with a frequency reuse of 1 and 3 with both sector antennas and beamforming antennas is given in a full-buffer traffic scenario. Subsequently, globally coordinated systems are evaluated, providing an estimate of the upper performance bound. This is complemented by a study of locally coordinated systems, which are enhanced by means of Fractional Frequency Reuse (FFR). Optimal parameters for the genetic algorithm involved in the proposed distributed interference coordination scheme are then derived by means of Monte-Carlo simulations. Furthermore, its convergence behavior is studied. Throughput and coverage results are then obtained and compared to the uncoordinated and globally coordinated reference systems. The Chapter is concluded by a performance comparison for the case of uneven load distributions, and by a performance evaluation with an on-off traffic model. Complementary simulation results can be found in Annex D.

Finally, the conclusion and an outlook to possible future work are given in Chapter 8.

# 2 Cellular Networks

This Chapter covers the fundamental concepts of wireless and cellular systems as well as their specific implementation in IEEE 802.16 based networks. After a brief overview of the history of cellular networks in Section 2.1, Sections 2.2 to 2.7 deal with the characteristics of the wireless channel and with basic concepts for data transmission in cellular networks. Special focus will be put on techniques that are currently under consideration within standardization bodies or that have recently been used for emerging cellular systems. Section 2.8 concludes the Chapter with an introduction to the state-of-the-art IEEE 802.16 technology.

## 2.1 History and Development of Cellular Networks

The first written reference describing the basic concept of a cellular telephone network is a technical memo by Douglas Ring, a researcher in the Bell Telephone Laboratories, from the year 1947 [Rin47]. It then took several decades until the first operative cellular network, the Nippon Telephone and Telegraph (NTT) system, was put into operation in 1979 [Stü00], followed by the Nordic Mobile Telephone (NMT) 900 system in 1981 [Stü00], and the Advanced Mobile Phone Service (AMPS) in 1983 [MD79]. By then, the cellular concept as described in [Rin47] had been amended by cell sectorization and other essential concepts, such as handover [MD79]. From then on, a large number of incompatible analog cellular systems were deployed, such as the German C-Netz in 1985. Roaming across country borders was not possible in most cases.

In 1982, the Groupe Spécial Mobile (GSM) was installed by the European Conference of Postal and Telecommunications Administrations (CEPT) in order to develop a pan-European digital cellular telephone system. This work resulted in the installation of the first digital cellular networks according to the Global System for Mobile communications (GSM) standard in 1991. Succeeding the first generation analog networks, the early fully digital networks such as GSM are seen as the second generation of cellular networks. GSM focuses on voice services, but also provides limited data services, such as Short Message Service (SMS) or Circuit Switched Data (CSD) with relatively low data rates. The increasing popularity of the Internet and the demand for multimedia data services led to the development of the General Packet Radio Service (GPRS), a packet switched extension to GSM, which was enhanced further towards higher data rates under the name Enhanced Data rates for GSM Evolution (EDGE). Similar developments took place in the USA with the introduction of IS-54/136 and IS-95.

Parallel to the enhancement of second generation cellular networks towards higher data rates for packet switched data, work began on third generation cellular networks, aiming at higher

data rates for multimedia services. While the development of first and second generation digital cellular networks resulted in a variety of incompatible systems, the goal for the third generation was a unique standard for worldwide interoperability. This led to the foundation of the 3rd Generation Partnership Project (3GPP), which is an international consortium of operators and vendors. Its original purpose was the development of the third generation cellular network standard, but it has also taken on the maintenance and further development of the GSM, GPRS, and EDGE standards. The work of the 3GPP resulted in the specification of the Universal Mobile Telecommunication System (UMTS), which is further enhanced for packet switched services by High Speed Packet Access (HSPA). Currently, 3GPP is working towards evolved HSPA with increased data rates and a simplified network architecture. Furthermore, a new air interface is currently standardized by the 3GPP under the name *Long Term Evolution* (LTE), aiming at even higher data rates. Accompanying the work on an advanced radio interface, a separate working group within the 3GPP studies aspects of the *System Architecture Evolution* (SAE), aiming at an all-IP based evolved packet core network. Again, in the USA, a parallel development took place with the introduction of cdma2000 and its packet switched extension Evolution Data Only (EV-DO).

A common global standard for cellular networks was also the goal of the International Telecommunication Union (ITU). It quickly became obvious that this was an unrealistic goal due to too many competing standardization efforts, such as cdma2000 developed by 3GPP2, the North American counterpart to 3GPP. As a consequence, the ITU began to promote common architectural principles under the name International Mobile Telecommunications – 2000 (IMT-2000). Several different radio access technologies have been approved as IMT-2000 technologies. Among these are Universal Terrestrial Radio Access – Frequency Division Duplex (UTRA-FDD) within UMTS as well as EDGE.

Later on, in October 2007, a subset of the IEEE 802.16e standard was included in IMT-2000. In particular, this was the Time Division Duplex (TDD) radio interface based on Orthogonal Frequency Division Multiple Access (OFDMA). IEEE 802.16e resulted as an extension of the IEEE 802.16 standard for fixed broadband wireless access, with the primary goal of delivering broadband mobile access. The IEEE 802.16 standard family has been developed within the IEEE in parallel to the 3GPP activities and can be seen as a late competitor to HSPA. However, while 3GPP-based systems define a full system architecture from the air interface back to the core network including management functions, the IEEE 802.16 standard focuses on the air interface technology. IEEE 802.16e is promoted by the WiMAX Forum, an industry-led organization that specifies system profiles, supplemental architectures, and interoperability tests. It is therefore also responsible for the certification of *WiMAX-compliant* devices.

The inclusion into IMT-2000 was a major success for IEEE 802.16e, giving it the same privileges as the other IMT-2000 technologies. This includes access to precious frequency resources, which are reserved for IMT-2000 technologies in some countries. Recently, ideas are being collected within the 3GPP for systems beyond IMT-2000. Alike, the IEEE 802.16e Task Group m (TGm) works on the further development of the IEEE 802.16 air interface. These next generation technologies are also handled under the term IMT-advanced, the dedicated successor of IMT-2000.

## 2.2 The Cellular Concept

Figure 2.1 shows a schematic view of a cellular mobile network. One *cell* is served by one *base station*, equipped with a *radio transceiver*, where the corresponding *base station site* is located in the center of the cell. The shape of a cell is determined by the area where the signal of the serving base station is the strongest among all received signals. If all base station sites are placed on a regular grid as in Fig. 2.1, and under uniform radio propagation conditions, the cells have a hexagonal shape. Naturally, in a real world scenario, the cell shapes are irregular due to uneven terrain, buildings, and the like. If all base stations transmit on the same radio resources, base stations will cause interference to mobile terminals in neighboring cells, as it is illustrated in Fig. 2.1. It may therefore not be possible to reuse the same channel in adjacent cells, where a channel may be defined by the periodic occupation of a particular air interface resource. This leads to idle channels and a waste of resources. Hence, an optimal assignment of resources to cells needs to be found during the network planning process, which may for example take into account spatial traffic distributions [Gam91].

In order to increase the *spatial channel reuse*, cells can be reduced in size while at the same time reducing the transmit power of each base station. A similar effect can be achieved by sectorizing the cells, which was originally proposed by another Bell Labs engineer, Phil Porter [Hoc99]. In a sectorized cell, the radio transceivers are at the edges of the respective *cell sectors*, as illustrated in Fig. 2.2. This increases the spatial channel reuse along with the spectral efficiency of the system [Cha92]. A cell can easily be sectorized by installing several transmitters at one base station site. Typical numbers are three or sometimes six transceivers per base station site, where other numbers are possible depending on the terrain to be covered. Further details can be found in [Stü00].

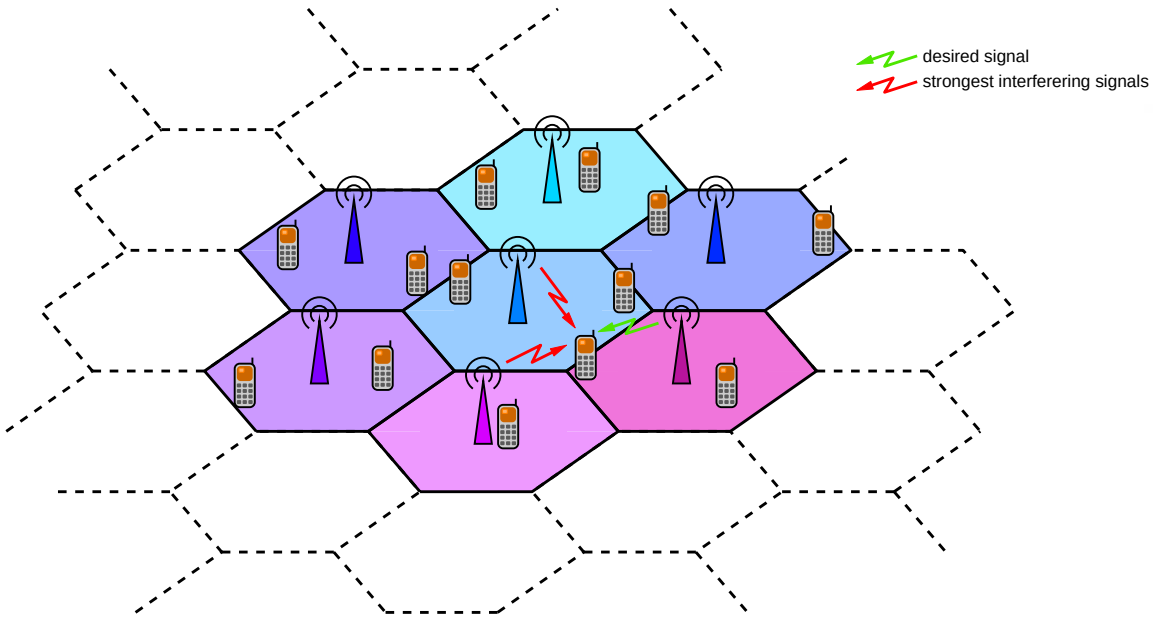



**Figure 2.1:** Illustration of a cellular system with omni-directional antennas.

The problem of interference is worst for mobile terminals at the cell edge, where the signal of their serving base station is relatively weak compared to the interfering signals. In order to serve these cell edge terminals, a base station needs a disproportionate amount of resources. The same resources could be used to serve mobile terminals in more favorable positions with higher data rates, such as terminals close to the base station. This leads to a tradeoff between fairness and the aggregate throughput within a cell sector. The aggregate throughput is often normalized to the system's bandwidth, resulting in the overall *spectral efficiency*. This metric is usually given in Bits per Hz per second, and allows a convenient comparison of radio technologies independent of the available bandwidth. The performance at the cell edge can be expressed by the throughput of cell edge terminals or by the spectral efficiency in the cell edge areas. Both the aggregate and the cell edge throughput have to be considered when evaluating the performance of cellular networks since it is trivial to increase one metric at the cost of the other.

Cellular networks can be designed for synchronous or asynchronous operation. In a synchronous network, the frame starting times and frame durations are synchronized among all base stations. This can be achieved by help of the Global Positioning System (GPS), as it is done in cdma2000 [Sha02]. Synchronization can also be done based on signals received from neighboring base stations [Ger96] or by help of signals received from mobile terminals [Ots96]. Additionally, a central node may assist in establishing synchronization based on base station measurements [Hul01]. In contrast, asynchronous networks do not need synchronization on the frame level. The decision for synchronous or asynchronous network operation depends on various different factors, above all on the radio technology. Some technologies such as cdma2000 mandate a synchronous network operation. Other technologies may get along without network synchronization, even though synchronization would increase their performance. In this case, the tradeoff between performance enhancement and added complexity for synchronization is left to the network operator. Finally, network synchronization may have no impact at all for some radio technologies.

## 2.3 The Wireless Channel

In a cellular network, any pair of transmitter and receiver is connected by a wireless communication channel. The data transmission over a wireless channel is one of the biggest challenges when designing a mobile communication system. This results from some very unfavorable characteristics of this type of channel. First of all, the channel is noisy and prone to distortions from

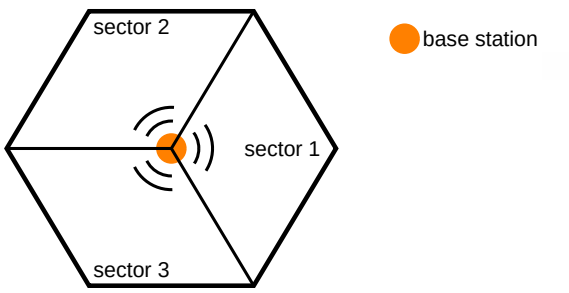



**Figure 2.2:** Sectorization of a cell into three sectors, each with 120° opening angle.

other external and internal effects, such as interference. Second, the signal attenuation is relatively high, and it may be variable over time and frequency. The following sub-sections focus on the transmission characteristics of a single wireless channel, specifically on the signal attenuation. The problem of interference in cellular systems will be treated later-on in Section 2.6.

### 2.3.1 Distance dependent path loss

The path loss describes the attenuation of the transmitted signal and is defined as follows (in linear notation):

$$L_{p\ (\text{lin})} = \frac{P}{S}\ , \tag{2.1}$$

where $P$ is the transmit power and $S$ is the received signal power. In a flat earth model with no obstacles, the path loss at distance $d$ can be approximated as

$$L_{p\ (\text{lin})} \approx \left[\left(\frac{h_b h_m}{d^2}\right)^2\right]^{-1}\ , \tag{2.2}$$

provided that $d \gg h_b h_m$, with $h_b$ being the height of the base station antenna and $h_m$ the height of the terminal's antenna. It can be seen that the signal power decays with the fourth power of the distance as compared to the second power in free space. In a real environment, the distance dependent path loss is generally given by

$$L_{p\ (\text{lin})} = \frac{d^{\beta}}{k}\ . \tag{2.3}$$

$k$ is a constant of proportionality, and $\beta$ is the *path loss exponent* which needs to be determined by measurements. Typical values of $\beta$ range between 3 and 4.4 in both urban and suburban environments [Stü00]. Hence, the signal power quickly decays as the distance to the base station becomes larger.

The distance dependent path loss describes the long term average path loss which is obtained when the mobile terminal moves at a certain distance $d$ from the base station. Shadowing and fast fading, which are introduced in the following sub-sections, describe long and short term variations around this mean value.

### 2.3.2 Shadowing

Besides the distance dependent path loss, buildings, trees, and other obstacles reduce the received signal power. If the mobile terminal is moving, this signal variation mostly occurs on a long time scale. It is therefore referred to as *shadowing* or *slow fading*. In some cases, for example if the mobile station turns around the corner of a building, the signal variation due to shadowing may also occur quickly. Nevertheless, shadowing describes the signal variation averaged over short to medium time periods and the whole considered frequency band. Fast fading, which is described in the following sub-section, is the effect of time and frequency dependent signal variation around this mean value on a short time scale that occurs in a multipath propagation environment.

### 2.3.3 Multipath propagation

Multipath propagation, as illustrated in Fig. 2.3, is inherent to all wireless environments unless highly directive antennas are used. The radio waves from the base station may reach the mobile terminal on a direct Line-of-Sight (LoS) path. Often, the direct LoS is obstructed, for example by buildings, and the mobile terminal can only be reached indirectly via reflection, scattering, or diffraction of the radio waves [Rap02]. The latter make up the Non-Line-of-Sight (NLoS) components of the radio signal, which add up together with a possible LoS component at the mobile terminal. All components, which are also referred to as echos, have an individual propagation delay $\tau$ and phase shift $\theta$.

Two basic effects result from multipath propagation. First, if the time difference between the arrival of the first and the last dominant echo (also referred to as *delay spread*) is in the order of a data symbol period or longer, *Intersymbol Interference (ISI)* distorts the transmitted signal. Second, depending on $\tau$, $\theta$, and the location of the mobile terminal, the echos add up constructively or destructively at the receiver. This leads to variations of the received signal power of as much as $20-30$dB within a distance in the order of the wavelength [Ber00]. This effect is known as *fast fading*.

While fast fading is dependent on the position of the mobile terminal, the fast fading variations in time result mainly from the movement of the mobile terminal. Moreover, the Doppler effect caused by the mobile terminal's movement adds another difficulty. The Doppler shift $f_D$ is dependent on the direction of arrival of an echo relative to the direction of movement. If the terminal is moving with velocity $v$, the maximum Doppler shift $f_{D,max}$ that can occur in the scenario is

$$f_{D,max} = \frac{v}{\lambda} = v\frac{f_C}{c} \quad , \tag{2.4}$$

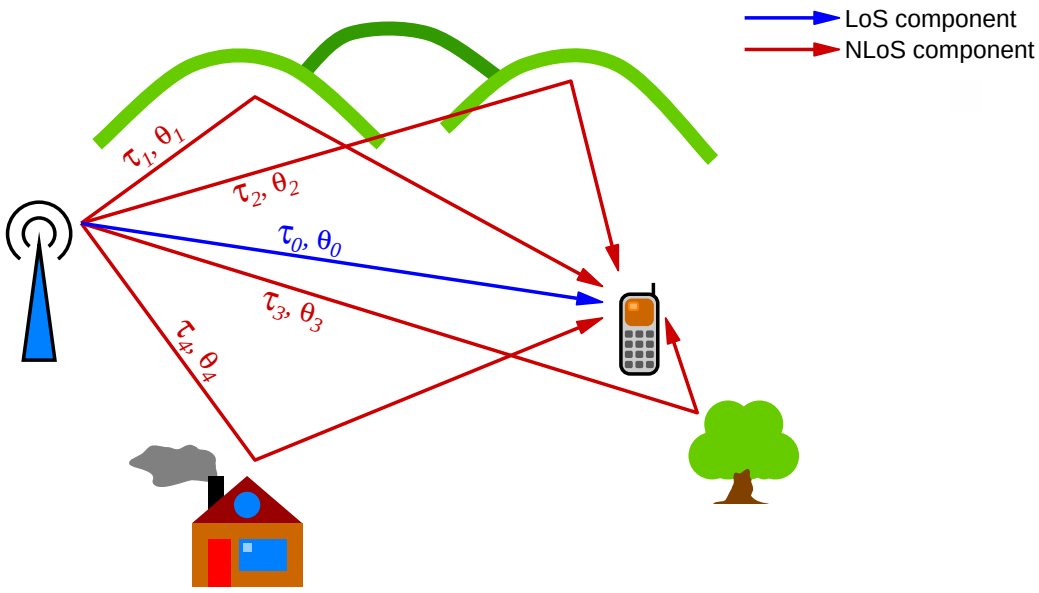



**Figure 2.3:** Illustration of multipath propagation.

where $\lambda$ is the wavelength of the radio waves, $c$ is the speed of light, and $f_C = c/\lambda$ is the radio signal's carrier frequency.

Even in wide-band systems, it is sufficient to consider only the carrier frequency $f_C$ for the Doppler effect. Nevertheless, fast fading varies depending on the frequency. If the delay spread is much smaller than the data symbol period, the wireless channel is said to experience *flat fading*. This means that the frequency response of the wireless channel is approximately constant within the signal bandwidth. Otherwise, the channel exhibits *frequency-selective fading*, meaning that the frequency response varies within the signal bandwidth. The magnitude of these fluctuations is in the same order as those of the location-dependent variations.

At first sight, multipath propagation seems like a nuisance. However, NLoS propagation is often the only way to reach a mobile terminal, since a direct LoS is not given in many cases. Moreover, multiple echos increase the average received signal power. Last but not least, since all mobile terminals experience independent fading, it is likely that at least one terminal experiences constructive interference. In a packet-switched system with time division multiple access (see Section 2.5.2), the system capacity can be increased by serving terminals whenever they experience particularly good fast fading conditions. This exploitation of *multiuser diversity* by means of *channel aware scheduling* may make data communication possible even if the average signal conditions would prevent it.

## 2.4 Modulation and Transmission Techniques

In a traditional digital communication system one or more carrier frequencies are modulated by a sequence of complex valued data symbols. The data symbols are taken from a fixed symbol alphabet, which results from the applied modulation scheme. The modulation scheme determines how a carrier frequency is modulated with respect to its amplitude and phase. All elements of the symbol alphabet can be mapped to a characteristic constellation diagram, representing all possible signal points of the modulation scheme. In a non-differential modulation scheme, a certain sequence of transmitted bits is directly mapped to a particular signal point, i.e., to a particular element of the symbol alphabet. Popular modulation schemes are Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM). These modulation schemes are further labeled as $n$-PSK and $n$-QAM. $n$ denotes the number of signal points in their constellation diagram, where $n$ is usually a power of two. A number of $\mathrm{ld}\,n$ bits can then be mapped to every signal point, allowing the transmission of $\mathrm{ld}\,n$ bits per baud period. Note that general numbers of $n$ are also possible, for example by puncturing of a superordinate error correction code [Nec04a, Nec04b].

The transmission of data symbols may be spread over a large frequency range by using several carrier frequencies in parallel. The data transmission thereby becomes more robust to channel fades or interfering signals on particular frequency blocks. One efficient realization of a wide-band communication system is Orthogonal Frequency Division Multiplexing (OFDM). In OFDM, a large number $N_{sc}$ of carrier frequencies with an equidistant spacing of $\Delta f$ is used for data transmission. If the spacing of these *subcarriers (SC)* with frequencies $f_i$ is chosen as

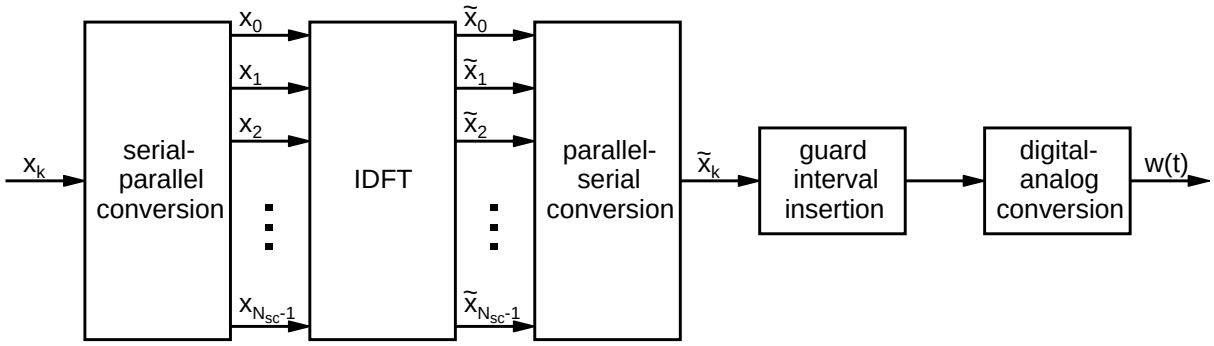



**Figure 2.4:** Block diagram of an OFDM transmitter realized with an IDFT.

$\Delta f = 1/T_S$, with $T_S$ the symbol duration, then they are orthogonal with respect to the following orthogonality condition:

$$\frac{1}{T_S}\int_0^{T_S} e^{j2\pi f_i t} e^{j2\pi f_l t} dt = \frac{1}{2\pi}\int_0^{2\pi} e^{ji\tau} e^{jl\tau} d\tau = \delta_{i-l} \ , \tag{2.5}$$

where $\delta_k$ is the discrete delta function. The orthogonality condition ensures that there is no interference among the subcarriers (Intercarrier Interference, ICI). An OFDM transmission system can be realized by using a Discrete Fourier Transform (DFT) and its inverse (Inverse DFT, IDFT), which can efficiently be implemented using a Fast Fourier Transform (FFT). The block diagram of an IDFT-based OFDM transmitter is shown in Fig. 2.4. It processes a continuous stream of incoming complex-valued data symbols $x_k$ and generates the analog equivalent base-band signal $w(t)$. First, the incoming data symbols $x_k$ are grouped into blocks of $N_{sc}$ symbols. A serial-to-parallel conversion is used to map one block to a parallel vector, which is then transformed to the time domain by the IDFT block. After a parallel-to-serial conversion, a guard interval is appended to each block of $N_{sc}$ symbols from the IDFT. The resulting block of symbols is referred to as OFDM-symbol. It is transmitted within one baud period after a digital-to-analog converter generates the analog base-band signal $w(t)$. A receiver can employ the inverse functions, in particular a DFT, to demodulate the signal.

The structure of an OFDM-symbol is shown in Fig. 2.5. The already mentioned guard interval is often a cyclic continuation of the usable portion $T_U$ of the OFDM-symbol, where the length $T_G$ of the guard interval is a system parameter[1]. In this case, the guard interval is also referred to as Cyclic Prefix (CP). When the signal $w(t)$ is transmitted via its equivalent base band channel with impulse response $h(t)$, the received signal $y(t)$ can be calculated as a convolution:

$$y(t) = w(t) * h(t) + n(t) \ , \tag{2.6}$$

with $n(t)$ being Additive White Gaussian Noise (AWGN). At the receiver side, the received signal $y(t)$ is sampled, making possible a discrete representation of equation (2.6):

$$\tilde{y}_k = w_k * h_k + n_k = \tilde{x}_k * h_k + n_k \ . \tag{2.7}$$

If the guard interval is longer than the channel impulse response, then equation (2.7) constitutes a cyclic convolution of the transmitted signal with the channel impulse response. After the

[1] Note that instead of the cyclic prefix zero padding may be used [Muq02].

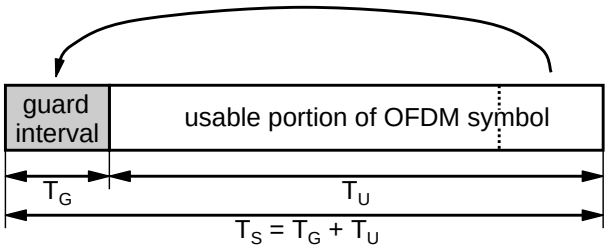



**Figure 2.5:** Schematic view of a transmitted OFDM symbol.

DFT at the receiver, this convolution results in a simple multiplication of the transmitted data symbols $x_k$ with the channel frequency response $H_k$ at a particular time instant, with $k$ being the subcarrier index. Consequently, the received signal can easily be equalized at the receiver. Furthermore, the signal is only distorted by ICI if $h(t)$ varies during the transmission of one OFDM symbol. For low speeds, this variation is very small so that ICI can be neglected.

In order to perform a coherent demodulation of the received data symbols, $H_k$ needs to be known. This requires a channel estimation at the receiver side, which can conveniently be done based on pilot symbols that are inserted into the two-dimensional grid of subcarriers and OFDM-symbol time steps. The pilot symbols are taken from a known sequence such that the receiver can calculate the channel transfer function at the respective time and frequency. If the two-dimensional sampling theorem is fulfilled, the receiver can estimate the channel transfer function at all other time and frequency steps by means of low-pass filtering.

The preceding explanations illustrate the inherent robustness of OFDM against multipath-propagation and the related ISI in the time domain. By choosing the guard interval at least as long as the largest possible delay spread in the scenario, interference between OFDM symbols can be avoided. Since a longer guard interval also implies a reduced spectral efficiency, OFDM-based systems usually foresee variable system configurations that allow choosing $T_G$ depending on the environment where the system is deployed.

A major drawback of OFDM is the high Peak to Average Power Ratio (PAPR) of the transmitted signal. This causes problems for the analog part of the transmitter, where amplifiers have to be operated with significant head room to account for power peaks without reaching the non-linear domain of the amplifiers. Several PAPR reduction techniques, such as clipping and precoding, exist. Further details on this issue and on OFDM in general can be found in [Li06].

## 2.5 Duplexing Methods and Multiple Access

The transceivers of the mobile terminals and the base stations all access the same shared transmission medium. Therefore, methods need to be applied in order to separate first the forward and backward direction (downlink and uplink), and second the transmissions of and for different mobile terminals. The former are referred to as duplexing methods and are overviewed in Section 2.5.1. The latter are known as multiple access schemes, which are detailed in Section 2.5.2.

### 2.5.1 Duplexing methods

There are two duplexing methods that are generally being applied in mobile networks, namely Frequency Division Duplex (FDD) and Time Division Duplex (TDD). In an FDD system, up- and downlink are found in separate frequency bands. A *paired spectrum* is required for this mode of operation. TDD systems operate the up- and downlink in the same frequency band. The two directions are separated in time by dividing the MAC frame into a downlink and an uplink part. Both concepts have advantages and disadvantages [Cha06]. 3GPP LTE is pushing the FDD variant, whereas IEEE 802.16e focuses on TDD operation. Therefore, the available spectrum may play a central role for an operator's technology choice.

### 2.5.2 Multiple access schemes

There are four basic multiple access schemes known from literature, namely Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA), and Space Division Multiple Access (SDMA). They are illustrated in Fig. 2.6. In an FDMA system, different users are separated in frequency. For this, the available spectrum is divided into multiple smaller frequency bands. In TDMA, users are separated in time by dividing a transmission frame into several smaller time slots. In CDMA, every user is assigned a code sequence, and the transmitted bits are multiplied by the code sequence, resulting in a longer sequence of chips to be transmitted. Users can then be separated if the code sequences are orthogonal, or if they have small cross-correlations at all delays, even if all users transmit at the same time and on the same frequency. Interleave Division Multiple Access (IDMA) is a variant of CDMA where different users are separated by different interleavers instead of different orthogonal codes [Pin03]. Multi-User Detection (MUD, see Section 2.6.2) can then be applied at the receiver to distinguish the different users. Finally, in SDMA, users are separated in space by means of directional antennas while they can transmit at the same time, on the same frequency, and possibly using the same codes. A detailed description of all basic multiplexing schemes can be found for example in [Rap02].

All above multiplexing schemes can be combined. For example, GSM employs a combination of FDMA and TDMA. More recently, Orthogonal Frequency Division Multiple Access (OFDMA) has gained increasing interest. OFDMA essentially is a combination of FDMA and TDMA, where the FDMA is based on an underlying OFDM transmission system. Two fundamentally different resource allocation strategies are possible in an OFDMA system. First, in a *frequency selective* allocation as shown in Fig. 2.7(a), the different data streams of the various mobile terminals are assigned adjacent subcarriers and time-slots, such that every data stream

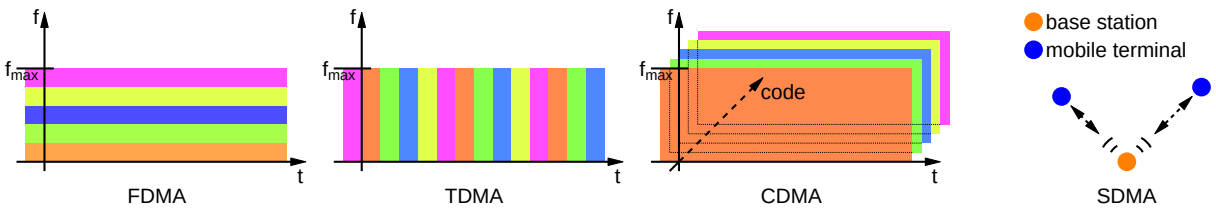
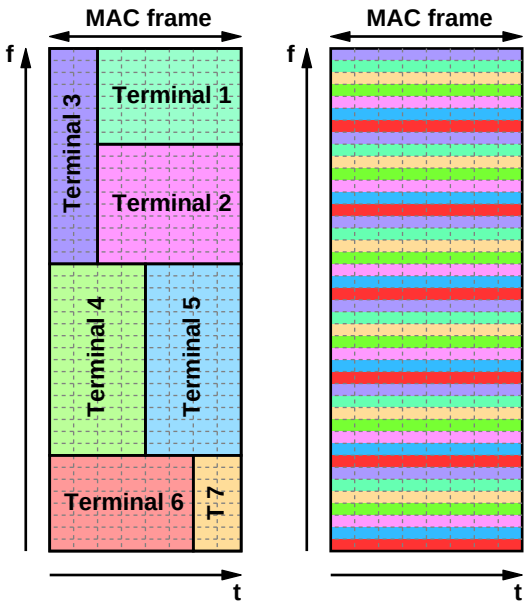



**Figure 2.6:** Basic multiple access schemes.

is served on a contiguous "area" in the frequency-time plane. This allows the implementation of frequency selective scheduling schemes, which, for every mobile terminal, try to allocate resources on subcarriers with good fast fading conditions while avoiding bad subcarriers. Besides the first possibility mentioned in Section 2.3.3, this is the second possibility to exploit multiuser diversity, and just like in the TDMA case, it can significantly increase the system capacity. However, it requires knowledge of the channel state, hence such a scheme is only feasible if the channel does not vary too quickly, i.e., if the velocity of the mobile terminal is not too high. If no detailed channel state information is available or if mobile terminals move too fast, a *frequency diverse* allocation as illustrated in Fig. 2.7(b) is possible. Here, the subcarriers allocated to one particular data stream are spread across the spectrum in order to equalize the frequency selective fading.

## 2.6 Noise and Interference

The capacity of cellular systems is essentially limited by two fundamental factors, namely noise and interference. The most dominant noise type is thermal noise, which depends on the hardware components and the environment. It is introduced by all parts of the system, such as antennas, amplifiers, filters, etc. Moreover, the ever present background radiation originating from cosmic or terrestrial sources adds up to the overall noise. Interference depends on the applied radio technology and the cellular layout. It can be divided into two major classes. First, *intra-cell interference* refers to all interference sources within the same cell sector. Second, *inter-cell interference* describes interference from other cells. Intra-cell interference is present regardless of the cellular layout, even in single isolated cells. In contrast, inter-cell interference becomes weaker the farther the base stations are spaced apart. Depending on whether noise or




(a) frequency selective (b) frequency diverse

**Figure 2.7:** Illustration of Orthogonal Frequency Division Multiple Access (OFDMA).

inter-cell interference is the dominating effect, the cellular system is referred to as *noise limited* or *interference limited*. Cellular systems in densely populated areas are usually interference limited. Noise and interference play a crucial role for the reception quality of a radio signal, which can be characterized by the *Signal-to-Interference-and-Noise Ratio (SINR)*, or, in the case of an interference limited system, approximately by the *Signal-to-Interference Ratio (SIR)*:

$$SINR = \underbrace{\frac{S}{N+I}}_{SINR} \overset{N \ll I}{\approx} \underbrace{\frac{S}{I}}_{SIR} , \tag{2.8}$$

with $S$ being the total received signal power, $I$ the sum power of all interfering signals, and $N$ being the noise power.

Sources of intra-cell interference are manifold. In FDMA systems, *adjacent channel* interference is caused by nearby frequencies leaking into the considered frequency band. This may for example be due to imperfect receive or transmit filters. The already mentioned ISI is caused by multi-path propagation and can only be avoided by choosing the data symbol period substantially longer than the channel's delay spread. Otherwise, a channel equalizer has to be used in the time-domain. In OFDM systems, ISI in the time-domain can efficiently be mitigated by the cyclic prefix. However, if the channel's impulse response varies during the period of one OFDM symbol, the subcarriers loose orthogonality, causing ICI. In CDMA systems, loss of orthogonality between different codes due to multipath-propagation leads to intra-cell interference between different users, known as Multiple Access Interference (MAI).

Like intra-cell interference, the characteristics of inter-cell interference depend on the applied technology. In CDMA systems like cdma2000 or UMTS, inter-cell interference occurs since neighboring cells are not separated by orthogonal codes, but by Pseudo Noise (PN) sequences. In FDMA or TDMA systems, inter-cell interference is caused by neighboring base stations transmitting on the same frequencies or time slots, i.e., on the same channels. This is also known as *co-channel interference*. In an interference-limited OFDMA scenario, inter-cell interference will be dominating compared to intra-cell interference. Therefore, in the remainder of this monograph, the focus will be on inter-cell interference.

The situation of a mobile terminal with respect to the inter-cell interference situation can be characterized by the *geometry factor* [R1-050738, R4-060117]. The geometry factor for a particular mobile terminal is defined as the fraction of total power received from the desired base station and the total other-cell interference plus noise. It can be formally defined as follows [Web08]:

$$G = \frac{E[S]}{E[I]+N} , \tag{2.9}$$

where $E[S]$ is the expected value of the received signal, $E[I]$ is the expected value of the total received interference, and $N$ is the noise. An estimate of the geometry factor can be obtained from pilot measurements [R1-050896]:

$$G = \frac{S_{serving}}{N + \sum I_{non-serving}} \approx \frac{P_{pilot,serving}}{N + \sum P_{pilot,non-serving}} , \tag{2.10}$$

with $S_{serving}$ the total received power of the serving base station, $I_{non-serving}$ the total received power of an interfering base station, and $P_{pilot}$ the received power of the pilots.

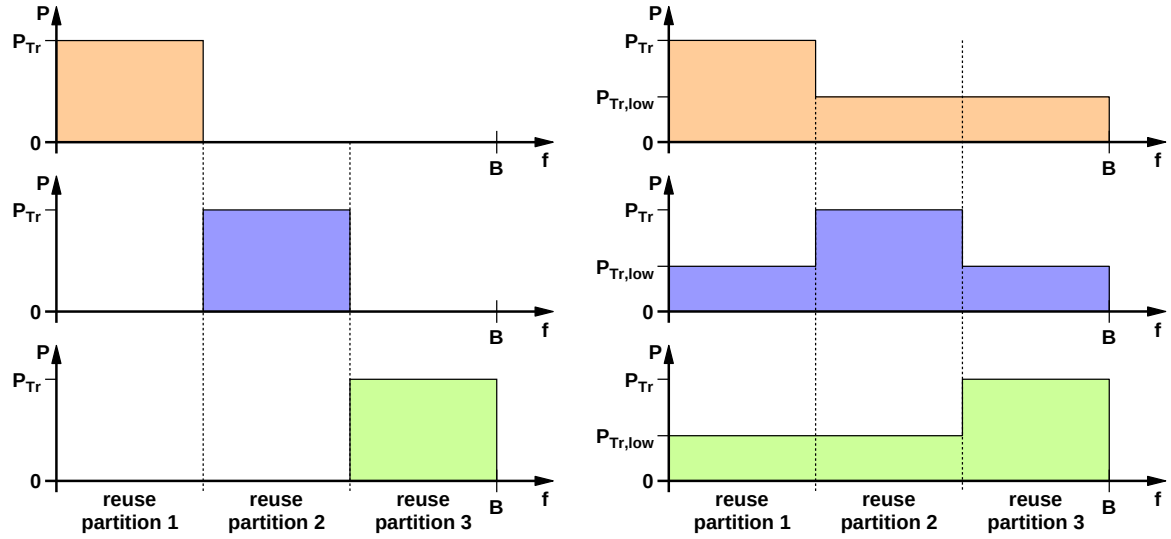
In interference limited systems, inter-cell interference is a serious capacity and coverage limiting issue. Depending on the air interface technology, it can be addressed by one of the various interference mitigation techniques. An overview of these techniques will be given in the following sub-sections.

### 2.6.1 Frequency reuse schemes

Conventional FDMA systems combat inter-cell interference by avoiding the reuse of the same frequency bands in adjacent cells. A certain frequency band is therefore only used in every $r$-th cell. $r$, or sometimes also $1/r$, is known as the *frequency reuse factor*. For example, in a scenario with three cell sectors per base station, a reuse factor of $r = 3$ is obtained by assigning disjoint frequency bands to all three cell sectors. This concept is also known as *hard frequency reuse* and is illustrated in Fig. 2.8(a). The overall bandwidth $B$ is divided into three reuse partitions, and every partition is served with a maximum transmit power $P_{Tr}$. An increase of the frequency reuse factor commonly increases the cell edge performance while at the same time reducing the overall spectral efficiency. In [R1-073567], the spectral efficiency at the cell edge was analytically estimated to increase by a factor of two when moving from $r = 1$ to $r = 3$ in a hexagonal cellular scenario with omnidirectional antennas.

*Soft frequency reuse*, which is sketched in Fig. 2.8(b), is a generalization in which every cell sector can transmit on all frequencies. However, the maximum transmit power is restricted to $P_{Tr,low}$ in the reuse partitions which would be unused with hard frequency reuse. It has thereby been recommended that the power reduction needs to be applied not only to data symbols, but also to pilot and control symbols [R1-070611].

Hard and soft frequency reuse schemes can be combined with other, more advanced concepts, such as reuse partitioning (see Section 4.5.2). They are sometimes also referred to as full trans-




(a) hard frequency reuse (b) soft frequency reuse

**Figure 2.8:** Illustration of hard and soft frequency reuse with $r = 3$.

mit power isolation, and partial transmit power isolation, respectively [R1-061391], thereby alluding to the way the interfering cells are isolated from each other. A generalization of frequency reuse is fractional loading [R1-073567] in which the whole frequency band is available in all cell sectors, but the duty cycle is smaller than 100% on a particular time/frequency resource. This leads to a statistical gain and essentially allows the realization of arbitrary frequency reuse factors.

### 2.6.2 Interference cancelation

In contrast to frequency reuse, which tries to avoid interference in the first place, interference cancelation aims at removing interference from the received signal. Two classes of co-channel interference cancelers exist [Hoe05]. *Filter-based* approaches try to mitigate co-channel interference by means of linear filters and interference models. In contrast, *Multi-User Detection (MUD)* directly includes the interfering signals in the decoding process. This is done by jointly decoding the signal of interest and the interfering signals, or by decoding and subtracting the interfering signals from the signal of interest. The implementation of interference cancelation does not require any modifications of the system standard, making it an attractive technique. It has successfully been applied to both CDMA systems [DH95] and TDMA systems [Hoe05].

Further possibilities for interference cancelation exist if the mobile terminals are equipped with multiple receive antennas. For example, a mobile terminal with two receive antennas can treat the signal of interest and the interfering signal as a Multiple Input Multiple Output (MIMO) $2 \times 2$ system and thus cancel one interfering source in the ideal case. This space-domain interference cancelation can be transferred to the frequency domain if only one transmit antenna is available. Within the 3GPP LTE standardization, [R1-050829, R1-051085] propose *coordinated symbol repetition*, where neighboring base stations transmit a data stream twice on two different frequencies. This allows interference cancelation by applying the same signal processing algorithms as in the MIMO case.

### 2.6.3 Interference mitigation using antenna arrays

Antenna arrays allow the application of beamforming techniques to mitigate co-channel interference. This can be regarded as a special case of the previously mentioned filter-based interference cancelation. An optimal radiation pattern corresponds to an optimal choice of the weights for the antenna array processing. Antenna array systems will be further detailed in Section 2.7.3.

While it is well feasible to install multiple antennas at the base station, this is often not possible in mobile terminals due to size constraints. For the uplink, *cooperative communication*, also known as *virtual Multiple Input Multiple Output* (virtual MIMO), has therefore been studied [Nos04, Jan04, Jun07], where multiple close-by mobile terminals either relay the information overheard by neighboring terminals, or even form a full virtual antenna array. Coordination among these terminals may be advantageous or necessary. It can be achieved remotely by the base station, or by using short range communication techniques, such as Bluetooth, IEEE 802.11, or cable based techniques like power lines [Kuh06].

### 2.6.4 Interference randomization

Frequency reuse aims at reducing the power of the interfering signals. Alike, interference cancelation tries to eliminate interfering signals in the received signal. In contrast, interference randomization, which is also known as interference whitening, tries to make interference appear like background noise, i.e., it averages the interference across the data symbols of a data block or the whole frequency band. This is widely applied in CDMA systems, where neighboring cells are separated by pseudo noise sequences. In FDMA/TDMA systems like GSM, frequency hopping in combination with interleaving can achieve a whitening effect. In OFDMA-based systems, interference randomization can be achieved in a number of different ways. First, if a frequency-diverse resource allocation is used, different spreading patterns can be used in adjacent cells. Second, combinations of CDMA and IDMA with OFDM have been proposed [Pin07]. IDMA-based schemes have also been proposed within the 3GPP LTE standardization [R1-050608].

### 2.6.5 Macro diversity

In contrast to the so far discussed schemes for interference reduction, macro diversity relies on the cooperation of neighboring base stations. When using macro diversity in the downlink, a mobile terminal receives data from several base stations of which it is within the receive range [Wei03]. In *transmit selection diversity*, the signal is always transmitted from the base station with the best momentary channel quality. In *simulcast*, the base stations transmit the same signal to the mobile terminal at virtually the same time. The mobile terminal then combines the different received signals to decode the transmitted data stream. Due to the independent fading of the received signals, a significant gain can be achieved. In the uplink, the signal from the mobile terminal is received by several base stations and then combined at a central point, such as the Radio Network Controller (RNC) in UMTS.

Macro diversity is relatively easy to implement in circuit switched networks, since a continuous data stream has to be transmitted from neighboring base stations. This allows for low complexity synchronization mechanisms. One example in existing circuit switched systems is the soft handover for Dedicated Channels in UMTS. It is also well feasible in a circuit or packet switched broadcast system, such as the Multimedia Broadcast Multicast Service (MBMS) [Par06]. The implementation of macro diversity is more difficult in a packet switched unicast system. In the uplink, multiple base stations can receive the signal and forward it to one master base station or a central node. This is already supported by UMTS High Speed Uplink Packet Access (HSUPA) systems. In the downlink direction, selection diversity is possible on a coarse time scale in UMTS High Speed Downlink Packet Access (HSDPA) [For05]. However, simulcast requires fully synchronized scheduling among the participating base stations, which is extremely difficult to achieve. Nevertheless, it has been proposed for HSDPA systems [Bar07], and it is also considered as a candidate technology for IMT-advanced in combination with MIMO techniques [R1-081877].

With respect to the performance of transmit selection diversity and simulcast in the downlink direction, a tradeoff exists between improved signal quality (achievable with simulcast compared to transmit selection diversity) and an increased amount of interference (which goes along with

simulcast). This issue has been investigated by Weiss [Wei03] based on a general evaluation of the reception conditions, and by Stähle in [Sta02] for UMTS handovers, where both works conclude that transmit selection diversity outperforms simulcast.

### 2.6.6 Interference coordination

Interference Coordination (IFCO), also known as Inter-Cell Interference Coordination (ICIC), has only recently gained attention to mitigate interference in cellular networks. In order to reduce interference, IFCO coordinates the usage of resources in neighboring cell sectors, either in a static, a semi-static, or in a dynamic fashion. Resources for coordination are space (with beamforming antennas), time slots, frequency ranges, code resources, or transmit power. For example, in an OFDMA system, IFCO can prevent two cell edge terminals in neighboring cell sectors to be served on the same time/frequency resource. While the main goal is the reduction of the interference, secondary goals determine exactly how the interference coordination is done. Secondary goals can for example be the maximization of the overall network capacity, or the maximization of the cell edge throughput. IFCO will be the focus of Chapters 4 and 5, and a performance evaluation of several IFCO variants will be given in Chapter 7.

## 2.7 Advanced Concepts for Data Transmission in Wireless Networks

Data transmission in cellular networks is a challenging task. In the past decades, various techniques have been developed in order to increase the overall spectral efficiency and transmission speed, as well as to mitigate interference and thus increase the cell edge performance. The following sub-sections introduce the most important concepts which are applied in today's networks.

### 2.7.1 Error detection and correction

As already indicated in Section 2.3, data transmission on a wireless channel is prone to errors. It is therefore necessary to apply error detection and correction schemes in order to protect the transmitted data. Error correction schemes can be divided in two classes. First, *Forward Error Correction (FEC)* refers to the application of error correction codes at the transmitter, such as Reed-Solomon block codes or convolutional codes. A decoder at the receiver side can then correct a certain number of errors. At lower layers, *soft decoding* [Pro07] can increase the performance by working on soft values instead of binary values, which additionally take into account the channel, noise, and interference situation for every received data symbol. The coding and decoding process is usually applied to finite length *FEC blocks*, even in combination with convolutional codes. Longer FEC-blocks allow for a better performance, but come at the cost of higher complexity and possibly a larger transmission delay. FEC introduces additional redundancy bits into the stream of information bits, resulting in the transmitted stream of coded bits. The *code rate* $r_c$ defines the number of added redundancy bits in the following way:

$$r_c = \frac{\text{number of information bits}}{\text{number of coded bits}} \quad . \tag{2.11}$$

The residual loss probability of an FEC block is typically quite high, demanding for a second layer of error protection. For example, in UMTS networks, the optimal loss probability in order to maximize network capacity lies between 10 and 30% [Hol00]. *Backward Error Correction (BEC)* as the second class of error correction schemes needs to be applied in order to reduce the residual error probability to an acceptable value. BEC is usually implemented by some sort of *Automatic Repeat Request (ARQ)*. With ARQ, the data stream is segmented into ARQ blocks, which are amended by a sequence number and an error detection code. The receiver can then detect corrupt or missing ARQ blocks and request a retransmission. Alike, correctly received blocks are acknowledged, allowing the sender to detect and retransmit completely lost blocks. Retransmissions can thus be initiated by both the receiver and the sender, where a combination yields the best performance [Mut04].

*Hybrid ARQ (HARQ)* extends regular ARQ by FEC. Classically, type I and type II HARQ is distinguished [Com84, Lin84]. In type I HARQ, an ARQ block is protected by an additional FEC. In type II HARQ, the ARQ block is first sent with no protective FEC. Subsequent retransmissions then do not contain the original ARQ block, but instead contain FEC parity-check bits. Therefore, with every retransmission, the amount of redundancy increases. Lin et al. [Lin84] already mention the possibility of soft-combining the retransmissions of an ARQ block. While regular ARQ completely discards corrupt ARQ blocks, HARQ with soft-combining stores incorrectly received ARQ blocks on the soft bit level and uses this information during the decoding process of subsequent retransmissions. *Chase combining* combines identical multiple transmissions of an ARQ block which is analog to type I HARQ. The analogon to type II HARQ is *incremental redundancy*, which uses additional parity bits obtained by retransmissions.

### 2.7.2 Link adaptation

Channel variations due to propagation loss and fading effects are a major challenge in mobile communications. Above all, this applies to variations due to fast fading. Channel variations can be faced by two different concepts. First, a fast variation of the transmit power can ensure an approximately constant receive power. This scheme is known as closed-loop power control and requires fast channel feedback from the receiver. It is widely applied in CDMA networks and allows data transmission with always the same modulation scheme. In contrast, a constant transmit power can be used in combination with *Adaptive Modulation and Coding (AMC)*. Based on feedback information from the mobile terminal, the transmitter selects the optimal Modulation and Coding Scheme (MCS) such that a certain Bit Error Ratio (BER) or target loss probability for FEC blocks (Block Error Ratio, BLER) is achieved. It is interesting to note that AMC has been proposed by J. Hayes already in 1968 [Hay68].

### 2.7.3 Advanced antenna techniques

In the past decade, Advanced Antenna Systems (AAS) have gained increasing attention as a powerful means to increase data link speeds and system capacity. Instead of the conventionally used single antennas, AASs employ an array of multiple antennas at the base station and/or the mobile terminal. Multiple antennas can be used for shaping the radiation pattern of the AAS in a beneficial way (beamforming), or they can be used for exploiting spatial diversity in a MIMO

system. MIMO is currently a hot-topic in communications research, as it allows to significantly increase the capacity of a data link with multiple transmit and multiple receive antennas. Several emerging wireless communication standards have standardized MIMO techniques, such as IEEE 802.11n, IEEE 802.16, or 3GPP LTE. An introduction to MIMO can be found in [Pau04] or [Spe05]. In the following, only beamforming systems will be considered, since they can efficiently be combined with interference coordination algorithms.

In beamforming, the radiation pattern of the antenna array is dynamically modified in order to achieve more or less gain into particular directions. This allows to suppress interfering signals and to amplify the signal of interest. Beamforming is achieved by phase-shifting the signals of the individual antennas in an array such that constructive or destructive interference can be observed in one particular direction. Beamforming antennas are therefore often referred to as *phased array antennas*. Figure 2.9 shows the basic block diagram of a phased array antenna consisting of $N_{ant}$ array elements. Before being transmitted via one of the array elements, the input signal is multiplied by the complex weighting factor $w_k$. The radiation pattern is therefore determined by the radiation pattern of the individual array elements, their locations, and by the complex weights $w_k$.

Beamformers can be classified into the following two categories:

- **Fixed Beamformers**

  Fixed beamformers, also known as switched beam systems, have a pre-configured set of radiation patterns which can be selected dynamically. They feature a relatively low complexity, since only a small set of pre-calculated weighting factors is required.

- **Adaptive beamformers**

  Adaptive beamformers allow to control the radiation pattern dynamically and without the limitation of the coarse steering granularity of switched beam systems. Adaptive beamformers are therefore more powerful, but also more complex than fixed beamformers.

A commonly used array type is the linear phased array, where the array elements are arranged on a straight line with a distance of $d$ between the elements. Alternatively, in a circular phased array, the array elements are arranged on a circle with a certain angular distance between them. This allows to steer the beam in any direction, while the linear array is restricted to the front direction. The element spacing $d$ determines the amount of mutual array element correlation. If $d$ is significantly smaller than the wavelength $\lambda$ (typically at most half the wavelength), the mutual antenna correlation is high, and the propagation loss is approximately equal for all array elements (except for the direction dependent phase difference). If $d$ is large, i.e., in the order of several wavelengths, the mutual antenna correlation is low, and the propagation loss cannot be assumed equal anymore. There exist several techniques to control the radiation pattern, which will roughly be outlined in the following. Further details can be found for example in [Mil05].

- **Beam steering**

  In beam steering, the direction of the antenna array's main lobe is steered into a particular given direction. In a Line-of-Sight (LoS) environment, this direction is usually the direction of the mobile terminal to be served. In a Non-Line-of-Sight (NLoS) environment, the beam is steered in the direction of the strongest scatter component.

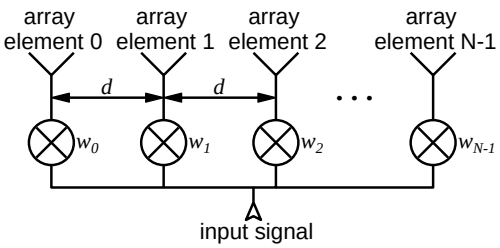



**Figure 2.9:** Block diagram of phased array antenna with complex weighting factors $w_k$.

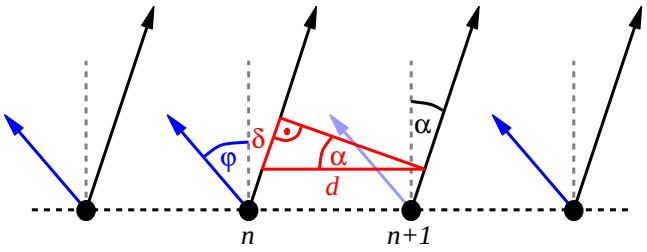



**Figure 2.10:** Calculation of radiation pattern in arbitrary direction φ with steering angle α.

For a linear phased array with high mutual antenna correlation, the calculation of the weighting factors $w_k$ can be done as follows. Let α be the angle towards which the array shall be directed, and φ be the angle towards which the resulting antenna gain shall be calculated. This means that constructive interference shall occur in the direction of α. Figure 2.10 illustrates the considered antenna array with an element spacing of $d$. Note that the sketch implies that the mobile terminal is located infinitely far away from the antenna array, since the rays from the array elements are assumed to be parallel. This is a valid and commonly made approximation, as the distance to the mobile terminal is much longer than the spacing of the antenna array.

In order to achieve constructive interference in the direction of α, the radio waves from array element $n+1$ need to be delayed by the time the waves need for the additional distance δ. Consequently, the multiplication with the weighting factor $w_{n+1}$ needs to perform a phase shift by a fraction of $\delta/\lambda$, where λ is the wavelength of the radio waves. The complex weighting factors then yield to:

$$w_k = w_{taper,k} \cdot e^{-jk\frac{2\pi d}{\lambda}\sin\alpha} \quad . \tag{2.12}$$

The $w_{taper,k}$ are additional constant weighting factors that are used for tapering the amplitudes and phases of the array elements. Tapering is used to reduce the side lobes of the array at the cost of an increased beam width and a reduced array gain. Commonly, amplitude tapering is used, which chooses the weight vectors such that the amplitude of the transmitted signal decreases towards the array edge.

In direction φ, the signal of array element $n+1$ arrives at the mobile terminal later than that of element $n$, corresponding to a phase shift $v_k$:

$$v_k = e^{jk\frac{2\pi d}{\lambda}\sin\varphi} \quad . \tag{2.13}$$

Combining both equations, the power gain $G_{array}(\alpha,\varphi)$ of the phased array antenna in direction φ when it is steered towards α can be calculated as follows (in the logarithmic domain):

$$\begin{aligned} G_{array}(\alpha,\varphi) &= G_{element}(\varphi) + 20\log \sum_{k=0}^{N_{ant}-1} w_k v_k \\ G_{array}(\alpha,\varphi) &= G_{element}(\varphi) + 20\log \sum_{k=0}^{N_{ant}-1} w_{taper,k} e^{jk\frac{2\pi d}{\lambda}(\sin\varphi - \sin\alpha)} \quad , \end{aligned} \tag{2.14}$$

where $G_{element}(\varphi)$ is the power gain of an individual array element in direction $\varphi$. Note that the beamforming effect is based on a superposition of the electric fields, hence the factor 20 when calculating the power gain in the logarithmic domain. It should also be noted that the above calculations are only approximate, since in reality there are complex interdependencies among the antenna elements. Furthermore, if the phased array has a low mutual antenna correlation, optimally chosen complex weights $w_k$ will not only modify the phase of the signals, but also the amplitude [Hot03, Dah07].

- **Null steering**

  Null steering can be used to minimize the received interference from particular directions by placing nulls in the radiation pattern. This allows to mask out interfering signals received from particular directions, or to avoid interference in certain directions when transmitting. Null steering can be combined with beam steering, and an optimal placement of nulls can be achieved with optimal beamformers.

- **Optimal beamformers**

  In contrast to beam and null steering, which both adjust the radiation pattern based on geometric assumptions, optimal beamformers choose the weights $w_k$ based on certain optimality criteria. This results in a radiation pattern where the main beam direction is not necessarily directed towards the user of interest, neither need the nulls be in the direction of the strongest interferers. The beamformer would rather place the main direction and the nulls such that for example the output SINR is maximized. Further details on this class of beamformers can be found in [God04].

Beamforming techniques are usually applied at the base stations. In contrast to antennas of mobile devices, base station antennas are stationary and usually located above the roof top, which is of great advantage. Practical issues in the design of beamforming systems are the scatter environment and the *Direction of Arrival (DoA)* estimation. In a LoS environment with only a small multipath component, it is obvious where to direct the beam of a phased array antenna. If there are several strong multipath components at the base station, this is much more difficult. Those particularly bad scatter environments can often be avoided by placing the antenna array above the roof top and by avoiding strong scatterers in the close neighborhood of the antenna. It then remains the problem of estimating the DoA of the strongest signal component, a problem known as DoA estimation. Various algorithms for DoA estimation have been developed. An overview can for example be found in [God96] and [God97]. The accuracy thereby depends, among other things, on the angular spread of the arriving signal, and on its Signal-to-Noise Ratio (SNR). Measurements by Kuchar et al. with a single arriving plane wave have shown that the DoA estimation error can be kept below $1^\circ$ if the SNR is above 0 dB [Kuc99a]. Non-zero angular spreads were considered by the same authors in [Kuc99b], revealing decreasing accuracy with increasing angular spread. If the mobile terminal is moving, further inaccuracies will occur due to an inherent delay involved in the beam steering process.

## 2.8 IEEE 802.16e Networks

The focus of this monograph is interference coordination in OFDMA networks. All discussed algorithms are generally applicable to any OFDMA or FDMA/TDMA based network that pro-

vides the necessary physical and MAC-layer mechanisms. They can therefore be developed using an abstract view of the system. However, the performance evaluation requires settling on a specific radio technology. Two OFDMA-based cellular systems have recently been developed, namely IEEE 802.16 and 3GPP LTE, where the latter is still in the final stage of the standardization process. The performance evaluation in Chapter 7 is therefore based on an IEEE 802.16e system.

The following sub-sections briefly outline the IEEE 802.16 family of standards and system specifications. Further details can be found in the respective standard documents, or in recently published books [And07, Nua07]. First, Sections 2.8.1 and 2.8.2 give an overview of the IEEE 802.16 family of standards and the WiMAX Forum Mobile System Profile, respectively. Subsequently, Sections 2.8.3 and 2.8.4 outline the physical and MAC layer of IEEE 802.16e. Finally, Sections 2.8.6 and 2.8.7 briefly discuss the architecture of mobile IEEE 802.16e networks and the interworking with 3GPP networks.

### 2.8.1 Overview of the IEEE 802.16 family of standards

The timeline of the IEEE 802.16 standard family is shown in Fig. 2.11. Work on IEEE 802.16 began in 1998 with the first standard published in 2001. The IEEE Standard 802.16-2001 specified a system for fixed wireless broadband access in the $10-66$ GHz band, which required direct LoS. The standard was later amended by IEEE 802.16a-2003, specifying the licensed operation in the $2-11$ GHz band. This gave rise to operation in NLoS environments, but still prohibited a fully mobile usage. Together with the IEEE 802.16c-2002 amendment on system profiles for the $10-66$ GHz band, the documents were consolidated in the standard IEEE 802.16d-2003. Eventually, 802.16d was superseded by the IEEE 802.16-2004 standard [IEEE80216], which is still the active standard for fixed operation. The wish for mobile operation led to the amendment IEEE 802.16e-2005 [IEEE80216e], which extends the standard by all features necessary for mobile operation. Moreover, it introduces many enhancements, such as improved support for MIMO and AAS. The further development towards a higher speed air interface is currently done within the IEEE 802.16 Task Group m (TGm).

IEEE 802.16-2004 and IEEE 802.16e-2005 are sometimes misleadingly referred to as *Fixed WiMAX* and *Mobile WiMAX*, which will be clarified in the subsequent Section. This terminology

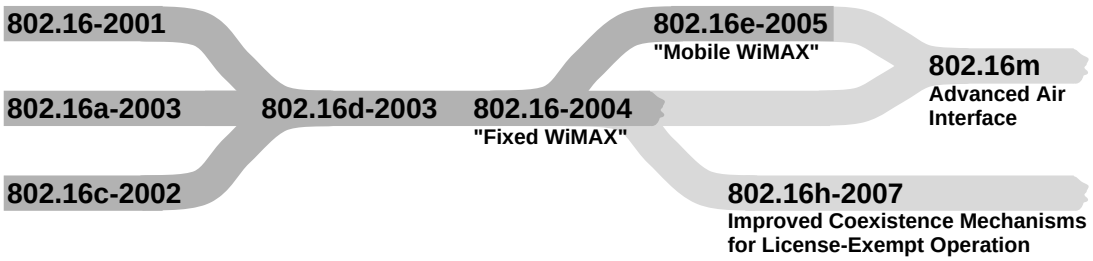



**Figure 2.11:** Timeline of most relevant IEEE 802.16 standard documents and amendments.

results from their major use cases. The application of IEEE 802.16-2004 lies in broadband wireless access for fixed subscribers, for example as a substitute for Digital Subscriber Line (DSL) access in remote areas. Moreover, it is used as a backhaul technology for hotspots and cellular networks. In contrast, IEEE 802.16e is seen as a competition for 3GPP HSPA, targeting nomadic and fully mobile terminals. Especially a nomadic usage with small portable devices, such as smart phones or notebooks, is believed to be a major use case.

Besides the air interface related standards and amendments mentioned above, several task groups have worked out specifications for management related procedures. Moreover, the Task Group h is currently preparing the amendment IEEE 802.16h [IEEE80216h], which specifies coexistence mechanisms for license-exempt network operation. The mechanisms described therein are highly related to interference coordination, as will be discussed in Section 4.4.1. They are particularly useful with license-exempt operation, where networks are deployed in a random, dynamic, and uncoordinated fashion. A summary of all past and present standards, drafts, and amendments is given by Table 2.1.

### 2.8.2 WiMAX Forum Mobile System Profile

The IEEE 802.16 standard has a large number of configuration options. Besides standard parameters like timer values, this comprises fundamental system properties like modulation and coding schemes. The WiMAX forum has therefore defined a system profile for operation and deployment of mobile IEEE 802.16e networks [WiM07a]. This profile fixes many of the various choices available in IEEE 802.16e in order to simplify product development and ensure the compatibility of products from different vendors. While IEEE 802.16 networks are often referred to as WiMAX networks, this name should actually only be applied to networks that conform with the WiMAX Forum Mobile System Profile. Nevertheless, fixed WiMAX and mobile WiMAX are sometimes used as synonyms for the respective standard documents.

### 2.8.3 IEEE 802.16e protocol architecture

The scope of the IEEE 802.16 standard family comprises the physical and the MAC layer. Hence, it specifies all algorithms for data transmission at the physical layer, as well as all relevant mechanisms at the MAC layer. A reference protocol architecture is specified by [IEEE80216] and depicted in Fig. 2.12. The left part of Fig. 2.12 shows the layering in the data and control plane, including the names of the data units exchanged between the different protocol layers in the data plane. Both are specified by the main standard [IEEE80216] and its amendment [IEEE80216e]. Apart from that, Fig. 2.12 also includes the relation to management entities. Management plane procedures are detailed in the standard amendment IEEE 802.16g [IEEE80216g] and will not be discussed here.

As illustrated in Fig. 2.12, the MAC layer is subdivided into three sublayers, namely the *Service-Specific Convergence Sublayer* (CS), the *MAC Common Part Sublayer* (MAC CPS), and the *Security Sublayer*. The resulting four protocol (sub-)layers will briefly be characterized in the following.

#### *2.8.3.1 Service-Specific Convergence Sublayer (CS)*

The CS acts as an interface to higher layer protocols. It is currently specified to handle Asynchronous Transfer Mode (ATM), IPv4, and IPv6. All incoming Service Data units (SDUs) are classified and mapped to their respective connection, which is identified by a Connection Identifier (CID). Additionally, the CS may perform header compression according to the Payload Header Suppression (PHS) algorithm specified in the IEEE 802.16 standard.

| Standard | Name | Latest release | Comment |
|---|---|---|---|
| terminated projects | | | |
| 802.16d | Detailed System Profiles for 2 – 11 GHz | not released | Amendment to 802.16-2001 |
| superseded standards | | | |
| 802.16-2001 | Air Interface for Fixed Broadband Wireless Access Systems | April 2002 | |
| 802.16a-2003 | Medium Access Control Modifications and Additional Physical Layer Specifications for 2 – 11 GHz | April 2003 | Amendment to 802.16-2001 |
| 802.16c-2002 | Detailed System Profiles for 10 – 66 GHz | January 2003 | Amendment to 802.16-2001 |
| active standards | | | |
| 802.16-2004 | Air Interface for Fixed Broadband Wireless Access Systems | October 2004 | |
| 802.16e-2005 | Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands | February 2006 | Amendment to 802.16-2004 |
| 802.16f-2005 | Management Information Base | December 2005 | Amendment to 802.16-2004 |
| 802.16g-2005 | Management Plane Procedures and Services | awaiting sponsor ballot | Amendment to 802.16-2004 |
| drafts under development | | | |
| 802.16h | Improved Coexistence Mechanisms for License-Exempt Operation | under development | |
| 802.16i | Management Plane Procedures and Services | under development | Amendment to 802.16-2004 |
| 802.16j | Media Access Control (MAC) Bridges – Multihop Relay Specification | under development | Amendment to 802.16-2004 |
| projects in pre-draft stage | | | |
| 802.16m | Advanced Air Interface | under preparation | |

**Table 2.1:** List of all IEEE 802.16 standards, standard amendments, and drafts.

#### 2.8.3.2 MAC Common Part Sublayer (CPS)

The classical MAC functionality is implemented in the MAC CPS. Most prominent, this is fragmentation and concatenation of SDUs, ARQ and HARQ, bandwidth allocation, resource allocation, and scheduling. Moreover, connection establishment and teardown, frame building (see Section 2.8.4), and advanced concepts such as interference coordination or Active Queue Management (AQM) can be implemented here.

#### 2.8.3.3 Security Sublayer

The security sublayer provides basic security mechanisms, such as authentication, encryption, key exchange, and integrity control. As the security functionality is not relevant for the discussion of interference coordination and its performance analysis, it will not be considered further.

#### 2.8.3.4 Physical Layer

The IEEE 802.16e standard defines three basic *phy modes*, namely a single carrier mode, an OFDM mode, and a mode based on OFDMA. The single carrier mode is further subdivided into a mode for $10-66$ GHz LoS operation (denoted as SC), and a mode for NLoS operation below 11 GHz (denoted as SCa). All modes can be used with FDD and TDD according to the standard, while the WiMAX Mobile System Profile [WiM07a] is currently only defined for TDD. Alike, OFDMA is the sole phy mode considered by the WiMAX Forum Mobile System Profile.

OFDMA as specified by IEEE 802.16e is also known as *Scalable OFDMA (SOFDMA)*. The term scalable thereby refers to the system bandwidth, which can be scaled from 1.25 MHz up to 20 Mhz. Note that the WiMAX Mobile System Profile specifies systems only up to a bandwidth

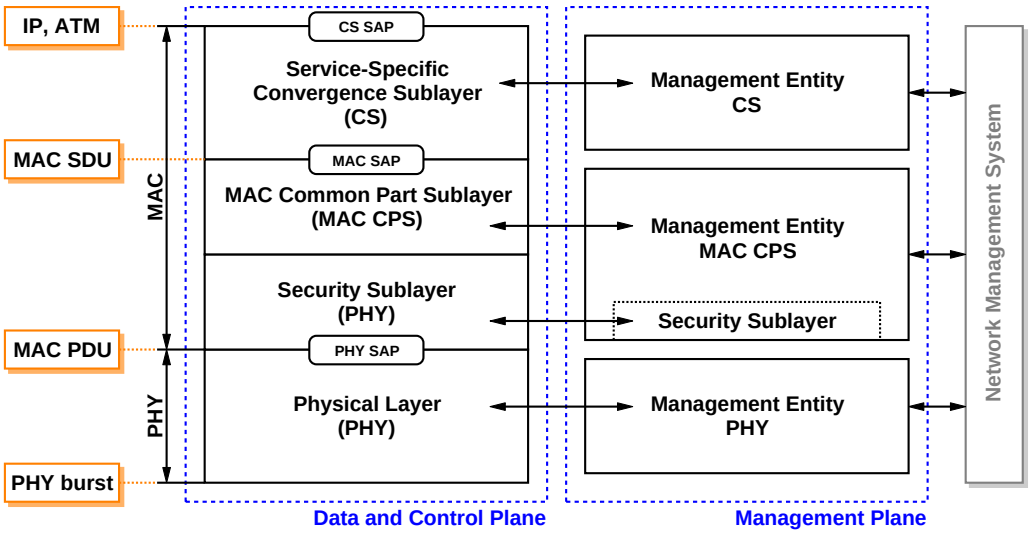



**Figure 2.12:** IEEE 802.16 reference protocol architecture according to [IEEE80216], Fig. 1.

of 10 MHz, corresponding to an FFT size of 1024. The subcarriers of the underlying OFDM transmission scheme are grouped into *subchannels*. This can be done such that the subcarriers of a subchannel are spread across the frequency spectrum in a pseudo-random fashion, or it can be done by allocating contiguous subcarriers. These strategies aim at an exploitation of frequency-diversity and frequency-selectivity, respectively, as discussed in Section 2.5.2. The basic resource assignment granularity for data transmissions is a *slot*, which is defined by a subchannel index over a certain number of OFDM symbols. A slot always contains 48 data subcarriers plus additional pilot symbols, which determines its dimensions in time and frequency direction. Several slots are grouped to form a *data burst*, within which one ore more connections of a particular mobile terminal (also referred to as subscriber station, SS) is served.

IEEE 802.16 supports AMC with Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 16-QAM, and 64-QAM. The most popular FEC schemes, which are also mandated by the WiMAX Forum Mobile System Profile, are Convolutional Coding (CC) and Convolutional Turbo Coding (CTC) with code rates ranging from $r_c = 1/2$ to $7/8$. The combination of modulation and coding scheme that is used for a particular burst is defined by its *burst profile*. The process of finding the most efficient burst profile for any transmission is called *burst profile management*. A basic burst profile management algorithm and further discussions of this issue are given in Section 6.7.

### 2.8.4 Frame structure

In IEEE 802.16e TDD systems, every MAC-frame is divided into an uplink and a downlink subframe. Both subframes are further divided into permutation zones allowing for different operational modes. A sample frame structure is shown in Fig. 2.13. The first OFDM symbol of a frame forms the preamble, which is a series of known symbols, and which is used for synchronization and equalization. Following the preamble, every frame begins with a mandatory PUSC zone (Partial Usage of Subchannels). Besides possible data bursts, this PUSC-zone contains important control information. First, the Frame Control Header (FCH) provides information on the length and the modulation and coding scheme of the following uplink and downlink maps. These maps are the second important control information in the header, as they describe the slot occupation and burst profile of the data bursts. Commonly, one or more optional PUSC-zones with data bursts follow, which may contain additional maps. PUSC is one of the basic zone types operating in a frequency diverse mode. The subcarriers of one subchannel are spread across the frequency band by a particular permutation scheme. This permutation can be parametrized by setting the permutation base parameter. If two base stations operate with the same permutation base, the subcarriers of a subchannel will all overlap in both cells. Otherwise, the subcarriers will overlap only partially, giving way to more diversity. Besides PUSC, the FUSC zone (Full Usage of Subchannels) provides frequency diversity based on a different permutation scheme. FUSC is designed to operate with a frequency reuse of $r = 1$, while PUSC also allows higher frequency reuse factors of $r = 2$ and $r = 3$. Moreover, FUSC cannot be used in combination with beamforming antennas. Hence, it will not be considered further.

A third important zone type besides PUSC and FUSC is the band AMC zone. In this zone, a set of contiguous subcarriers in frequency direction and OFDM symbols in time direction are allocated to one mobile terminal, i.e., a frequency selective resource allocation is done. The

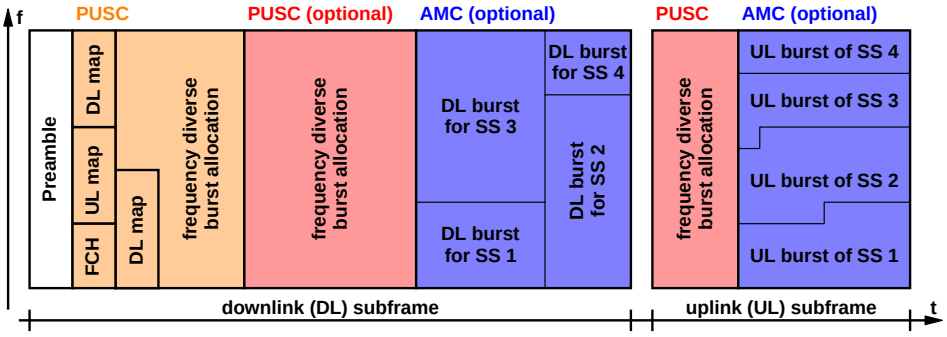



**Figure 2.13:** Schematic view of a TDD MAC frame.

subchannel size for the band AMC zone can be configured to either 8, 16, or 24 data subcarriers, but only the AMC 2x3 mode with 16 subcarriers is mandated by the WiMAX Forum Mobile System Profile. This yields slots of 16 data subcarriers by 3 OFDM symbols. In the downlink, the standard requires the allocation of rectangular areas, as indicated in Fig. 2.13. In the uplink, a linear allocation of slots is foreseen. Similar allocation restrictions exist for the PUSC and FUSC zone, albeit with a less obvious geometric analogy. The placement of bursts within the permutation zone is an NP complete optimization problem (see Section 3.6.1) and can be treated as a strip packing problem if no frequency selective scheduling is considered [Nec08c].

### 2.8.5 Beamforming support

IEEE 802.16e provides AAS support for both beamforming and MIMO. For beamforming systems, it is important that the receiver is still able to estimate the channel transfer function. Commonly, broadcast pilot symbols are used, and the mobile terminal uses all the pilot symbols for channel estimation. Beamforming requires dedicated pilots, where every mobile terminal receives pilots specifically beamformed for itself. In the downlink, PUSC and band AMC both support dedicated pilots. Besides, fast channel quality feedback is supported in order to control the quality of downlink beamforming.

### 2.8.6 Network architecture

The standards IEEE 802.16-2004 and IEEE 802.16e cover the physical and the MAC layer. The architecture of IEEE 802.16-based networks is outside the scope of these standards. Architectural aspects have therefore been addressed by the Network Working Group (NWG) within the WiMAX Forum as well as by the working group *IP over IEEE 802.16 Networks (16ng)* of the Internet Engineering Task Force (IETF).

The 16ng working group focuses on the transmission of IP traffic in IEEE 802.16 networks. This can be done either directly via the IEEE 802.16 convergence sublayer (described in [Mad08] for IPv4 and in [RFC5121] for IPv6), or with an intermediary Ethernet layer (described in [Jeo08]). Moreover, [RFC4968] studies different link models for the transmission of IPv6 traffic, such as a shared link for all mobile terminals or individual point-to-point links. All

these documents treat very specific IP and Ethernet related issues, such as the Maximum Transmission Unit (MTU) or addressing issues. They do not address issues related to the network architecture in an operator-driven environment. This task has been taken on by the WiMAX Forum, which adopted the three-stage design process according to ITU-T Recommendation I.130 [ITU-T I.130]. Stage 1 is a requirements document, stage 2 describes the overall architecture on a functional level [WiM08a], and implementation specific details such as protocols are specified in stage 3 [WiM08c].

One of the requirements addressed by the NWG is the logical separation between the Network Access Provider (NAP), the Network Service Provider (NSP), and the Application Service Provider (ASP). The NAP is the operator of the actual access network, while the NSP closes the contract with the end-user and provides the IP service. Furthermore, ASPs may provide specific IP-based services. The logical view of a typical WiMAX network architecture is shown in Fig. 2.14. The mobile terminals connect to a base station via the air interface defined by [IEEE80216] and [IEEE80216e]. The base stations are connected to an Access Service Network Gateway (ASN-GW). Together they form the Access Service Network (ASN), which is owned and operated by the NAP. Connection to the Internet or to an ASP is established via the Connectivity Service Network (CSN). Note that the functional entities within the ASN may or may not be implemented within the same device. The WiMAX architectural description specifies three different ASN profiles, allowing for a single box implementation or for implementations with separate devices.

Figure 2.14 also shows the protocol stack between a mobile terminal and its base station for the case of an all-IP network. Analog to the IETF 16ng documents, IP transport can be done either directly via the MAC CS, or with an intermediary Ethernet layer. In both alternatives, the IP or Ethernet protocol is tunneled using Generic Routing Encapsulation (GRE) from the base station to the ASN-GW, from which there is direct IP connectivity to the CSN.

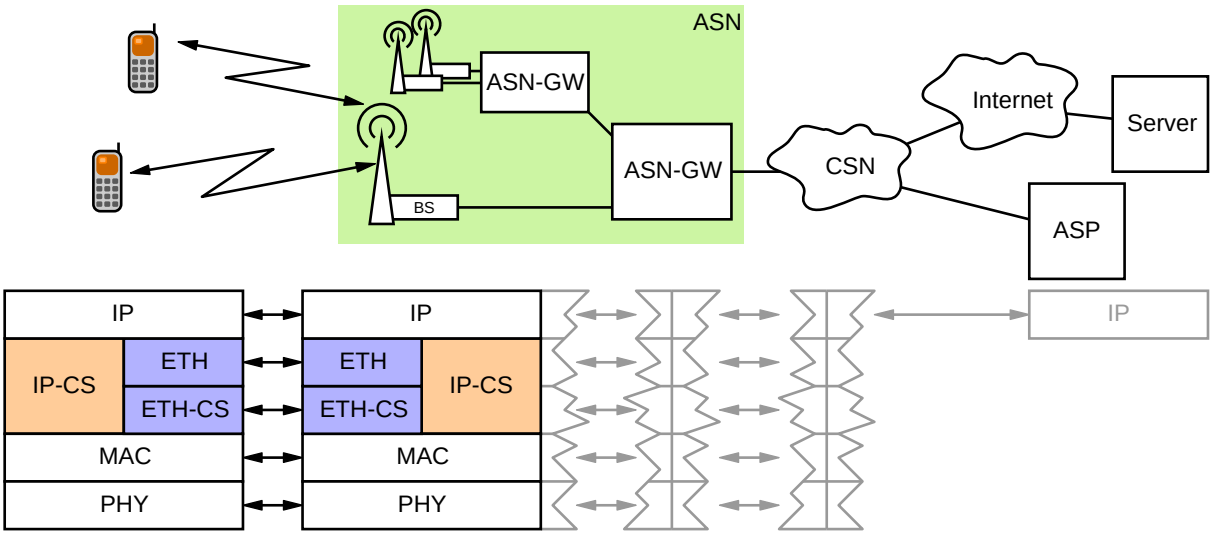



**Figure 2.14:** Logical view of the architecture of a typical WiMAX network, including the protocol stack between mobile terminal and base station.

The ASN performs all functions related to the direct connectivity with the mobile terminal. This includes mobility-related functions, such as location management and handover, Radio Resource Management (RRM) and provisioning of a required Quality of Service (QoS). The CSN performs higher layer functions. It allocates IP addresses to the mobile terminals, is responsible for Authentication, Authorization, and Accounting (AAA), and for mobility management between different ASNs.

### 2.8.7 Interworking with 3GPP networks

3GPP-based networks have a very high market penetration. The interworking of alternative radio technologies with 3GPP networks has therefore been a topic of great interest in recent years. Most importantly, the integration of Wireless LAN (WLAN) according to IEEE 802.11 gained a lot of attention, resulting in the standardization of Interworking WLAN (I-WLAN) [TS 23.234]. Based on this specification, the WiMAX forum has defined an initial concept for the interworking of IEEE 802.16 and 3GPP networks [WiM08b]. Concepts that are more sophisticated are currently under development in the course of the work on the 3GPP Service Architecture Evolution (SAE).

# 3 Methods for Simulation, Evaluation, and Optimization of Cellular Networks

The high complexity of state-of-the-art and future cellular mobile networks makes their optimization and performance evaluation a challenging task. While some aspects in cellular networks may be evaluated analytically, testbeds or computer simulations are often the only way to go. Testbeds naturally deliver the most reliable results. However, they are difficult and expensive to install and maintain. Moreover, measurement trials are very time consuming, and it is very difficult to explore a large parameter space. Computer simulations are therefore a commonly applied method when it comes to evaluating the performance of cellular networks. All relevant standardization bodies and industry consortia have specified reference scenarios, model parametrizations, and metrics in order to allow a comparison of simulation results generated by different research groups. These specifications can be found in [TS 25.814] for the 3GPP, in [WiM07b] for the WiMAX forum, and in [NGMN2007] for the Next Generation Mobile Networks (NGMN) Alliance. The latter is an open alliance of leading network operators with the goal of influencing and pushing the development of next generation mobile networks.

The first half of this Chapter is devoted to the most important methodological basics for the simulation of cellular networks. Section 3.1 classifies computer simulations of wireless and cellular networks with respect to their level of abstraction, and it introduces basic simulator categories. Section 3.2 treats models for path loss and fading processes of wireless channels, followed by Section 3.3 on the modeling of cellular scenarios. Further aspects have to be considered when taking into account higher layer protocols, which will be discussed in Section 3.4. Finally, Section 3.5 introduces and classifies the most important metrics that are used during the performance evaluation of cellular networks.

The optimization of cellular networks can be done on many different levels and for a large number of different aspects. This includes static optimizations during the network planning process, as well as highly dynamic optimizations such as optimized queue management or scheduling decisions at the frame level. Many optimization problems that occur can be solved by means of simple problem specific heuristics. However, certain problems demand for a more generic solution approach. This is the topic of the second half of this Chapter in Section 3.6. In particular, Section 3.6.1 gives a brief introduction to optimization problems in general. Subsequently, Section 3.6.2 overviews the field of evolutionary algorithms. Finally, Section 3.6.3 introduces genetic algorithms, which can be used as a generic tool to solve a large number of combinatorial optimization problems.

## 3.1 Simulation of Wireless and Cellular Networks

Computer simulations of cellular networks can be classified with respect to their level of abstraction and with respect to their simulation type. This will be described in Sections 3.1.1 and 3.1.2, respectively. Subsequently, Section 3.1.3 introduces the concept of Variance-Reduction Techniques, which are a powerful tool for treating the problem of long simulation times caused by complex models of cellular networks.

### 3.1.1 Level of simulation

Simulations in wireless and cellular networks can be classified in three categories with respect to the targeted layers in the protocol stack and with respect to the considered level of abstraction.

- **Physical / link level simulations**

  Physical level simulations, also known as link level simulations, are used for optimization and evaluation of physical layer properties and algorithms, such as symbol constellations or error correction codes. This requires a detailed modeling of the wireless link and a simulation at the symbol level. Their time scale of operation is therefore the baud rate of the system. Results are often Bit Error Ratios (BER) or Block Error Ratios (BLER), where the BLER may for example refer to FEC or ARQ blocks. Due to the involved computational complexity, it is difficult to extend link level simulations to higher protocol layers, whose effects are on longer time scales.

- **Network level simulations**

  In order to simulate protocol layers above the physical layer, network level simulations usually use an abstract physical layer model. They often comprise the MAC layer and the IP layer, but may also extend to transport layer protocols or even applications. Subject to investigation may be ARQ performance, scheduling, queue management, protocol interactions, and the like. The physical layer abstraction makes necessary an appropriate interface, which will be discussed in Section 3.4.1. Effects from other cells are often neglected or represented by statistical models.

- **System level simulations**

  While physical and link level simulations commonly target only one or a few transmission links within a cell, system level simulations cover several cell sectors and transceivers within a certain geographic area. Classically, they use a simple model of the physical layer and abstract the effect of MAC protocols. System level simulations are used to evaluate system level aspects such as network capacity and coverage, handover algorithms, admission control, and the like.

The choice of the right simulator category depends on the investigated problem and the desired results. Usually, it is advantageous to decouple simulators working at different levels of abstraction in order to reduce complexity. In this case, appropriate interfaces are required, a problem

which will be treated in Section 3.4.1 for the interface between physical and network level simulations. However, some problems may require the combination of different simulation levels in one simulator. Recently, network and system level simulations have grown together, owing to the increasing complexity of MAC protocols and algorithms, which more and more interact with adjacent cells, or depend on adjacent cells in their performance. Examples include interference coordination algorithms or certain aspects of scheduling algorithms.

### 3.1.2 Simulation type

Mattern [Mat96] distinguishes three basic simulation paradigms, namely the *event-driven simulation*, the *time-driven simulation*, and the *stochastic simulation*:

- **Event-driven simulation**

  In an event-driven simulation, the system state changes only at certain time points called events. It must therefore be possible to associate every state change with one event. Events take place in zero time. Since per definition nothing happens between two events, the simulation time can directly advance to the time of the next event once the current event has been processed. Event-driven simulations are commonly used for physical and network level simulations. They are also applied to system level simulations when the progress of time plays a role. System state changes are the transmission instants of new data symbols, the generation of a new packet or the arrival of a new call, the expiry of a retransmission timer, etc.

- **Time-driven simulation**

  Time driven simulations can be applied when it is not possible to associate every system state change with a corresponding event. Instead, the simulation time progresses in predefined constant increments. After every increment, all system state changes that occurred since the last increment are aggregated and committed. Time-driven simulations are for example applied for climate simulations, where it would be impossible to identify events. They are commonly not used for the simulation of communication networks.

- **Stochastic simulation (Monte Carlo simulation)**

  Monte Carlo simulations are characterized by the fact that the progress of time plays no substantive role [Law91]. Instead, random numbers are used as triggers and inputs to stochastic models. The final result is obtained by averaging over all results obtained from the generated random triggers and inputs, where all results must be obtained from uncorrelated sets of input values. In the context of mobile networks, every step that generates an output from a random input is referred to as *drop*. In this area, the Monte Carlo method is used for certain types of system level simulations where the progress of time is not important. One example is the evaluation of the network coverage. In this case, mobile terminals are randomly placed in the scenario in every drop (the mobiles are *dropped* onto the cells), and the resulting signal and interference situation is determined. This results in a particular coverage and throughput result for that drop. By performing a large number of drops, it is possible to obtain for example the average throughput at a particular position.

The above described paradigms can be combined to hybrid approaches. A popular example is the combination of event-driven simulation with Monte Carlo simulation. In such an approach, an event-driven simulation is performed for every drop. For example, the mobile terminals could be repositioned for every drop, and a short event-driven simulation could be executed to obtain certain performance metrics including time-dependent processes such as ARQ.

### 3.1.3 Variance-Reduction Techniques

The computational complexity of a cellular network simulation depends on many factors, such as the channel models and the size of the scenario (i.e., the number of cells and the number of mobile terminals). Moreover, the layers that are spanned in the protocol stack have a strong impact on the simulation time. Large-scale simulations of cellular networks may quickly entail long computation times. Increasing the simulation's efficiency is therefore of great interest. *Variance-Reduction Techniques* (VRT) can be used to increase the *statistical efficiency*, which is measured by the variances of a simulation's output random variables [Law91]. Several VRTs haven been proposed in literature. In this Section, the technique of *Common Random Numbers* (CRN) is presented, which can be applied to cellular network simulations with little extra effort. Further information on VRTs can be found in [Law91].

CRN can be used when comparing different configurations of a system in different simulation runs. The basic idea behind CRN is to evaluate different system configurations under the same statistical conditions. Variations in the results should result from differences in the system configuration, not from differences in the statistical input data. CRN solves this problem by using the same set of random number sequences in every simulation run. It is important to take care that a particular random number is used for the same "purpose" in every simulation run. For example, the same sequence of random numbers should be used for the inter-arrival time of arriving packets in a queuing network. This can often be solved by using a set of independent Random Number Generators (RNGs), each driving a certain set of input random variables. When using CRN in the context of cellular networks, different RNGs have to be used for the mobility model (Section 3.3.2), the channel models (Section 3.2), the traffic generators, and possibly other parts of the simulation model. Using a separate RNG for the mobility model forces all mobile terminals to move along the same paths in different simulation runs. Using a separate RNG for the channel models additionally forces the channels towards the mobile terminals to behave identically in each simulation run. Thus, every system configuration has to handle the same difficult or beneficial constellations of mobile terminals at each time point in the simulation. Eventually, this avoids that the results of some simulation runs are influenced by particularly good or bad terminal constellations, allowing for shorter simulation times when comparing for example different algorithms in the network.

## 3.2 Modeling of Wireless Channels

Wireless propagation models can be classified into two categories. *Deterministic models* usually make use of raytracing or related techniques [Lan99]. They are based on a more or less precise topographical representation of a certain area, which includes buildings and elevation information. The path loss is calculated by considering the physical paths along which the

electromagnetic waves propagate, taking into account reflections, diffractions, and scattering. Deterministic models therefore deliver a very good map of the signal strength in a specific terrain, such as a particular city, which comes very close to reality. Usually, deterministic models include the effects of shadowing, but not the effects of fast fading. They are computationally very expensive and require a detailed database of the environment. They are commonly used for system level simulations and radio network planning, where exact results for a particular area are most important.

*Empirical models* are statistical models, which are generally based on results from measurement campaigns. In contrast to deterministic models, they do not take into account terrain features of a particular area. Empirical models have a much lower computational complexity compared to deterministic models and can be used for link, network, and system level simulations. The remainder of this Section details some of the most important empirical models for the distance-dependent path loss (Section 3.2.1), the effects of shadowing (Section 3.2.2), and for fast fading (Section 3.2.3).

### 3.2.1 Empirical path loss models

Various empirical path loss models have been developed, usually by curve fitting of data obtained by measurements. A measurement-based approach implies that the model is valid for a certain parameter range, namely the range that the measurement data sets cover. The Okumura-Hata model is one popular empirical path loss model. In [Hat80], Hata presented the following empirical formula for the path loss calculation:

$$L_{p\ (\mathrm{dB})} = A + B \log_{10} d \ . \tag{3.1}$$

Hata obtained the parameters $A$ and $B$ from measurements performed by Okumura et al. in [Oku68]. Influencing factors are the terrain category, the carrier frequency, and the antenna heights. The Okumura-Hata model can be used to model the path loss in macro cells with path lengths between 1 and 20 km and frequencies between 100 and 1500 MHz, though extrapolations to higher frequencies are possible. Mogensen et al. [Mog91] proposed a modified Hata prediction model for the frequency range $1500-2000$ MHz based on measurements in urban areas, which was further refined in [COST231]. Since this model was developed in the course of COST action 231, it is known as COST231-Hata model. COST 231 also proposed a combination of the Walfisch [Wal88] and the Ikegami [Ike84] models, known as COST231-Walfisch-Ikegami model. In contrast to the models before, it is suitable for smaller cells with a path loss distance in the range of 0.02 and 5 km.

More recently, Erceg et al. [Erc99] proposed a path loss model based on measurements in the 1.9 GHz band with possible path lengths between 0.1 and 8 km. It can be parametrized to model a variety of different scenarios. In particular, the authors also account for the variation of the path loss exponent between different cells. The path loss is given as follows [Erc99]:

$$L_{p\ (\mathrm{dB})} = A + \underbrace{10(a - bh_b + \frac{c}{h_b}) \log_{10}\left(\frac{d}{d_0}\right)}_{\text{distance-dependent path loss}} + \underbrace{10x\sigma_\gamma \log_{10}\left(\frac{d}{d_0}\right)}_{\text{cell individual variation}} + \underbrace{y\mu_\sigma + yz\sigma_\sigma}_{\text{shadowing}} \ ; \ d \geq d_0 \ , \tag{3.2}$$

with

$$A = 20log_{10}(4\pi\frac{d_0}{\lambda}) \ , \tag{3.3}$$

and $d_0 = 100$ m. $a$, $b$, $c$, $\sigma_\gamma$, $\mu_\sigma$, and $\sigma_\sigma$ are model parameters which depend on the terrain category. Three terrain categories are defined in [Erc99], namely *Hilly/Moderate-to-Heavy Tree Density* (terrain category A), *Hilly/Light Tree Density or Flat/Moderate-to-Heavy Tree Density* (terrain category B), and *Flat/Light Tree Density* (terrain category C). $x$, $y$, and $z$ are independent zero-mean Gaussian variables with unit standard deviation. $x$ and $z$ vary from cell to cell, where $y$ varies from location to location. The latter models the shadowing variations. A more detailed overview of shadowing models in general is given in the following sub-section.

### 3.2.2 Empirical shadowing models

Gudmundson proposed a simple empirical model for temporally correlated shadowing in [Gud91]. He used a simple correlation function, which exponentially decreases with the traveled distance. A model verification showed that the results obtained with this model are consistent with measurements obtained from urban and sub-urban environments. The correlation of the shadowing values in the logarithmic domain can be expressed depending on the decorrelation length $d_{corr}$ as follows [TR 101 112]:

$$R(\Delta x) = e^{-\frac{\Delta x}{d_{corr}}\ln 2} \ . \tag{3.4}$$

The decorrelation length $d_{corr}$ is the distance, where the correlation has dropped to $1/e$, i.e., $R(d_{corr}) = 1/e$. If a lognormal distribution of the shadowing values is desired, an auto-regressive filter of first degree can be used for their generation as shown in Fig. 3.1. If the filter operates in the logarithmic domain, its input are independent and identically distributed (i.i.d.) random variables, which are normal distributed with mean 0 and standard-deviation $\sigma_1$. The output of the filter is also normal distributed with zero mean and standard deviation $\sigma_2$. The multiplication factor $R(\Delta x)$ determines the correlation between two subsequent samples at the filter output, where $\Delta x$ is the distance between the two locations the samples are drawn for. In order to calculate the multiplication factor $a$, a further constraint has to be defined. It is thereby practical to require $\sigma_1$ to be identical to $\sigma_2$. From Fig. 3.1, it yields:

$$\sigma_2 = \sqrt{R(\Delta x)^2\sigma_2^2 + a^2\sigma_1^2} \tag{3.5}$$

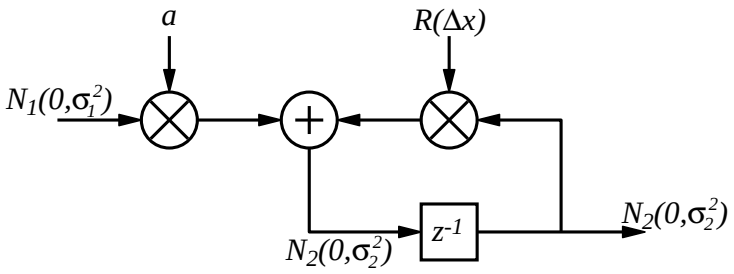



**Figure 3.1:** Auto-regressive filter of degree one for generation of correlated shadowing values.

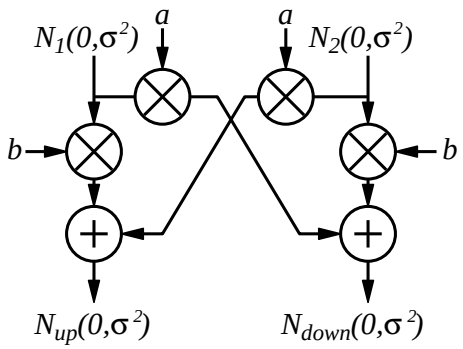



**Figure 3.2:** Generating correlated shadowing values for up- and downlink.

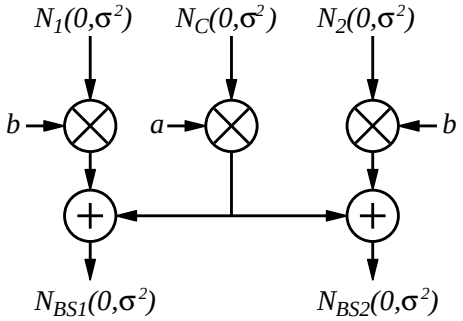



**Figure 3.3:** Generating correlated shadowing values for links to different base stations.

and with $\sigma_1 \stackrel{!}{=} \sigma_2$:

$$a = \sqrt{1 - R(\Delta x)^2} \ . \tag{3.6}$$

This form of spatial correlation does not guarantee that two mobile terminals at the same position have the same shadowing value. Even one mobile terminal that passes the same position several times will receive different shadowing values every time. Nevertheless, the Gudmundson-model is widely accepted and required by the 3GPP, the NGMN, and the WiMAX forum. If shadowing values with a location-dependent or even an angular-dependent correlation are desired, other models have to be used, such as the model proposed by Cai and Giannakis in [Cai03].

Besides the spatial correlation of one particular channel, there is a correlation between different channels of one mobile terminal [Kli99]. First, the shadowing in up- and downlink direction is correlated. In a TDD system, the shadowing is identical for up- and downlink, that is, the correlation is one. In an FDD system, the correlation depends on the frequency gap between the up- and downlink bands, but is typically in the order of 0.9. Second, the shadowing towards different base station sites is correlated with a correlation coefficient of 0.5 according to NGMN and WiMAX forum requirements. The generation of correlated random variables for this purpose is analog to the generation of temporally correlated random variables and will be described in the following.

Figure 3.2 shows a sample block diagram for the generation of correlated shadowing values in the uplink and the downlink between one mobile terminal and one base station. The shadowing in each direction is a superposition of two normal distributed random variables $N_1$ and $N_2$. The weighting factors $a$ and $b$ can be calculated by taking into account the correlation coefficient $\rho_{up,down}$ for links of one mobile terminal:

$$\rho_{up,down} = E\left(\frac{1}{\sigma^2}(ax_{1,i} + bx_{2,i})(ax_{2,i} + bx_{1,i})\right) \tag{3.7}$$

$$\rho_{up,down} = 2ab \ , \tag{3.8}$$

where $E()$ is the expected value, and $x_{1,i}$ and $x_{2,i}$ are realizations of $N_1$ and $N_2$, respectively, at time index $i$. Note that $N_1$ and $N_2$ are i.i.d. random variables. Since the standard deviation of

the input distributions $N_1$ and $N_2$ is again required to be equal to the standard deviation of the output distributions $N_{up}$ and $N_{down}$, the weighting factors $a$ and $b$ yield to:

$$a = \sqrt{\frac{1}{2}}\sqrt{1-\sqrt{1-\rho^2_{up,down}}} \tag{3.9}$$

$$b = \sqrt{\frac{1}{2}}\sqrt{1+\sqrt{1-\rho^2_{up,down}}}\ . \tag{3.10}$$

Similarly, correlated random variables for shadowing on links towards different base stations can be calculated according to Fig. 3.3. In this case, the shadowing of each link is a superposition of a common normal distributed random variable $N_C$, and individual normal distributed random variables $N_1$ and $N_2$, one for each link. The weighting factors $a$ and $b$ can be calculated by taking into account the correlation coefficient $\rho_{BS}$ for links of one mobile terminal:

$$\rho_{BS} = E\left(\frac{1}{\sigma^2}(ax_{C,i}+bx_{1,i})(ax_{C,i}+bx_{2,i})\right) \tag{3.11}$$

$$\rho_{BS} = a^2\ . \tag{3.12}$$

The weighting factors $a$ and $b$ then yield to

$$a = \sqrt{\rho_{BS}} \tag{3.13}$$

$$b = \sqrt{1-\rho_{BS}}\ . \tag{3.14}$$

### 3.2.3 Empirical fast fading models

Flat fading is the simplest form of fast fading. An appropriate fast fading model has to take into account the temporal correlation of the fading values. Two popular models for flat fading channels are *filtered Gaussian noise* and the *sum of sinusoids* method [Stü00]. The sum of sinusoids method is very intuitive, as it constructs a complex fading envelope from the superposition of several complex sinusoids with different angular Doppler frequency and phase shift.

In a frequency selective environment, the variation of the channel transfer function in frequency direction has to be taken into account. Frequency selective fast fading models are usually based on tapped delay lines, where the complex fading envelope results from a superposition of several *taps* representing signal components arriving at different times. Höher proposed a model where the complex envelope is a superposition of several complex sinusoids with different angular Doppler frequency, time shift, and phase shift [Hoe92]. The model proposed by the ITU-R in [ITU-R M.1225] constitutes of a superposition of a much smaller number of taps, where each tap undergoes Rayleigh fading [Stü00].

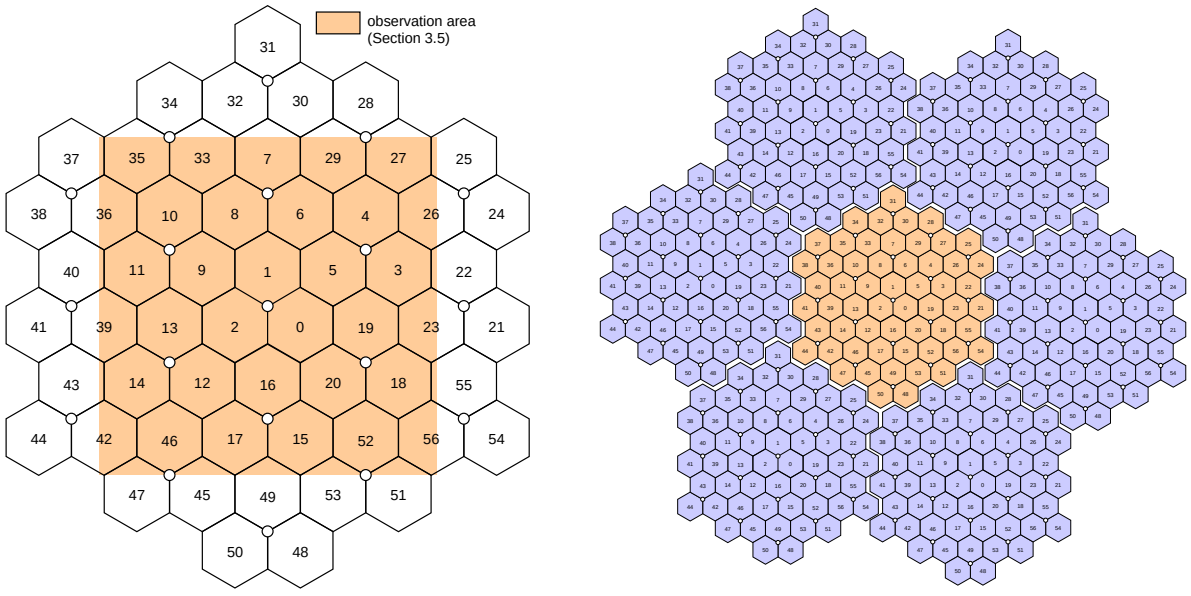
For MIMO systems, the spatial correlation has to be considered as a further component. A popular model for this purpose proposed by the 3GPP and 3GPP2 is the *Spatial Channel Model* (SCM) [TS 25.996]. SCM was developed for a carrier frequency of 2 GHz and a signal bandwidth of up to 5 MHz. The *Spatial Channel Model Extended* overcomes these limitations, with a carrier frequency range of $2-5$ GHz and a bandwidth of up to 100 MHz [Bau05], and it also adds further extensions. SCME was developed in the course of the European IST Winner project, which also developed two new models in project phase 1 [Bau05, ES06] and phase 2 [Kyö2006], respectively. A comparison of all four models can be found in [Nar07].

## 3.3 Modeling of Cellular Scenarios

Cellular scenarios can be modeled in a number of different ways. The choice of the model depends on the problem under investigation and the required level of detail. The simplest possibility is a single-cell model in which one cell is modeled in detail and effects from other cells are taken into account by abstract models. For example, other-cell interference can be modeled by traces or statistical models. There may even be problems where effects of other cells only have an insignificant impact and can therefore be neglected. If a single-cell model is not sufficient, there are two different possibilities to model a scenario with multiple base stations. First, a real environment with actual buildings, streets, etc. can be used. Base stations are placed at their real geographic coordinates, and mobile terminals move along distinct paths, such as streets. Such a model is typically used in conjunction with a deterministic propagation model. The computational complexity of this approach is relatively high. Alternatively, a more abstract model based on hexagonal cells and empirical propagation models can be used. An example is illustrated in Fig. 3.4(a), which shows a scenario with 19 base stations, each serving three cell sectors. This results in a total of $N_{cells} = 57$ cell sectors. Even though such a hexagonal model does not represent a real world scenario, it allows for an easy modeling of a multi-cellular scenario. This makes it well suitable as a standard scenario, and it has been adopted as the basic scenario in all three mentioned bodies ([TS 25.814], [NGMN2007], [WiM07b]).

### 3.3.1 Treatment of scenario borders

In the scenario of Fig. 3.4(a), the border cells receive less interference than the center cells. Hence, the results will differ in each cell depending on its position. There are several possibili-




(a) scenario without wrap-around

(b) identical layout with wrap-around

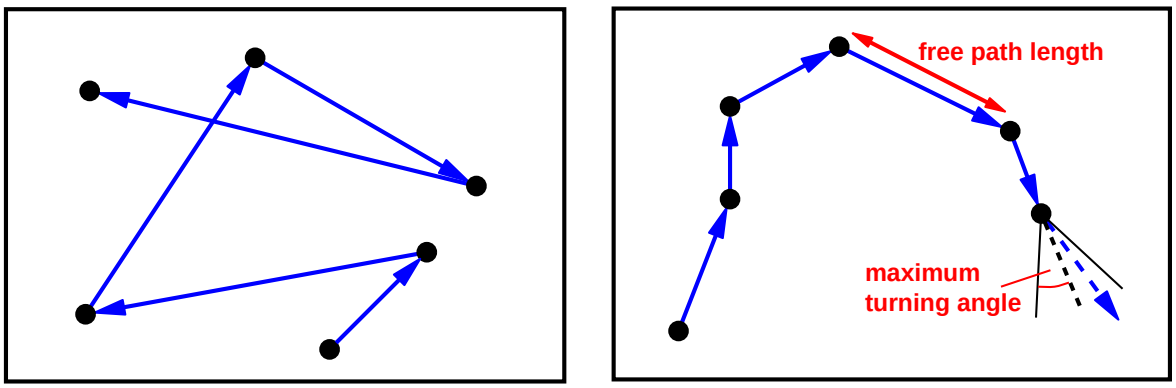
**Figure 3.4:** Cellular scenario with 19 base station sites and three sectors per base station.

ties to treat this problem. The easiest way is to evaluate only the center cells while the results of all other cells are ignored. This, however, wastes a lot of computing power, since only a fraction of the simulated cells is evaluated. A second possibility is to surround the scenario by interference generating cells, which generate interference based on a statistical model. In this case, all cells that are simulated in detail can be evaluated, even though border cells may still operate under different conditions compared to center cells. A variant of such a method was presented by Berger et al. in [Ber04b], where other-cell interference within a cell was modeled based on the cell geometry factor and the terminal's line of sight angle to its serving base station.

Last but not least, *wrap around* is an attractive technique. Wrap around virtually repeats all base stations infinitely often in a regular grid, as it is illustrated in Fig. 3.4(b). This corresponds to a mapping of the scenario in Fig. 3.4(a) onto a torus. Mobile terminals and electromagnetic waves that leave the scenario re-enter it at the opposite side, where the impact of every transceiver is taken into account only once, namely by choosing the one closest to the point of interest. Hence, cell sectors number 27 and 47 now have a common borderline and cause interference on each other. With wrap around, all cells are equal, and measurements can be taken in every cell. This makes it a very efficient and widely used model.

### 3.3.2 Terminal mobility

Simulations with moving terminals require a suitable mobility model. For a scenario based on empirical channel models with no distinct terrain features, two popular mobility models are the *random waypoint* and the *random direction* mobility model. In the random waypoint mobility model illustrated in Fig. 3.5(a), a target location within the movement plane is randomly determined. The terminal then moves to that position on a straight line and with a randomly chosen constant speed. The major drawback of this model is its non-uniform spatial terminal distribution, as terminals are more likely to be encountered in the center area [Bet02]. In the random direction mobility model, the direction of movement and the length of the path segment are randomly chosen. The terminal moves with a randomly chosen constant speed until the end of the path segment is reached. Then a new direction is randomly chosen where a maximum




(a) Random waypoint (b) Random direction

**Figure 3.5:** Illustration of commonly used terminal mobility models.

turning angle can be set, as shown in Fig. 3.5(b). The free path length is selected based on a given distribution function resulting in a particular mean free path length. Special actions have to be taken if the mobile terminal hits the scenario border. A uniform node distribution is achieved if the terminal bounces off the scenario border like a billiard ball, or if a wrap-around technique is used [Bet01].

In all models described above, a randomly chosen velocity may lead to problems, since slowly moving terminals take a much longer time to reach their destination compared to fast moving terminals. Therefore, the scenario will contain more slowly moving terminals than fast moving terminals once it has reached its steady state. If all terminals move at a constant speed, this problem naturally does not arise.

### 3.3.3 Interference and noise modeling

As discussed in Section 2.6, noise and interference are the fundamental capacity limiting factors in a wireless network. An adequate model for these effects is therefore indispensable. Thermal noise is commonly modeled as Additive White Gaussian Noise (AWGN). This means the noise directly adds to the received signal's amplitude, its power spectral density is not frequency dependent, and the noise samples are i.i.d. and follow a Gaussian distribution. While this is quite simple, interference modeling is more difficult. As discussed in Section 2.6, there exist different types of interference. Intra-cell interference essentially adds a constant interference offset and is mostly independent of the applied scheduling and resource assignment strategy, except for MAI that occurs when several terminals are scheduled at the same time in a CDMA system. In contrast, inter-cell interference largely depends on the scheduling and resource assignment strategy, and thus also on possibly applied interference coordination schemes. It therefore has a large impact on the system performance. Consequently, in the remainder of this monograph, only inter-cell interference will explicitly be considered.

The modeling of inter-cell interference depends on the air interface technology, and there mostly on how neighboring cells are separated. In CDMA systems like cdma2000 or UMTS, neighboring cells are separated by Pseudo Noise (PN) sequences, and the data signals are spread across the whole bandwidth. It is therefore valid to assume the inter-cell interference as additional AWGN. In FDMA/TDMA or OFDMA systems, this is more difficult. A commonly used approach to model inter-cell interference is to assume it as an additional source of additive Gaussian noise. The interference power is calculated using the same propagation models that are used for the desired signals. This needs to be done separately for every time and frequency slot depending on the resource allocation in the interfering cells. The interference is therefore not white. Plass et al. have evaluated the accuracy of this approach in a 3GPP LTE environment [Pla06] and found it to be a good approximation of the real interference situation. Exceptions are the application of advanced signal processing algorithms that rely on the exact interfering signal properties, such as interference cancelation or optimal combiners. Furthermore, the Gaussian approximation also does not hold if there is a highly dominating interferer. However, in such a situation, the SIR is relatively bad anyway, leading to a very low or even no throughput, and hence to a relatively low impact of this inaccuracy.

## 3.4 Modeling of Data Transmissions on Wireless Links

On top of the model of a wireless channel and a cellular network, protocols and other building blocks need to be modeled in order to be able to evaluate the performance of an end-to-end connection in a network level simulation. This involves a large number of modeling decisions. First of all, the level of abstraction needs to be determined. Usually, network level simulations extend from the MAC layer to some higher protocol layer, making necessary an interface between the physical and the MAC layer. However, other levels of abstraction are possible. A higher level was chosen in [Pei01] and [Nec05a], where the authors present models for the delay of IP packets in UMTS networks. On the one hand, such models require more simplifications. On the other hand, they allow a fast and efficient simulation of higher layer protocols. If the model reflects all delay and loss effects that have an impact on a specific higher layer protocol, it can efficiently be used for their design and evaluation [Nec05a]. It is also possible to choose the abstraction below the MAC layer. In this case, certain parts of the physical layer would explicitly be simulated. Even though this brings along a high computational complexity, it may be necessary for systems where it is difficult to find suitable abstractions for certain physical layer processes. MIMO systems are an example where the abstraction of physical layer issues becomes more and more difficult.

Nevertheless, network level simulations starting at the MAC layer are very common and can efficiently be applied to many wireless technologies. A suitable interface between the physical and the MAC layer is required. Section 3.4.1 reviews different possibilities for this interface and the modeling of block errors on the physical layer. Subsequently, Section 3.4.2 discusses the problem of HARQ modeling.

### 3.4.1 Interface between physical and MAC-layer

The interface between the physical and MAC-layer depends very much on the considered radio technology. Typically, an adequate model for the block error process on the MAC layer is developed, which is then parametrized by separate link level simulations. The model has to reflect all relevant characteristics of the block error process. This includes second-order statistics, such as the correlation of block errors. Models for block error processes can be put into one of the following three basic classes.

- **Statistical models**

  The simplest statistical model is a random drop model with constant drop probability. Such a model is well suitable for certain systems, such as CDMA systems with perfect power control [Nec05b]. Even if it does not accurately reflect the link behavior in other systems, it may be a valid simplifying assumption for certain studies. More advanced statistical models were originally developed to represent bursty bit errors in link level simulations. The first model was presented by Gilbert in [Gil60]. It is based on a Markov chain with two states, representing a *good* and a *bad* channel state, respectively. In the good channel state, no errors occur. In the bad channel state, symbol errors occur with a certain probability. This model captures the correlation and burstiness of symbol errors. Elliott [Ell63] extended the model such that symbol errors may also occur in the good

state with a certain probability, yielding the well-known Gilbert-Elliot channel model. The same approach can also be used to model the block error process on the MAC layer. For example, Wang and Moayeri [Wan95] and Bergamo et al. [Ber02] discuss finite-state models for the Rayleigh fading channel. Tralli and Zorzi developed models for the block error process within a WCDMA system based on $N$-state Markov chains that adequately capture the correlation between block errors [Tra02, Tra05].

Statistical models are easy to implement and have a low computational complexity. However, their development and parametrization is difficult. Moreover, they are restricted to relatively simple interfaces between the physical and MAC layer. Advanced features such as HARQ can be modeled only roughly, and it is time-consuming to model correlations among different connections.

- **Traces**

  In contrast to statistical models, traces constitute a direct interface between a link and a network level simulation. Typical traces obtained from link level simulations contain a block error probability for every time step. Trace-based models are easy to implement and have a low computational complexity. However, different traces must be generated for different scenarios, such as different terminal velocities, and trace files may be very large. It is also difficult to capture feedback effects from upper layers to the physical layer. In contrast to statistical models, it is easier to model advanced features such as HARQ with traces. On the other hand, traces also lack a detailed interference model, making it more difficult to model correlations among different connections.

- **Lookup tables**

  A very common approach to a link level interface is based on lookup tables, which map an SINR-value to a specific loss probability. The loss probability is usually given per FEC-block. If a data block on the MAC layer is composed of multiple FEC blocks, a separate calculation of the overall loss probability has to be done. Lookup-table based approaches require an SINR value for every FEC-block, which can either be obtained from traces, or by a direct calculation of the SINR in the simulation. Fast fading can be considered in two ways. Hämäläinen et al. [Häm97] refer to these two possibilities as *average value interface* and *actual value interface*, respectively. The average value interface was originally discussed in [Mal94] for system simulations of next generation TDMA systems. This interface type includes fast fading in the lookup tables, whose input is thus an average SINR value. In the actual value interface, instantaneous lookup tables are used, which do not average out the effects of fast fading. This requires the effects of fast fading to be included in the input SINR values.

  While [Häm97] and [Mal94] consider flat fading channels, additional steps may be necessary when considering wide band frequency selective fading channels. If the data symbols of an FEC block experience different fading conditions, an effective SINR value for the table lookup has to be calculated. This may be the case if the FEC block is spread over a large enough frequency or time range. A commonly used approach is the *Exponential Effective SINR Mapping* (EESM) [R1-031303].

  Lookup table based approaches are relatively accurate, and they allow modeling advanced features such as HARQ (see next Section 3.4.2). They have a low computational complexity, despite the possibly time consuming calculation of (effective) SINR values. On

the downside, link level simulations are necessary to obtain the lookup tables. Moreover, separate lookup tables have to be obtained for different modulation and coding schemes, for different FEC block sizes, and for different scenarios.

### 3.4.2 Modeling of hybrid ARQ mechanisms

Advanced MAC layer algorithms in wireless networks often require a close interaction of the physical and MAC layer. This contradicts a clear abstraction of the physical layer as discussed in the previous sub-section. Indeed, the modeling of certain MAC layer algorithms poses a big challenge on the design of network level simulations and may even require at least a partial simulation of the link layer. Hybrid ARQ is a prominent example, which requires a tight interaction of MAC and physical layer. In the following, the modeling of hybrid ARQ in combination with a link layer interface based on lookup tables is discussed.

Frederiksen et al. presented a model for determining a combined SINR after $n$ transmissions of a hybrid ARQ data block with chase combining or incremental redundancy [Fre02]:

$$\left(\frac{E_S}{N_0}\right)_{C,n} = \varepsilon^{n-1} \cdot \eta(\mathrm{ECR},M)_n \cdot \sum_{k=1}^{n} \left(\frac{E_S}{N_0}\right)_k \tag{3.15}$$

$$\eta(\mathrm{ECR},M)_n = \begin{cases} \eta(\mathrm{ECR},M) & \text{if } n \geq 2 \\ 1 & \text{otherwise} \end{cases}, \tag{3.16}$$

where $M$ is the order of the modulation scheme and the Effective Code Rate (ECR) is the code rate of the first transmission. $E_S$ is the received energy per data symbol, and $N_0$ is the noise power per data symbol. $(E_S/N_0)_{C,n}$ is the combined $E_S/N_0$ after $n$ transmissions. The chase combining efficiency $\varepsilon$ describes the loss compared to an ideal summation of the SINR values of all (re)transmissions. [Fre02] lists values of $\varepsilon$ between 0.92 and 0.95 depending on the channel profile and modulation scheme. $\eta(\mathrm{ECR},M)$ is the additional gain achieved by incremental redundancy. It can be set to 1 if only chase combining is used.

A simplified version for a chase combining hybrid ARQ is obtained by setting $\varepsilon$ to 1, as it was done in [Nec07e]. This overestimates the combined SINR by about $5-9\%$ after the first retransmission, by about $11-18\%$ after the second retransmission, and by approximately $17-28\%$ after the third retransmission. According to previous studies in [Nec07f], the probability for a second retransmission is about 7% in a frequency reuse 1 scenario, and approximately 2% in a reuse 3 scenario with random scheduling. The probability for a third retransmission falls to approximately 2% in a reuse 1 scenario, and to about 0.4% in a frequency reuse 3 scenario. The error introduced by choosing $\varepsilon = 1$ is therefore negligible in most cases and needs only be addressed when investigating the tails of message delay distributions.

## 3.5 Evaluation and Metrics

This Section overviews relevant performance metrics and discusses their merits and limitations. The different metrics are grouped into two classes by the ITU-T and the IETF. The first class consists of *objective* performance metrics, referred to as *Grade of Service* by the ITU-T and

*Quality of Service (QoS)* by the IETF. Objective metrics comprise all data that can be measured directly, such as packet delay and jitter, throughput, packet loss, and the like. Moreover, metrics that can be inferred from this data are considered objective metrics. In contrast, the second class of *subjective* performance metrics cannot be measured directly. Subjective metrics are collected under the terms *Quality of Service (QoS)* by the ITU-T and *Quality of Experience* by the IETF. The ITU-T defines subjective metrics as "the collective effect of service performances which determine the degree of satisfaction of a user of the service" [TR 101 112]. Subjective metrics are always specific to a particular application or service. For some applications, it is possible to infer a subjective performance metric from measured objective metrics, such as for speech. However, most applications, like for example video applications, require extensive test trials with test users.

In the following, only objective metrics will be considered. Two important metrics have already been introduced in Section 2.2, namely the overall spectral efficiency and the spectral efficiency at the cell edge. When multiplied with the available bandwidth, they yield the aggregate cell sector throughput and the cell edge throughput, respectively. While it is straightforward to determine the aggregate throughput in a cell sector, this is not the case for the performance at the cell edge. Clearly, the performance has to be measured for terminals that are close to the cell edge, but it is not obvious where exactly the cell edge area begins. One approach for this problem is to measure the cumulative distribution function (cdf) of all terminal throughput values in a cell sector, which basically shows how often a particular throughput result can be expected. In other words, if the cdf is obtained from a random and uniform user distribution, the cdf allows determining the probability that a randomly placed mobile terminal will achieve a certain throughput. [TS 25.814] proposes to take the 5% quantile from this cdf as a scalar metric for the cell edge throughput. The 5% quantile indicates the throughput values of those mobile terminals with the lowest throughput in the scenario, which are in general located close to the cell edge. This metric has also been adopted by the NGMN Alliance [NGMN2007].

In a Monte-Carlo simulation, the 5% quantile is easy to determine, since every drop results in a set of throughput values for the randomly placed mobile terminals. This allows a simple calculation of the cdf. In an event-driven simulation with full terminal mobility, this is more difficult. In order to determine the quantile, the individual throughputs of the terminals within a short time period $T_{STP}$ have to be measured. Subsequently, the cdf and the quantile of these short-term averages can be calculated. Obviously, the short-term period $T_{STP}$ is an important parameter, which has a strong impact on the result. It must be long enough to even out effects of the MAC procedures, such as segmentation and retransmissions, but short enough to capture holes in the coverage.

In addition to the just discussed throughput metrics, there are a number of indirect metrics that give an indication for the expected throughput without requiring a detailed system model. Most important, the SIR and SINR directly relate to the maximum achievable spectral efficiency by the Shannon–Hartley theorem. Along with the SIR and SINR, it is important to study the occupied bandwidth in order to judge the resulting throughput. The occupied bandwidth is reflected in the resource utilization, where an occupation of the full available bandwidth corresponds to a resource utilization of 100%. Alike, a network with frequency reuse 3 has a resource utilization of up to only $1/3$. It is obvious that an increase of the resource utilization counteracts an improvement of the SI(N)R, leading to a tradeoff between these two metrics. Note that the

SI(N)R's mean value is a problematic metric, since it is dominated by high values close to the base station. If a single scalar value characterizing the SINR or SIR conditions is desired, it is better to consider the median, which is not susceptible to outliers of the SIR.

Besides the above explained scalar metrics, it is also interesting to study position-dependent metrics. Position-dependent metrics illustrate the performance depending on the location of a mobile terminal. This is somewhat the case already for the throughput quantile, which depends on the position in the way that it corresponds to the throughput close to the cell edge. In order to obtain further insight into network coverage and position-dependent performance, it is convenient to plot for example the throughput or the SINR over the area. For this, an observation area is defined as indicated in Fig. 3.4(a), within which the respective metric is plotted depending on the position.

## 3.6 Optimization of Cellular Networks

Optimization describes the process of finding, for a particular problem, the best solution with respect to a particular goal from a set of possible alternatives [Lit92]. Optimization is an inherent part of many real-life problems. For example, the production of most industrial goods requires a sophisticated, geographically distributed supply chain. An optimal planning of such a supply chain could for example be done with respect to minimizing the overall cost of transportation and temporary storage. This problem is known as the warehouse problem [Sou73]. A large number of optimization problems also exist in the area of mobile communications [Boo07]. Offline problems occur during the network planning process. Examples are the assignment of frequency blocks to cells such that a certain minimum SINR is achieved at any location [MD79], or the optimization of the size of location areas [Tho88]. Online problems occur during the resource assignment process, for example when using channel borrowing schemes that allow a cell to use radio channels originally assigned to neighboring cells [Sal95]. At the frame level, the scheduling of radio blocks constitutes a highly dynamic optimization problem, often aiming at maximizing the system throughput while taking into account several constraints, such as QoS requirements [Nec06a].

### 3.6.1 Introduction to optimization

In general, the formulation of an optimization problem consists of a vector $\vec{x}$, an objective function $f(\vec{x})$, and a number of constraints which limit $\vec{x}$ to a set $\boldsymbol{S}$ of feasible solutions:

$$\text{minimize / maximize} \quad f(\vec{x}) \tag{3.17}$$

$$\text{subject to} \quad \vec{x} \in \boldsymbol{S} \ . \tag{3.18}$$

If $f()$ is non-linear in $\vec{x}$, or if the constraints have to be expressed with an arbitrary non-linear relationship, the problem is referred to as a non-linear optimization problem. Problems of this class are often difficult to solve. If $f()$ is linear in $\vec{x}$, and if the constraints can be formulated as a linear inequality system, the optimization problem is called a linear optimization problem,

which is easier to solve. The optimization problem can then be formulated as a *Linear Program* (LP) with the following standard form:

$$\max\{\vec{c}^T\vec{x} \mid \boldsymbol{A}\vec{x} \leq \vec{b},\ \vec{x} \geq 0\} \ . \tag{3.19}$$

The vector comparisons in the constraints are meant as a component-wise comparison for every vector index, i.e. $x_i \geq 0\ \forall i$ and:

$$\sum_j a_{ij}x_j \leq b_j \ . \tag{3.20}$$

The vector $\vec{c}$ directly follows from $f()$. The matrix $\boldsymbol{A}$ and the vector $\vec{b}$ follow from the constraints $\boldsymbol{S}$. Many real-life problems of this type can be solved efficiently, for example by the well-known Simplex algorithm [Lit92].

A further class are discrete optimization problems, where the elements of $\vec{x}$ are required to be integer. A special sub-class are *Integer Linear Programs* (ILPs) with the additional constraint of linearity in the objective function and the constraints[1]. The optimization problem can then be formulated as an ILP with the following standard form:

$$\max\{\vec{c}^T\vec{x} \mid \boldsymbol{A}\vec{x} \leq \vec{b},\ \vec{x} \geq 0,\ \vec{x} \in \boldsymbol{Z}^n\} \ . \tag{3.21}$$

If the integer requirement is valid only for some elements of $\vec{x}$, the problem is referred to as *Mixed Integer Program* (MIP). If the elements of $\vec{x}$ are required to be binary variables, the problem is a *Binary Integer Linear Program* (BILP). Besides, *combinatorial optimization problems* are a special problem type closely related to ILPs. In a combinatorial optimization problem, only a countable number of solutions from a discrete set exists. Often, solutions are created by combining a certain number of elements from a set of discrete elements. For example, in the well-known Traveling Salesman Problem, the discrete elements are places, and the discrete set of solutions consists of all possible permutations of places. Combinatorial optimization problems can often be formulated as ILPs [Dom05], and consequently, they can be solved by the same algorithms.

Due to their integer restrictions, ILPs are much more difficult to solve than regular LPs. In particular, they belong to the class of NP-complete (Nondeterministic Polynomial-time complete) problems. This type of problem cannot be solved in polynomial time on a regular Turing machine, given that the problem class $P \neq NP$, where $P$ is the class of problems for which there exist polynomial-time solution algorithms. In general, the computational effort to find the optimal solution highly depends on the specific ILP. Owing to their binary restrictions, BILPs are a particularly difficult type of ILP. One possible solution approach is a branch-and-bound algorithm ([Lan60, Dak65]), which is used for example by the BILP-solver of the Matlab Optimization Toolbox [Mat07]. Every time a branch decision is made, an LP relaxation of the original BILP is solved using well-known solution algorithms for LPs. This procedure guarantees to find the optimal solution to the BILP if it is solvable under the given constraints. The downside of this method is its big computational complexity. Besides an exact method, a heuristic can be applied if it is sufficient to obtain an approximate solution. Often, problem specific heuristics are developed to efficiently solve a particular problem type. In contrast, *metaheuristics* can be used as a generic tool to solve arbitrary integer and combinatorial optimization problems. This includes methods like tabu search, simulated annealing, or evolutionary algorithms. If these algorithms

[1] Strictly speaking, linearity is only achieved if the elements of $\vec{x}$ were real and not integer.

are enriched with problem specific knowledge, near-optimal solutions can be obtained for many problems in an acceptable time.

Evolutionary algorithms are a particularly generic but also powerful tool to solve arbitrary integer and combinatorial optimization problems. The following Section 3.6.2 gives a brief introduction to the field of evolutionary algorithms. Subsequently, Section 3.6.3 focuses on genetic algorithms, which are a popular instance of evolutionary algorithms.

### 3.6.2 Introduction to evolutionary algorithms

In the past decades, evolutionary algorithms have attracted increasing interest for the solution of optimization problems. In particular, this applies to combinatorial optimization problems, which would otherwise be hard to tackle by exact methods. Evolutionary algorithms belong to the class of nature-inspired algorithms. They are based on the evolution theory in nature, whose modern foundations were laid by Charles Darwin (1809-1882) [Dar59]. According to the general understanding of evolution theory, life forms undergo a constant (self-)optimization while trying to adapt to their changing habitat. For example, polar bears have adapted to their cold environment by developing a special kind of fur where every hair functions as a small optical fiber conducting sun light to the skin for heating purposes. This process takes place during a long time period by developing new *generations* of *individuals* (also referred to as *solutions*), where every new generation contains the offsprings of the individuals of the previous generation. All individuals of one generation are referred to as *population*. Over the course of the generations, the individuals develop new characteristics, which may or may not be favorable for their survival. This leads to one of the basic principles of evolution theory, namely the *survival of the fittest*. When moving from one generation to the next, a *natural selection* takes place. Individuals that are well adapted to the environment (having a good *fitness*) are more likely to survive and reproduce than individuals that are badly adapted (having a bad fitness). The beneficial characteristics of the surviving individuals are passed on to their offsprings. New characteristics are introduced by random mutations that may occur during the reproduction, thus creating offsprings with a better or worse fitness. Over the course of time, this leads to a constant adaptation and optimization of a species to its environment.

The physical appearance, properties, and characteristics of an individual are determined by its genetic information. It consists of a certain number of *genes*, which may take on certain values. The possible values of a gene are referred to as *alleles*. The genetic information can be regarded as coded information, which is called *genotype*. In contrast, the decoded information yields the actual observed physical appearance and characteristics of an individual. This is referred to as *phenotype*.

Several factors govern the course of the evolution process. The *mutation* determines how genes are mutated from one generation to the next. A mutation in nature always targets the genotype, i.e., the mutation varies the genes of an offspring. Some of these variations will then alter the phenotype of the new individual, which has an impact on the second important factor, namely the *selection*. The selection process determines which individuals are chosen to reproduce for the following generation. In nature, it is likely that individuals with a high fitness will mate, but also individuals with a low fitness value have a chance of creating offsprings. The third major

factor is the *recombination*, often referred to as *crossover*. It determines how the genomes of two individuals are merged during the mating process.

Besides these important factors, a large number of other factors influence the evolution. In nature, the phenomenon of *co-evolution* plays a significant role. It refers to the mutual dependence of different species, such as for example in predator–prey relationships. Furthermore, the inner structure of a population is important. Natural populations are often divided into smaller sub-populations (called *demes*), separated by rivers or mountain ridges. A limited exchange of genetic information takes place by *migration* of individuals from one sub-population to the other.

It is obvious that the evolution as described above involves a large amount of randomness. This applies, among others, to the mutation, the crossover, and the selection. Therefore, the offsprings of a certain generation are random to a certain degree. However, this randomness is channelized due to the inheritance of features from one generation to the next. Colloquially, one could therefore describe the evolution process as structured probing of the possible solution space.

When applying evolutionary algorithms to real-world optimization problems, several tasks have to be performed. First, a suitable *representation* of real-world solutions has to be found. This goes along with finding a method for determining the fitness of an individual, where the fitness is usually a real number. Furthermore, the *genetic operators* have to be defined, such as selection, mutation, and crossover. Alike, the management of the population has to be decided on. Other considerations may be necessary, depending on the problem and the particular instance of the evolutionary algorithm. Finally, the resulting algorithm may exhibit a large number of parameters. Section 5.3 will highlight this process for an optimization problem occurring in the proposed distributed interference coordination scheme. At this point, a brief overview of the different flavors of evolutionary algorithms and a more detailed introduction to genetic algorithms will be given.

Within the overall field of evolutionary algorithms, three main schools of thought have developed in the early years (see Fig. 3.6). These are *genetic algorithms*, *evolution strategies*, and *evolutionary programming* [Poh99, Nis97]. Initially, these schools were developed independently by relatively small research groups with no significant exchange among them. This has changed since the 1980ies, during which the individual schools were united under the all-embracing term of evolutionary algorithms. Since then, the differences have become somewhat diffuse[2]. In the following, the three main streams are introduced and characterized. More detailed information and a historic synopsis can for example be found in [Nis97].

- **Genetic Algorithms (GAs)**

  Genetic Algorithms (GAs) date back to the works of John Holland in the 1960ies. Classically, the representation of real-world solutions is done by means of bit strings, i.e., a binary solution coding is used. The bit strings are the genomes of the individuals, hence GAs operate on the level of the genotype. The representation is not limited to bit strings, and many applications suggest a more “natural” representation, for example by arrays of integer or real numbers. There are several standard genetic operators that are more

[2] Interestingly, this can be regarded as one example of sub-populations and migration.

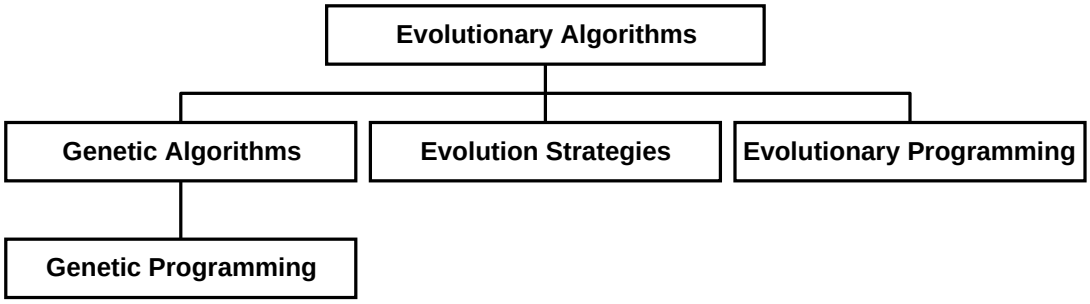



**Figure 3.6:** Classification of evolutionary algorithms.

or less suitable depending on the specific problem. Selection is done in a non-extinctive way, meaning that all individuals of a generation have the chance of producing offsprings. The main search operator to traverse the possible solution space is the crossover operator. Mutation is only a background operator whose major responsibility is to prevent the algorithm from being trapped in a local optimum. Genetic algorithms will be described in greater detail in Section 3.6.3.

A special form of genetic algorithms is *genetic programming*. It is used for program induction, which describes the process of automatic computer program generation. The induction is based on given training data, and the program shall solve a particular problem. The initial population is a set of randomly generated programs. The original work by John Koza is based on the programming language LISP, which has some favorable properties for genetic programming [Nis97]. For example, in [Koz91], Koza evolved a LISP program to generate random numbers. However, genetic programming was also used in conjunction with more common programming languages. In [Nor94], machine code programs were automatically generated which were able to distinguish between Swedish nouns and verbs based on their spelling. The authors showed the superior performance of their approach compared to neural networks.

- **Evolution Strategies**

  In contrast to genetic algorithms, which operate on the level of the genotype, evolution strategies operate on the level of the phenotype. Evolution strategies thus have a higher degree of abstraction and are often used to optimize continuous decision variables [Nis97]. The representation of solutions is always an array of real numbers. In contrast to genetic algorithms, the selection is deterministic and extinctive, meaning that individuals with a bad fitness value have no chance of producing offsprings. The main search operator is the mutation, which is done by adding normal distributed random numbers to the solution array elements. The standard deviation of the random numbers can be adjusted adaptively during the run time of the algorithm. Besides the mutation operation, the crossover is also used as a second important operator.

- **Evolutionary Programming**

  Originally, evolutionary programming was used to generate artificially intelligent finite state machines. These state machines were used for the prediction of time series. Given a series of elements from a finite alphabet, they should predict the next element. The only search operator is mutation, crossovers are not considered. The mutation step-size, which

is a real-numbered random variable like in evolution strategies, depends on the parent's fitness value. As the algorithm approaches the optimum, the mutation step-size becomes smaller. The selection process is stochastic and extinctive.

As mentioned before, the borderlines between these originally separate schools of thinking have become diffuse. Still, there are some distinct features characterizing each school. However, when applying an evolutionary algorithm to a particular real world problem, it is not beneficial to focus on only one of them. [Nis97] suggests a modular design principle, where specific evolutionary algorithms are created by taking building blocks from possibly different schools. These building blocks comprise population models, solution codings, concepts for handling constraints, search operators, selection operators, and modules for statistical evaluation [Nis97]. The following sub-section will introduce the most important building blocks at the example of genetic algorithms.

### 3.6.3 Overview of genetic algorithms

In this Section, genetic algorithms are studied in more detail. First, the basic principle of genetic algorithms is reviewed, followed by an overview of the most important population models. Subsequently, different selection strategies are discussed. Finally, standard crossover and mutation operators are presented. Only a brief overview of the various aspects is given in the following sub-section. More details can be found in one of the many books on genetic algorithms (e.g., [Nis97, Poh99, Bäc00a, Bäc00b]).

#### *3.6.3.1 Basic procedure of the genetic algorithm*

The flow chart of a generic genetic algorithm is shown in Fig. 3.7. In the beginning, an initial population $\boldsymbol{P}$ of solutions (individuals) is generated, which is usually done stochastically. These initial solutions are then evaluated. Afterwards, the evolution process begins. First, individuals are selected from the population for reproduction by a *selection operator*. These individuals are then manipulated by the *crossover* and the *mutation operator*. Finally, a new population is created and the algorithm checks whether it should terminate. Upon termination after a total number of $N_{gen}$ generations, the so far best solution is selected as the final solution. In the following, each of these steps will be described in more detail. Note that the size $|\boldsymbol{P}|$ of the population is identical after every generational step.

The generic procedure in Fig. 3.7 serves as a basis for a particular instance of a genetic algorithm. They differ in how the genetic operators are applied and how populations are managed. The flow chart for a simple *generational genetic algorithm* is shown in Fig. 3.8. This simple GA uses non-overlapping populations, meaning that every individual in the population set $\boldsymbol{P}_N$ of generation $N$ is replaced by a new individual in the following generation $N+1$. This is also referred to as *generational replacement*. During the transition from generation $N$ to generation $N+1$, a fraction of $R_C$ individuals is newly created by applying the crossover and mutation operators. The remaining individuals are modified solely by the mutation operator. The fraction $R_C$ is referred to as *crossover rate*. The selection operator (see below) determines which individuals are selected for the mating with the crossover operator. One disadvantage of this

strategy is obvious: The best individuals of generation $N$ will be lost with a certain probability, since every individual is modified by a genetic operator. This can be avoided by *elitism*, which always copies the best individual (or the best $i$ individuals) to the following generation.

An alternative is to use *overlapping populations*, so done by the *steady state genetic algorithm*. The basic idea is to replace at most a certain fraction $R_R$ of the population during each generational cycle. Thus, only $R_R \cdot |\boldsymbol{P}_N|$ individuals are selected from $\boldsymbol{P}_N$ for reproduction, which takes place like in the simple GA variant above. Classically, only one or two offsprings are created and inserted into the new population [Nis97, Dav91]. The general form of the steady state algorithm allows values of $R_R$ of up to 100% [Dav91]. A further degree of freedom is the set of individuals that compete for survival when the population of a new generation is finalized. In its original form, a sufficient number of individuals is deleted from the population of the old generation (usually the worst) to make room for the newly generated individuals (see for example [Nis97, Sys89]). Alternatively, all members of the old generation can compete with all offsprings for survival in the new generation, possibly deleting some of the newly generated offsprings. This was for example described in [Sar00] and [Rud00]. The corresponding flow chart is shown in Fig. 3.9.

Niching techniques constitute a further variant of genetic algorithms with respect to their population structure. One of the key aspects for the success of a genetic algorithm run is the het-

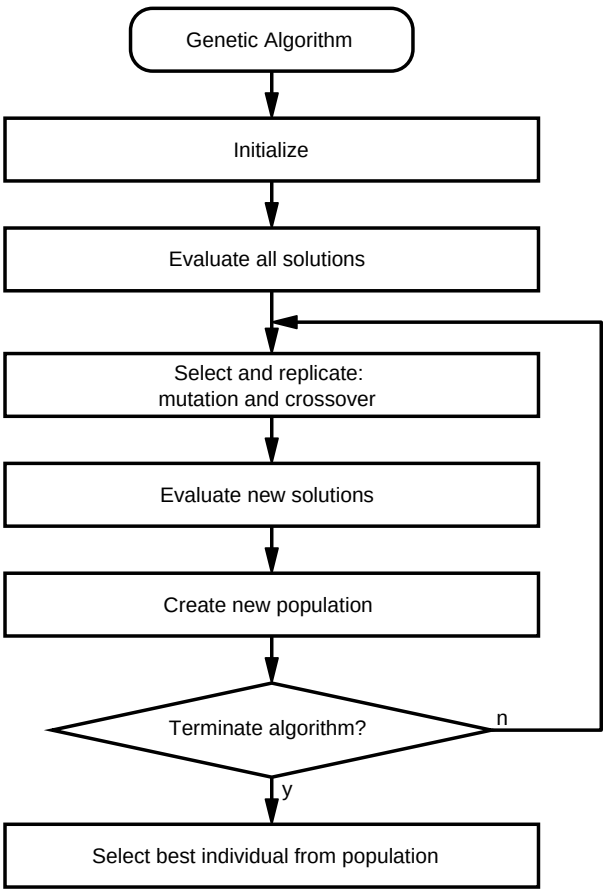



**Figure 3.7:** Generic flow chart of a genetic algorithm.

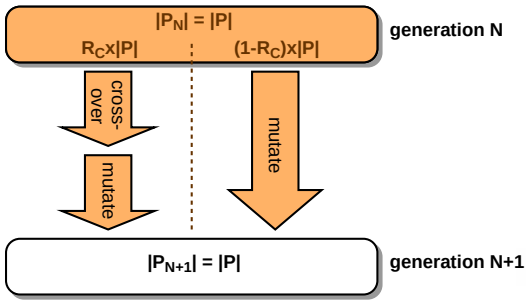



**Figure 3.8:** Illustration of one generation in a simple genetic algorithm.

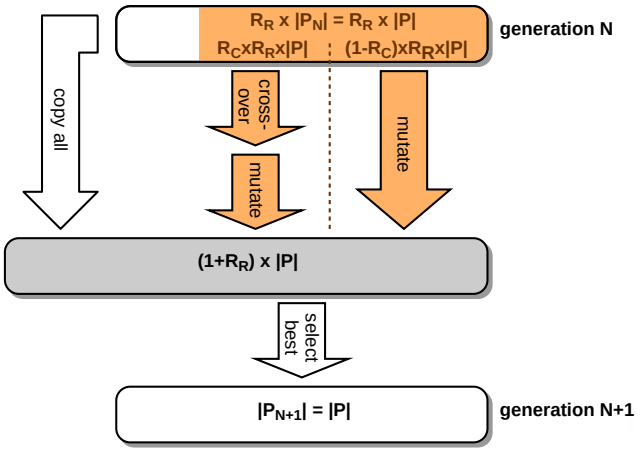



**Figure 3.9:** Illustration of one generation in a steady state genetic algorithm.

erogeneity of the population. If the individuals become too similar, the algorithm will stagnate. Niching techniques can be used to increase the heterogeneity of the population by imitating the formation of niches in natural eco systems. In such a technique, offsprings are combined with their most similar parent in order to avoid too many similar solutions within one population. Furthermore, the fitness value of solutions is reduced if there are many similar solutions. Details of this technique can be found in literature (e.g., [Nis97]).

Finally, deme genetic algorithms emulate sub-populations with migration of individuals between the sub-populations. They are particularly suitable for parallel implementation, since the sub-populations can be evolved independently between migration steps. Further details on the population structures in genetic algorithms can be found in [Bäc00b].

#### *3.6.3.2 Solution representation and evaluation*

The coding of the solution in a genetic algorithm depends very much on the problem under investigation. Among the data structures used for genomes are one- or multi-dimensional arrays, lists, and trees. The genes within these data structures may be of different types. Besides the classical binary genes, integer or real numbers may be used. In the most general form, the genes constitute of references to arbitrary objects. In general, the genes can take on any of their alleles. That means some solutions may contain gene values that others do not contain. In contrast, *permutation coding* is a special form of coding where all genomes are permutations of a fixed set of genes. This means that every solution can be obtained from any other solution by permuting the genes of that solution's genome. Permutation coding is particularly useful when solving combinatorial optimization problems. For example, in the traveling salesman problem, the solutions are permutations of the different places the salesman has to visit in a particular order.

Independent of the representation, there has to be an *evaluation operator* that calculates a single real-valued fitness value $F$ for every possible individual. This evaluation process may make it

necessary to determine the individual's phenotype from the genome, which may be a non-trivial procedure. Note that the evaluation operator is highly problem specific just like the representation.

#### *3.6.3.3 Selection operators*

The selection operator chooses individuals from the population $\boldsymbol{P}_N$ that are then used for reproduction. This is usually done based on the fitness of the solutions. While it is desirable to select good solutions for reproduction, the relatively bad solutions of a population should also be able to reproduce in order to not lose genetic diversity. A selection algorithm that heavily favors good solutions is said to have a high *selective pressure*. Genetic algorithms usually apply a stochastic selection algorithm that selects good solutions with a high probability, allowing bad solutions to reproduce, too, albeit with a smaller probability. This is in contrast to deterministic selection operators used for Evolution Strategies, which for example select the best $n$ solutions while ignoring the bad ones. In the following, the most important stochastic selection operators are summarized.

- **Fitness proportionate selection**

  Solutions are selected such that the expected value of the number of times a certain solution is chosen is proportional to its fitness value. This can be done by the well-known roulette wheel selection, or by the more sophisticated Stochastic Universal Sampling [Nis97]. Fitness proportional selection is a commonly used technique, but has a few drawbacks. First, the selection probability depends on the absolute fitness value of an individual. This prevents negative fitness values, and it usually makes it necessary to scale the fitness values such that they fall into a suitable value range. Moreover, selection becomes random when the fitness values of all solutions are more or less equal, for example towards the end of the algorithm run or when the fitness values are scaled badly. Another problem may arise from a *supersolution* [Bäc00a] with an extremely high fitness value, which will then be selected with an equally high probability. This may reduce the genetic diversity and may lead to the genetic algorithm converging to a sub-optimal solution.

- **Ranking selection**

  Ranking selection sorts the solutions with respect to their fitness values. The selection probability of an individual then depends on its position in this sorted list. Ranking selection gets along without scaling and also allows negative fitness values. However, it ignores the absolute difference in the fitness values.

- **Tournament selection**

  In a tournament selection, a number $N_{ts}$ of solutions is randomly picked from the population. These solutions then compete in a tournament where the solution with the best fitness value is finally selected. $N_{ts}$ is called the tournament size. It can be used to adjust the selective pressure, where larger values of $N_{ts}$ lead to a larger selective pressure. Tournament selection has become a popular selection algorithm. It is independent of the fitness scaling, allows for negative fitness values, has an adjustable selective pressure, and is easy to implement.

Various other selection operators have been proposed in literature, such as Boltzmann selection, which will not be detailed here. A good overview of selection operators including many details can be found in [Bäc00a].

#### *3.6.3.4 Search operators*

The two search operators mutation and crossover are used to explore the solution space. Traditional genetic algorithms use the crossover operator as the main search operator, while the mutation is only a background process [Nis97]. This view has changed over time, giving the mutation operator more attention in research. In general, all search operators should be tailored to the representation of the solution and the problem under investigation. Various variants have been proposed in literature. This is particularly the case for search operators working on list or array representations, which are the most popular representation form for genetic algorithms. In the following, the most important operators known from literature will be outlined.

##### *3.6.3.4.1 Mutation operators*

During the mutation process, the mutation rate $R_M$ determines the degree of mutation. If the mutation is considered a background process only, it ought to be chosen small (typically 1% or less). $R_M$ is interpreted differently depending on the specific mutation operator. Popular mutation schemes include flip mutation, swap mutation, and inversion mutation:

- **Flip mutation**

  The flip mutator is easiest applied to bit strings. In this case, it randomly inverts every binary gene with probability $R_M$ [Poh99]. An example is shown in Fig. 3.10. When applied to non-binary representations, it randomly selects a new value for every gene out of its alleles, where every gene is modified with probability $R_M$.

- **Swap mutation**

  The swap mutator randomly swaps two genes of the genome. This is illustrated in Fig. 3.11. There are different possibilities to implement such a mutator. In the implementation used for the performance evaluation in this monograph, the genome is traversed and every gene position is selected with probability $R_M$ for swapping with another gene position. The counterpart position is selected randomly. It is thereby possible to select the same position, leaving the gene unmodified.

- **Inversion mutation**

  This mutator inverts the genome between two randomly selected positions. It is particularly useful for the traveling salesman problem.

Many more mutation schemes have been proposed, such as simply adding or removing genes. When selecting an existing or designing a new mutator, several constraints from the representation or the problem under study have to be considered. For example, some representations rely

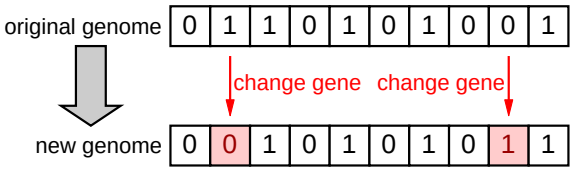



**Figure 3.10:** Example of a flip mutation on a binary bit string.

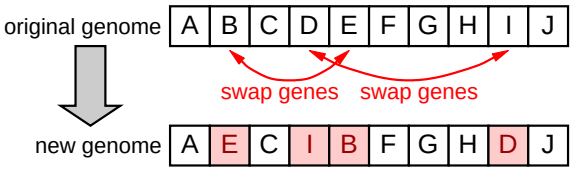



**Figure 3.11:** Example of a swap mutation with array representations.

on a fixed number of genes in a genome, rendering all mutators useless that change the number of genes. Other representations may limit the number of certain alleles in the genome, which poses restrictions to mutators like the flip mutator.

#### *3.6.3.4.2 Crossover operators*

Crossover operators are responsible for the recombination of solutions. They exchange genetic information between two or more parent solutions, thereby producing child solutions. If the parents have a high fitness value, the offsprings may inherit the good properties of the parents, leading to even better solutions. A large number of crossover operators have been proposed in literature for list and array representations. A brief overview of the most important ones will be given in the following paragraphs.

A simple example is the *N-point crossover* illustrated in Fig. 3.12. It selects $N$ random positions in the parent genomes that divide them into $N+1$ segments. These segments are copied into the offsprings while alternating the source parent from segment to segment. The $N$-point crossover has a positional bias, meaning that not all gene positions are treated equally [Nis97]. In particular, the leftmost positions are less likely to be modified during the crossover than a randomly chosen position. This is particularly the case for small values of $N$. It can be alleviated by shuffling the genes before the crossover, and un-shuffling them afterwards. This is also known as *shuffle crossover*. A generalization of the $N$-point crossover is the *diagonal crossover*, which produces $M$ childs from $M$ parents. Its operational sequence is similar to the $N$-point crossover, but takes alternating segments from all parents for each child [Nis97].

The $N$-point crossover changes the number of particular gene values within a solution. However, this is not desirable when working with permutation coding. In this case, the number of times a particular gene value occurs within a genome is predefined. Changing this number would result in an invalid solution. *Sequencing operators* have been developed as a special kind of crossover that explicitly take into account this constraint. Many sequencing operators were developed for particular problem sets. The traveling salesman problem and other scheduling and resource allocation problems are prominent examples. For example, the *edge recombination operator* takes into account adjacencies of genes, making it well suitable for the traveling salesman problem. Another example is the *Partially Matched Crossover* (PMX). It was originally introduced in [Gol85] under the name *Partially-Mapped Crossover*, but later on referred to as Partially Matched Crossover by the same author [Gol89]. It is based on the 2-point crossover, with a subsequent step to "repair" the resulting genomes. This is illustrated in Fig. 3.13. First, a 2-point crossover is executed, which defines the *matching region* in-between the two randomly

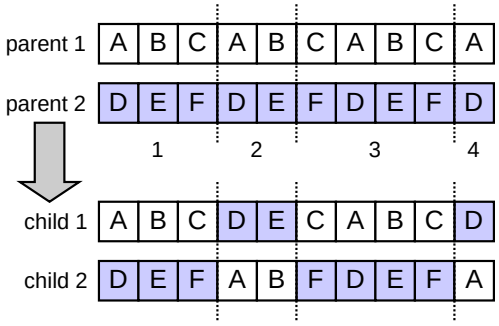



**Figure 3.12:** Example for an $N$-point crossover with $N = 3$.

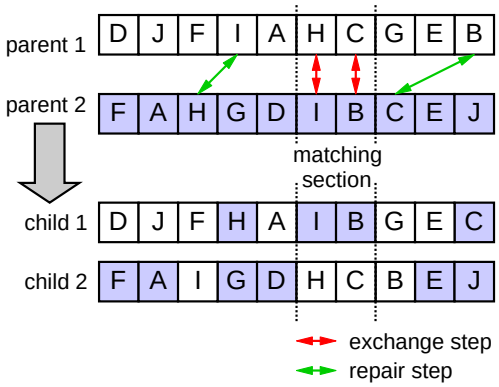



**Figure 3.13:** Example for a Partially Matched Crossover (PMX).

selected points. Subsequently, the pairs of exchanged elements (H–I and C–B in Fig. 3.13) are also exchanged outside the matching region, resulting in an unmodified number of a particular gene value. Many more sequencing operators exist. A good overview including a comparison for the traveling salesman problem and a warehouse problem can be found in [Sta91].

#### *3.6.3.5 Design of problem specific genetic algorithms*

For a good performance of the GA, the components presented in the previous sub-sections require coordination. For example, the search operators need to be adapted to the solution representation. This includes choosing appropriate search operators from literature, as well as developing custom problem specific ones. For example, if a permutation coding is used, only certain search operators make sense, while others would produce invalid solutions, thus weakening the performance of the GA. Section 5.3 will demonstrate this process for a BILP, which is part of the proposed distributed interference coordination algorithm.

# 4 Interference Coordination in Cellular OFDMA Networks

As mentioned in Chapter 2, Interference Coordination (IFCO) is a powerful tool to solve the problem of inter-cell interference in a cellular network. Moreover, interference coordination algorithms can be designed to also perform load balancing between cell sectors. Besides cellular networks, IFCO can also be applied to multi-hop and mobile ad hoc networks, where it has been an active research area for several years. In the area of cellular networks, IFCO algorithms have been proposed for circuit switched networks as early as 1971 by Cox and Reudink [Cox71]. For packet switched networks, it has become popular only recently, in particular in the course of IEEE 802.16e and 3GPP LTE standardization.

This Chapter first overviews the design space of IFCO algorithms. In particular, Section 4.1 deals with input parameters and controllable resources, followed by a classification of IFCO schemes in Section 4.2. Subsequently, Section 4.3 discusses the interaction of IFCO and scheduling. Section 4.4 briefly introduces architectural frameworks for the implementation of IFCO algorithms as they are currently being standardized by the IEEE and the 3GPP. An overview of the state-of-the-art in IFCO is then given in Section 4.5. A novel graph-based scheme for IFCO is then introduced in Section 4.6, which will be the basis for the proposed distributed IFCO algorithm later on in Chapter 5. Finally, the Chapter is concluded by a discussion of related work in the area of ad hoc and mesh networks in Section 4.7.

## 4.1 Input Parameters and Controllable Resources

In order to achieve the goal of coordinating transmissions in neighboring cells, IFCO algorithms may exercise control over various different resources and variables in the cellular network. Thereby, the decisions of the algorithm can be based on various input parameters. The following sub-sections overview commonly used input parameters and controllable resources and variables in cellular networks.

### 4.1.1 Input parameters

While there exist many possible input parameters for IFCO, the following input parameters are most commonly used:

- **Position of mobile terminals**

  The position of a mobile terminal in absolute or relative coordinates may be obtained by means of triangulation [Caf98, Nis00] or by help of satellite based positioning systems, such as GPS.

- **Direction of Arrival (DoA)**

  In a multi-path fading environment, signal components may arrive at the transceivers from different directions. The IFCO algorithm may take into account the DoA of only the strongest signal component, or of the first few strongest signal components.

- **Signal measurements**

  A further level of detail is obtained by taking into account signal levels, such as the level of the signal of interest $S$, the interference levels $I_i$ of the $n$ strongest interferers, or the sum $I$ of all interfering signals. Instead of separate signal components, the algorithm may use SINR values, or alternatively the geometry factor. Measurements can be taken from the received data signal, from pilots, or from the preamble. They may be frequency selective or they may be averaged over a certain frequency band. Moreover, they may reflect fluctuations due to fast fading in the time domain, or they may be average values over a certain time period, thus only reflecting the current shadowing conditions. The availability of signal measurements is highly system dependent, and practical issues like measurement overhead may often prevent the availability of detailed information such as measurements over the complete frequency range in an OFDMA network.

- **Channel quality**

  Directly related to signal measurements, the channel quality can be used. It can be derived in a way analog to the signal of interest or the interfering signals, and thereby allows the same degrees of freedom with respect to instantaneous or averaged measurements.

Further input parameters exist and may be used, such as the activity factor of mobile terminals, scheduling maps of other base stations, QoS profiles and measurements of the achieved QoS (for example the number of packet deadline violations).

### 4.1.2 Controllable resources and variables

Depending on the specific radio technology, a variety of resources and variables may be subject to control by an IFCO algorithm. The most important ones are overviewed in the following:

- **Time, frequency, and code resources**

  Transmissions in different cells may be coordinated in the time and frequency domain in TDMA and FDMA systems, respectively, by an optimized assignment of the usage of frequency sub-bands or subcarriers and time slots. Naturally, this also applies to OFDMA systems. Furthermore, coordination in time may be applied to virtually any other radio technology by coordinating the occupation of MAC frames in the time domain. In CDMA systems, code resources may be an additional resource for coordination.

- **Space**

  If beamforming antennas are used, the spatial domain may be used for coordination. The general operation is illustrated in Fig. 4.1, where Fig. 4.1(a) shows a situation where two transmissions in neighboring cells cause high interference on one another as it may occur in an uncoordinated system. In contrast, the transmissions are coordinated in Fig. 4.1(b) in order to minimize interference. The coordination in the spatial domain may leverage all degrees of freedom that the installed beamforming systems allow, such as placing nulls or arbitrarily shaping the radiation pattern (see Section 2.7.3).

- **Transmit power**

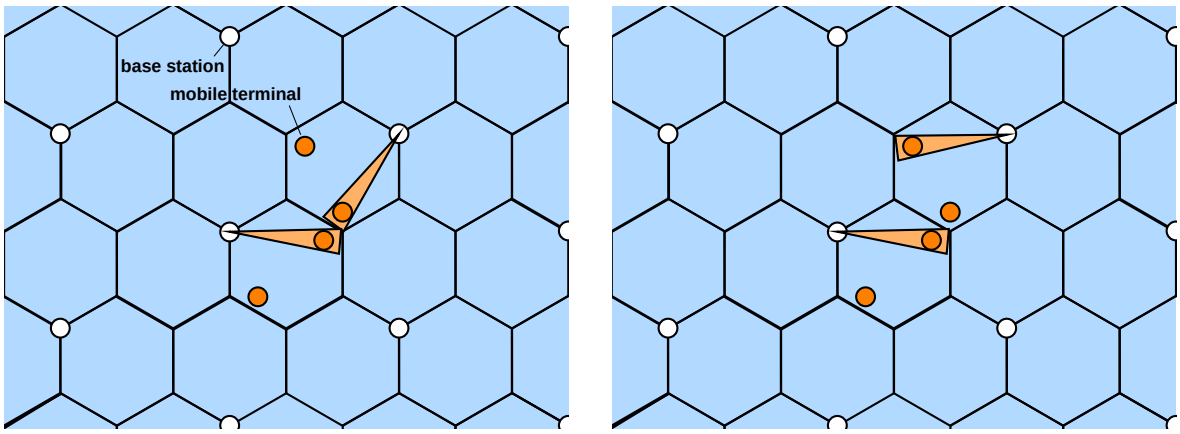
  The transmit power for certain transmissions may be raised or lowered. Lowering the transmit power reduces the amount of interference generated by one cell on another cell. Increasing the transmit power allows for the selection of a less robust MCS resulting in a lower usage of time, frequency, or code resources. This facilitates the resolution of interference conflicts in the time, frequency, or code domain as described above.

- **Modulation and Coding Schemes (MCS)**

  Closely related to the control of the transmit power, an IFCO algorithm may adapt the MCS for particular transmissions. A more robust MCS may be chosen to combat increased interference caused by other cells or to reduce the transmission power, and a less robust MCS may be chosen to achieve a lower time, frequency, or code utilization as described above.

## 4.2 Classification of Interference Coordination Schemes

Interference coordination schemes can be classified according to a variety of properties. This Section presents classifications with respect to the two most important criteria. First, Section 4.2.1 presents a classification with respect to the degree of distribution, followed by a classification with respect to the time scale of operation in Section 4.2.2.




(a) uncoordinated operation (b) coordinated operation

**Figure 4.1:** Illustration of IFCO in spatial domain.

### 4.2.1 Classification with respect to degree of distribution

The system architecture that follows from the interference coordination scheme can be divided into four major classes with respect to the degree of distribution [Nec08a].

- **Global Interference Coordination Schemes**

  Global schemes require an omniscient central device with full system knowledge. The global device is capable of acquiring the global system state in zero time, and it can distribute scheduling and other decisions instantly to all nodes (e.g., base stations) in the network. Naturally, a global scheme is not implementable in a real network. However, it provides an important reference for the system performance when studying actually implementable algorithms. The algorithm proposed in Section 4.6 is an example for a global scheme, which is based on a system-wide interference graph.

- **Distributed interference coordination schemes**

  Distributed schemes rely on one or more central components, which exchange information potentially with all nodes in the network. The central components thereby collect state information and distribute information relevant for coordination to the network nodes. Additionally, direct coordination-related communication among the network nodes may take place. In UMTS networks, the Radio Network Controller (RNC), which serves a large number of Node Bs, could be used as a central coordinator. In LTE, the meanwhile abandoned Control Plane Server (CPS) was proposed (see for example [R2-052906]), which could have taken over such functions.

  Distributed schemes can be investigated under ideal signaling conditions or under non-ideal signaling conditions. Ideal signaling refers to zero delay for any signaling message, whereas non-ideal signaling includes message delay and/or message loss. Often, signaling delay and message loss are system constraints. In principle, a global interference coordination scheme can be mapped to a distributed scheme under ideal signaling conditions.

- **Decentralized interference coordination schemes**

  Decentralized schemes are distinguished from distributed schemes by there being no central entity involved in the coordination process. The coordination is performed solely by communication and information exchange among the network nodes, which all have equal rights. If the network nodes elect one or more leaders that perform central coordination functions, the decentralized scheme turns into a distributed scheme.

  Like distributed schemes, decentralized schemes can be investigated under ideal or non-ideal signaling conditions. In the context of LTE, the signaling of a distributed algorithm takes place on the X2 interface [TS 36.300][1]. [R1-072456] mentions a message delay of 20 ms on the X2 interface as working assumption, which gives way to rather dynamic decentralized interference coordination schemes.

- **Local interference coordination schemes**

  Local schemes are based purely on information that is available locally in every base station. No coordination-related communication and information exchange with neighboring base stations takes place. Local state information includes data received from

[1] See Appendix A for an introduction to the 3GPP LTE network architecture

the mobile terminals, such as for example pilot measurements or information about the strongest interferers.

Local schemes may exercise either an implicit or an explicit coordination. Explicit coordination takes direct measures to become aware of interfering base stations or mobile terminals and tries to avoid interference-polluted resources for particular mobile terminals. In implicit coordination, the scheduling algorithms running in every base station are designed such that they automatically achieve an increased performance by knowing the scheduling algorithm of all other base stations.

In general, local interference coordination schemes are most desirable from an operator's perspective since they require no separate network equipment and provide the most flexibility with respect to the network design. On the other hand, local schemes have the least potential performance gain. Information exchange among base stations or with a central coordinator opens the way to much larger possible performance gains, however at the expense of increased system complexity. Global schemes have the highest potential performance improvement, albeit not being implementable. Their investigation is still very valuable since they can provide an estimate of the upper performance bound. Moreover, they allow studying some basic system parameters, and it may be possible to derive distributed or decentralized schemes as will be done later in this monograph.

### 4.2.2 Classification with respect to time scale of operation

Besides the degree of distribution, the time scale of operation is an important property of an interference coordination scheme. Three basic classes can be identified, namely *static schemes*, *semi-static schemes*, and *dynamic schemes*. This classification is illustrated in Fig. 4.2 together with further sub-classes of the semi-static schemes as defined in [R1-060586].

- **Static interference coordination schemes**

  Static schemes have no time variant component. They rely on a static *planning* of the interference situation in the network, usually during the network planning process. The time scale of operation is in the order of days or longer [R1-060586], and the scheme does not adapt to the present load situation. One example that falls into this category is Fractional Frequency Reuse (see Section 4.5.2).

- **Semi-static interference coordination schemes**

  In contrast to static schemes, semi-static schemes are able to handle uneven and variable load distributions among cells. Moreover, they may account for an uneven terminal distribution within a cell, for example when a large number of edge terminals cause interference to a particular neighbor cell. Semi-static schemes operate with a time interval between at least 200 ms and several seconds [R1-060135] or even minutes [R1-060586].

  *Cell load adaptive* coordination requires load estimates of the cells and operates on the level of minutes [R1-060586]. *User load adaptive* coordination takes into account the instantaneous traffic demand of the mobile terminals, which can be captured faster than the cell load. Therefore, user load adaptive schemes will operate on a time scale in the

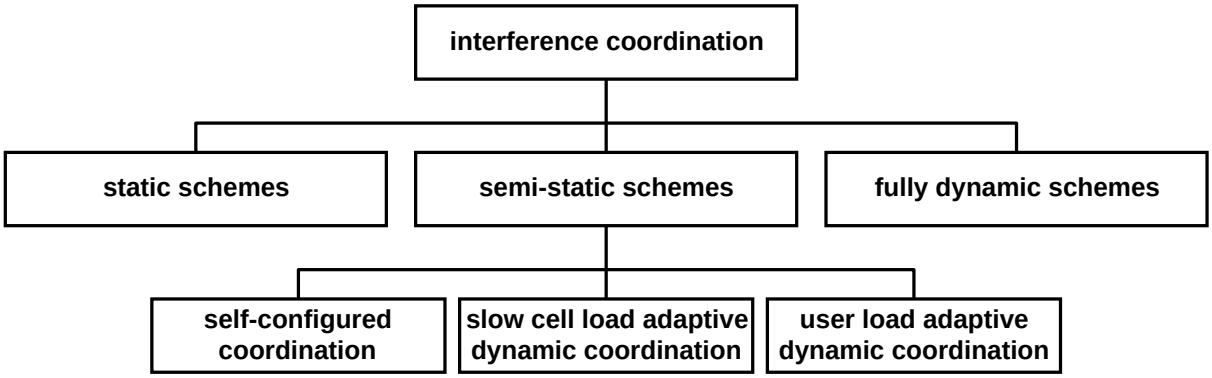



**Figure 4.2:** Classification of interference coordination schemes with respect to time scale of operation.

order of seconds or slightly below [R1-060586] or even several hundred milliseconds [R1-060135]. Metrics for the traffic demand can for example be the buffer occupancy or the QoS profile of a particular traffic flow.

[R1-060586] further identifies *self-configured coordination* as another possibility for semi-static coordination. Such a scheme is basically a self-optimizing static scheme, where the self-optimization cycle is in the order of days. One possibility given in [R1-060586] is to calculate the interference relations among the base stations from measurements obtained from the mobile terminals and use those relations to optimize the otherwise static network planning.

- **Dynamic Interference Coordination Schemes**

  Fully dynamic interference coordination schemes can instantly adapt to changing network conditions, such as changing traffic or load distributions. Their time scale of operation is in the order of one or only a few MAC frames. [R1-060586] discusses *synchronized scheduling* as the only dynamic interference coordination scheme, where the scheduling decisions of several base stations are fully synchronized. Synchronized scheduling is automatically realized by global interference coordination schemes. It is technically extremely challenging for distributed or decentralized schemes and may require a relaxation of the time scale of operation to several tens of milliseconds. On the other hand, it is rather simple if only a local coordination with synchronized scheduling among the cell sectors of one base station site is considered. Such a scheme was proposed for example by Sternad in [Ste03a] and can operate fully dynamically based only on local state information[2].

  Synchronized scheduling is not the only conceivable dynamic coordination scheme. Other algorithms may achieve dynamic coordination without fully synchronizing their scheduling decisions. Examples are implicit local coordination schemes, as they will be discussed later on.

[2]Note that the realization in an actual system may be restricted by implementation issues. The design of current base stations usually separates a cell sector completely from its neighboring cell sectors, which makes synchronized scheduling difficult. However, an integrated hardware design serving all cell sectors of one base station site would be well feasible.

The border between these classes is fuzzy. For example, a decentralized scheme could be considered dynamic if the signaling and processing delay is small, and it could be considered semi-static if the signaling and processing delay is large. However, in both cases, the algorithm itself may have remained unchanged.

## 4.3 Interaction between Scheduling and Interference Coordination

[R1-060135] notes a tradeoff between frequency-selective scheduling and interference coordination since they restrict the degrees of freedom of each other. If the allowed frequency range for a mobile terminal is limited by an interference coordination mechanism, frequency selective scheduling may not utilize the whole frequency spectrum and it will thus achieve a smaller performance gain. On the other hand, if only certain frequencies may be used, signaling overhead like channel feedback for frequency selective scheduling is reduced. In order to increase the performance of interference coordination in combination with frequency selective scheduling, the frequency-dependent channel transfer function can be considered during the interference coordination process. This is problematic for distributed and decentralized architectures, since moving terminals will exhibit a fast time-varying channel transfer function. However, it is well feasible for stationary terminals.

Furthermore, scheduling service disciplines may guarantee a certain QoS for traffic flows, which imposes additional constraints on the IFCO algorithm. Eventually, it may require the transmission of data packets of a particular data flow even if this is unfavorable with respect to the interference coordination. This issue will further be discussed for graph-based global interference coordination in Section 4.6.5.

## 4.4 Architectural Frameworks for Interference Coordination

In order to realize IFCO in real networks, the network needs to provide the necessary framework with respect to physical and MAC layer mechanisms, protocols, interfaces, and the overall network architecture. For IEEE 802.16 and 3GPP LTE, these issues have been addressed separately by the respective standardization bodies. First, physical and MAC layer mechanisms provide means to obtain relevant input parameters, such as signal levels, and they must allow separating transmissions in neighboring cells as described in Section 4.1.2. Second, an architectural framework must be provided with respect to network protocols and interfaces and with respect to the overall network architecture. The network must facilitate the communication of network nodes for the purpose of interference coordination, and it must foresee all required network components, such as a central network node for a distributed IFCO algorithm. In the following, architectural frameworks for IEEE 802.16 and 3GPP LTE are overviewed.

### 4.4.1 IEEE 802.16h

The IEEE 802.16h standard amendment [IEEE80216h] deals with coexistence mechanisms for license-exempt network operation. Essentially, it defines mechanisms for interference coordi-

nation among overlapping networks of different operators, or among networks generating significant interference on one another. These networks, which are coordinated in order to resolve interference problems, are referred to as *Community*. Mechanisms are specified for the entry of a new base station to an existing community. All base stations in a community have synchronized MAC frames. Interference coordination is then done within a newly added co-existence zone (CXZ) of the MAC frame, whereby the defined mechanisms may also be used for load balancing.

[IEEE80216h] defines three basic mechanisms for coexistence. First, a frame synchronization on the MAC layer is done, including the synchronization of downlink and uplink frame portions, thus preventing interference between downlink and uplink subframes. Furthermore, adaptive channel selection may be performed in order to find less interfered frequency subbands. Finally, interferers may be separated in the time domain.

In order to facilitate the coordination, inter-system communication may be performed on three different levels. First, cognitive radio signaling provides the simplest form of communication between systems operating with different profiles at the physical layer. This allows only a basic coordination of the systems, such as the coordination of transmit and receive intervals. More information can be communicated between systems using the second signaling mechanism of MAC-level messages. However, this is only usable if both systems operate with the same physical layer profile. Finally, information may be communicated using IP-level messages. This mechanism allows communication between systems that use different physical layer profiles, and it allows communication with a central coordinator. On the downside, communication on the IP-layer is usually slower than communication using cognitive radio signaling or MAC-level messages. Note that a mobile terminal may also relay control messages in certain situations.

### 4.4.2 3GPP LTE

The 3GPP RAN Working Group 2 study item on *UTRA UTRAN Long Term Evolution* identified three basic operational modes for IFCO in the respective document TR 3.018 *Radio Access Architecture and Interfaces* [TR 3.018]. The simplest form is IFCO based on un-solicited or explicitly requested status information, which is exchanged between base stations. This status information may contain a list of time/frequency slots that are currently used by a base station, a list of slots that will be used with a high probability, or that will not be used (with a high probability). While this option is relatively simple, it does not provide any means for a base station to actively influence the resource occupation of a neighbor base station. The second option defined in [TR 3.018] therefore provides a mechanism to issue explicit mute requests to another base station. This requires more sophisticated IFCO algorithms, and it also requires a mechanism to identify dominating interfering base stations. The third option in [TR 3.018] combines status information exchange and explicit mute requests.

## 4.5 State-of-the-Art in Interference Coordination

This Section overviews the state-of-the-art of interference coordination in cellular networks. Section 4.5.1 introduces early techniques for dynamic channel assignment in circuit switched

cellular networks. Next, Section 4.5.2 details the concept of reuse partitioning. Implicit and explicit local IFCO schemes are discussed in Sections 4.5.3 and 4.5.4, respectively. Subsequently, Section 4.5.5 surveys decentralized IFCO algorithms, followed by Section 4.5.6 on distributed algorithms. Finally, Section 4.5.7 discusses related work on global interference coordination.

### 4.5.1 Early dynamic channel assignment techniques

In circuit switched mobile phone systems, techniques for Dynamic Channel Assignment (DCA) have been investigated as early as 1971 by Cox and Reudink [Cox71]. Compared to a classical frequency reuse concept where a certain set of channels is statically assigned to a cell sector, DCA allows to dynamically reassign channels from one cell to another. The basic idea is thus the same as it is in interference coordination. In [Cox71], a one-dimensional cellular network was investigated, in which channels are taken from a global pool. This leads to increased flexibility and trunking efficiency and finally to lower blocking probabilities. Different strategies for the selection of idle channels were presented in [Cox72a] by the same authors. Furthermore, they extended their studies to two-dimensional cellular networks in [Cox72b], also with the investigation of different channel selection strategies. While Cox and Reudink used a criterion based on a reuse distance for the channel assignment, Nettleton and Schloemer proposed to assign channels based on real-time signal strength and interference measurements [Net89], with a more extensive discussion given in [Net90]. They also investigated the possibility of rearranging all assigned channels periodically in order to optimize system performance. Takenaka et al. combined a DCA scheme with transmit power control, thus further increasing the system capacity [Tak93].

A second basic scheme besides DCA is Borrowing Channel Assignment (BCA), which has been proposed by Engel in [Eng73a, Eng73b]. In contrast to DCA where all cell sectors may principally use all the channels, BCA begins with a fixed channel assignment to cell sectors and then allows a fully loaded cell sector to borrow unused channels from one of its neighbor sectors. Further enhancements to this scheme have been proposed, such as an approach based on an ordered list of channels for each cell sector where the first channels in the list are used for serving calls in the respective cell sector, and the last channels in the list are used for borrowing [Eln82]. Hence, this approach has no distinct set of own fixed channels and channels to be borrowed. In addition, improved techniques have been developed and applied to specific network architectures such as GSM [Mar92, Mar94].

While DCA and BCA were designed as flexible reuse schemes to cope with non-uniform spatial terminal distributions, they can be considered as the first proposals of interference coordination in circuit switched cellular networks. DCA has the character of a global or distributed coordination scheme, whereas BCA is more related to distributed or fully decentralized schemes. Both operate on a relatively long time scale, compared to IFCO algorithms in packet switched networks which have to react on packet arrivals in a much more dynamic way and thus operate on a time scale as small as the frame-level.

### 4.5.2 Reuse partitioning

Reuse partitioning is a form of static planning, which aims at increasing the capacity of a cellular network by using multiple different frequency reuse factors for certain mobile terminals. It was first introduced by Halpern in 1983 [Hal83] under the name *reuse partitioning* and re-invented in the late 1990ies as *intelligent underlay-overlay* [Lin96, Wil96, Nie97, Wig97]. Within the activities of the WiMAX forum and 3GPP LTE standardization, it is known as *Fractional Frequency Reuse (FFR)* (e.g., [WiM06]). While reuse partitioning and intelligent underlay-overlay stem from the era of circuit switched networks, the term FFR was coined in combination with packet switched networks.

#### *4.5.2.1 Basic principle*

The original proposal in [Hal83] was based on the idea of dividing the available spectrum into two or more disjoint frequency bands each corresponding to one resource partition. Each band $i$ is operated with an individual frequency reuse factor $r_i$. Mobile terminals in good reception conditions are then served on frequency bands with a low reuse factor, while mobile terminals in bad reception conditions are served on bands with a higher reuse factor. This optimizes the terminal's SIR by avoiding excessively good or excessively bad SIR values, and it optimizes the frequency reuse by assigning each mobile terminal to its most suitable frequency band $i$. Figure 4.3(a) illustrates this concept for the division into two frequency partitions with reuse factors $r_0 = 1$ and $r_1 = 3$. $p$ denotes the fraction of the total available bandwidth $B$ which is assigned to the reuse 1 band, i.e., $B_{reuse1} = p \cdot B$. Likewise, the overall bandwidth reserved for the reuse 3 band is $(1-p)B$, i.e., $B_{reuse3} = 1/3(1-p) \cdot B$ in each of the three reuse 3 sub-bands.

[Hal83] defines an effective frequency reuse factor $r_{eff}$:

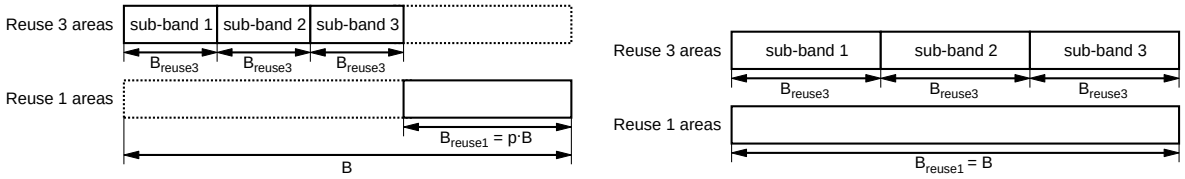
$$r_{eff} = \frac{B}{B_{cell}} \quad , \tag{4.1}$$

where $B_{cell}$ is the bandwidth available in every cell. In the case of two reuse partitions $B_{cell}$ yields to:

$$B_{cell} = \left( \frac{p}{r_0} + \frac{1-p}{r_1} \right) B = B_{reuse1} + B_{reuse3} \quad . \tag{4.2}$$

Hence, the effective reuse factor can be set in relation to $p$:

$$r_{eff} = \frac{3}{2p+1} \quad . \tag{4.3}$$




(a) Disjoint reuse 3 and reuse 1 bands

(b) Overlapping reuse 3 and reuse 1 bands

**Figure 4.3:** Partitioning of frequency domain for reuse partitioning.

From this follows:

$$B_{reuse1} = B \cdot \frac{3 - r_{eff}}{2r_{eff}} \tag{4.4}$$

$$B_{reuse3} = B \cdot \frac{r_{eff} - 1}{2r_{eff}} \; . \tag{4.5}$$

Reuse partitioning comes along with a tradeoff between increased spectrum utilization and SIR improvement. This tradeoff can be adjusted by $p$ in the case of two reuse partitions, resulting in a maximum system capacity for a certain $p$. It is important to note that the optimal value of $p$ varies depending on the distribution of terminals in the cells and the load of individual cells in the network. Moreover, it is not possible to choose the value of $p$ locally in every base station, since $p$ must be chosen identical in all cells.

In [R1-060135], Siemens presents a performance evaluation of the described tradeoff in the context of 3GPP LTE. The authors assumed a uniform terminal and load distribution throughout the network. The maximum system capacity was achieved for $r_{eff}$ between 1.3 and 1.8, depending on the fairness requirements for the mobile terminals in the different reuse partitions. Higher reuse factors are needed if mobile terminals in the reuse 3 partitions shall receive the same data rate as terminals in the reuse 1 partition in the long run, while lower reuse factors are sufficient if a certain unfairness is accepted.

#### *4.5.2.2 Partitioning of available frequency spectrum*

The original proposal in [Hal83] used disjoint frequency bands for the different reuse partitions. This was taken on by other publications, for example by [Ste03a]. Later work also proposed to use overlapping frequency bands as it is illustrated in Fig. 4.3(b). Examples are [R1-050841] and [R1-050764]. In principle, the whole spectrum can be used in every cell when using overlapping frequency bands. This makes the definition of the effective reuse factor problematic, since strictly speaking, the effective reuse is 1.

Overlapping frequency bands overcome the problem of choosing $p$. On the other hand, the SIR performance of mobile terminals served with reuse 3 is decreased, since they receive additional interference from reuse 1 terminals in adjacent cells. This problem can be alleviated by combining the reuse partitioning scheme with soft frequency reuse (see Section 2.6.1) such that mobile terminals are served with reduced power in the reuse 1 partition [R1-050507, R1-050841, R1-060291, Sim07]. [R1-072974] reported a power offset of $-4$ dB to deliver the best performance, which was obtained by simulation. Alternatively, [R1-050764] allows a cell to transmit at full power in that part of the reuse 1 partition which overlaps with its corresponding reuse 3 partition. The performance gain of soft reuse schemes has recently been questioned. [R1-072974] and [Sim07] both report that a pure frequency reuse 1 system performs better, where [R1-072974] reports particular performance advantages with respect to the cell edge throughput.

As an alternative to overlapping frequency bands, overflow techniques can be used as they were described in [Whi85, Sal88]. These techniques allow mobile terminals served with reuse 1 to

also use resources of the reuse 3 partition if all resources in the reuse 1 partition are already occupied. This allows the reuse 1 partition to be chosen smaller since traffic may overflow to the reuse 3 partition, which increases the efficiency in the reuse 1 band at the expense of the reuse 3 capacity.

#### *4.5.2.3 Assignment of mobile terminals to reuse partitions*

The assignment of mobile terminals to the different reuse partitions can be done based on various criteria. Typically, those terminals that are close to the base station experience good SIR conditions and are therefore assigned to partitions with a small reuse factor. Vice versa, mobile terminals close to the cell edge are usually assigned to partitions with a large reuse factor. This circumstance is illustrated in Fig. 4.4. The exact border between the reuse 1 and the reuse 3 areas is determined by the algorithm that assigns mobile terminals to the respective resource partition. Four basic schemes have been reported in literature:

- **Assignment based on location**

  Mobile terminals are assigned to resource partitions based on their geographic position. This requires knowledge of the terminal positions, which can be obtained for example via the Global Positioning System (GPS) or by means of triangulation [Caf98, Nis00]. Location by GPS requires additional capabilities by the mobile terminal and usually does not work inside buildings, while triangulation can be done by the network itself. The geographic borders of the resource partitions must be known, for example from network planning. A simple approach was taken by Sternad in [Ste03a] by using the distance of the mobile terminal to its serving base station to determine the resource partition.

  The disadvantage of location-based assignment is that it does not take into account the actual SINR situation. Instead, the borders of the reuse areas are pre-determined based on certain traffic assumptions. Therefore, mobile terminals may be assigned to non-optimal resource partitions, for example if neighbor cells are only lightly loaded.

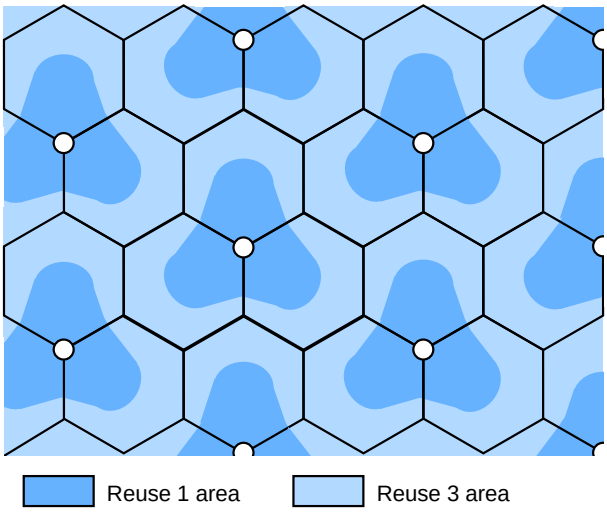



**Figure 4.4:** Illustration of Fractional Frequency Reuse with two reuse areas.

- **Assignment based on path loss**

  Mobile terminals are assigned to resource partitions based on the path loss to their serving base station. This is identical to the distance-based approach in [Ste03a] if no fading is considered and automatically adapts to terrain characteristics in a fading environment. [Sim07] reported an inferior performance of path loss based assignment compared to an assignment based on the terminal's geometry factor, which will be described next.

- **Assignment based on pilot measurements or geometry factor**

  Mobile terminals are assigned to resource partitions based on measurements of pilot signals from the serving base station and the major interfering base stations. No knowledge of the reuse area borders is required, since they are implicitly established by the pilot measurements. This automatically takes into account fading, but does not take into account the present load situation.

  In [R1-050738], [R1-060135], and [Sim07], the geometry factor is used instead of pilot measurements for the reuse area assignment. However, in practice, the geometry factor will most likely be approximated by pilot measurements (see Section 2.6).

- **Assignment based on recent SINR conditions**

  Mobile terminals are assigned to resource partitions based on the present SINR conditions. The SINR can for example be determined from the last transmission towards a particular mobile terminal, as it was done in [Nec07c]. The advantage of this method is the implicit consideration of fading and load variations. The major disadvantage is the strong variation of SINR measurements on short time scales. In particular, if beamforming antennas are used, the interference from neighboring cells varies heavily from transmission to transmission, leading to strong SINR fluctuations.

For LTE systems, Texas Instruments proposed to use the Channel Quality Indicator (CQI) for the same purpose [R1-060369, Pap07]. The CQI is a measurement value reported by the mobile terminal to its serving base station. It indicates the present channel quality and is already used for AMC or channel aware scheduling. [TS 36.213] defines the CQI index in terms of the achievable BLER and requires a mobile terminal to always report a CQI value "for which the transport block error probability would not exceed 0.1". It is thereby vendor specific how the CQI value is actually determined [Kam08]. One option are for example measurements on pilot symbols as described in [Pap07].

#### *4.5.2.4 Interaction with scheduling*

Reuse partitioning has a direct impact on scheduling since the scheduler must take into account the respective resource partition. Above all, the performance gain of frequency selective scheduling is reduced (see Section 2.5.2), since only a limited portion of the total frequency spectrum is available for scheduling certain particular mobile terminals. This leads to a tradeoff between fast frequency selective scheduling and interference coordination [R1-060135]. Alternatively, and to improve the performance in combination with frequency selective scheduling, [R1-070284] and [R1-072840] propose *Fractional Time Reuse*, where the air interface resources are not partitioned in the frequency domain, but in the time domain. This allows

to achieve the full gain of frequency-selective scheduling for all mobile terminals, since the full bandwidth is available in every resource partition. The major disadvantage of this solution is that it requires synchronous networks (see Section 2.2). [R1-072974] additionally investigates QoS aware schedulers that allow adjusting a certain minimum bit rate per traffic flow.

#### *4.5.2.5 Variations*

Further variations with more sophisticated partitioning concepts for the available frequency spectrum have been proposed. One example are the already mentioned overflow techniques [Whi85, Sal88]. In [R1-050896], Qualcomm Europe defined partially overlapping frequency bands as illustrated in Fig. 4.5, with $B$ being the full available spectrum, and $F_1$, $F_2$, and $F_3$ being three frequency bands, each of which is associated with one cell sector. Hence, the basis of this arrangement is a frequency reuse factor of 3. Mobile terminals in a particular cell sector are allowed to use the full available frequency band $B$ except for one of the frequency bands $F_1$, $F_2$, or $F_3$. In a cell sector associated with $F_1$, the usable frequency band is $B \setminus F_1$. If a mobile terminal comes close to a neighboring cell sector, for example a cell sector associated with $F_2$, the frequency usage is further restricted to $(B \setminus F_1) \cap F_2$. In the worst case, the usage may be restricted to $(B \setminus F_1) \cap F_2 \cap F_3$. This restriction is based on the potential geometry gain for a particular mobile terminal, showing a significant increase in the cell edge throughput at the cost of a decreased aggregate throughput. Furthermore, a scheduler is developed, which exploits the differences of the channel conditions in the above defined frequency bands, leading to a reduction of the aggregate throughput penalty.

In [R1-050407], Alcatel proposed a variant based on soft frequency reuse, where in every cell sector a certain portion of the frequency spectrum is used at a reduced power. The total available frequency band is divided into $s$ subsets. While all cell sectors may principally utilize the whole frequency band, each cell sector is assigned one of the $s$ frequency bands, which it may use only at reduced power. For example, $s$ can be set to 7, as illustrated in Fig. 4.6, where the subsets are then assigned to the cell sectors analog to a classical frequency reuse pattern. As the mobile terminal in Fig. 4.6 moves towards the cell edge of cell sector B, it measures as

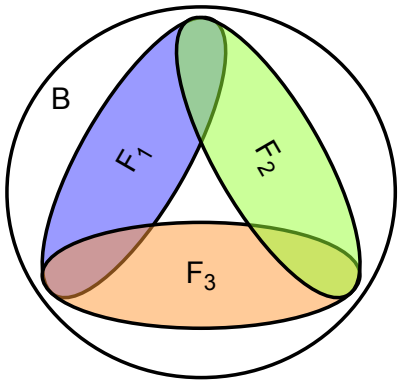



**Figure 4.5:** Partially overlapping frequency bands according to [R1-050896].

strongest interferer cell sector A, which it reports back to its serving base station. Subsequently, the mobile terminal is served on the frequency sub-band that is used only at reduced power in cell sector A while it is located close to cell sector A. In [R1-060209, R1-061032, R1-061455], the performance of this proposal was evaluated with a 10 dB power reduction in the soft reuse areas. It was demonstrated, that the average throughput can be increased by up to 37% while maintaining the same cell edge throughput. For the uplink, a similar scheme was proposed in [R1-050593] and [R1-072921].

#### *4.5.2.6 Optimal Static Planning*

A generalization of FFR as described in the Section before was studied by Bonald et al. in [Bon05, Bon06]. The authors define arbitrary geographic regions within a cell area, where in each region a terminal is served with a certain transmission profile. A transmission profile thereby corresponds to a particular combination of active transceivers. For example, for a network with three cell sectors, there are 7 different transmission profiles, namely three with only one active transceiver, three with two simultaneously active transceivers, and one with all transceivers active. For every transmission profile and geographic position $\vec{x}$, a feasible transmission rate is specified. This allows finding optimal boundaries for geographic regions served with a particular transmission profile. The authors present numerical examples for a two-cell and a three-cell network, and for infinitely large linear and hexagonal networks. Liu and Virtamo present a very similar study in [Liu06], taking further into account an arbitrary geographic traffic distribution.

Kaschub investigated a sub-optimal division of the cell sectors into sub-areas [Kas08]. The frequency assignment to these areas was formulated as an optimization problem, which was subsequently solved by a simulated annealing process. A uniform spatial terminal distribution was assumed. However, the calculations can easily be extended to non-uniform spatial terminal distributions. In combination with soft frequency reuse, the proposed static planning scheme achieved a performance gain of approximately 15% with respect to the aggregate throughput while maintaining the same cell edge throughput.

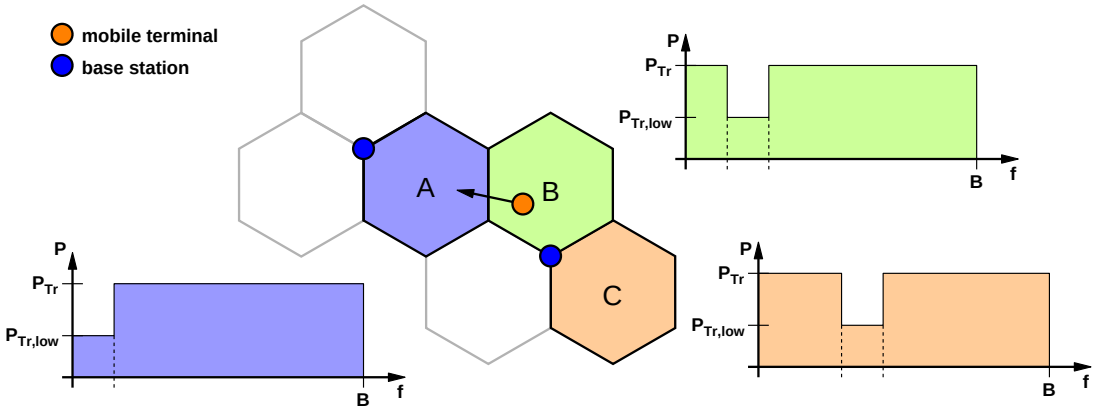



**Figure 4.6:** IFCO based on static network power planning according to [R1-050407].

### 4.5.3 Implicit local interference coordination

In an implicit coordination scheme, every cell sector locally runs a scheduling and resource assignment algorithm, which achieves coordination by implicit knowledge of the scheduling and resource assignment in neighboring cell sectors. No communication among the base stations is required for coordination purposes, and coordination is solely achieved by local decisions. As an important constraint, it is usually required to deploy the same scheduling and resource algorithm in all of the cell sectors in a certain neighborhood. In the following, the most interesting implicit schemes found in literature will be highlighted.

#### *4.5.3.1 Implicit coordination based on resource allocation priority*

In [R1-050833], LG Electronics proposed an implicit coordination scheme for the downlink, which is derived from FFR. The basic principle is illustrated in Fig. 4.7. As in FFR, the available frequency spectrum is subdivided into three pre-defined sub-bands. In the given example, all sub-bands were chosen equally wide with bandwidth $B_{reuse3} = 1/3 \cdot B$. In every cell sector, the sub-bands are ordered by assigning a priority such that one particular frequency sub-band has the highest (or the second highest) priority in only one of the three cell sectors of a base station site. In every cell sector, the mobile terminals are assigned to the different sub-bands based on their required transmit power. This is done by sorting all terminals according to their required transmit power and then assigning the terminals with the highest required transmit power to the priority 1 sub-bands, and assigning the terminals with the lowest required transmit power to the priority 3 sub-bands. This approximately corresponds to assigning the mobile terminals to the sub-bands according to their distance, if fading effects are not too strong. In any case, it is very unlikely that cell edge terminals of neighboring cell sectors will be assigned resources in the same frequency sub-band. Instead, it is likely that a cell center terminal, which is served at low power, will share the same frequency resource with a cell edge terminal, which is served at high power. This achieves an improved SIR compared to random resource assignment, and it allows using the full available frequency spectrum in all cell sectors.

The opposite is proposed by Xiao et al. in [Xia06, R1-060401] for the uplink direction. They propose to align terminals with similar channel quality across the cell sectors, i.e., to assign the same resources to terminals with similar channel conditions. They claim that this alignment will improve the cell edge performance, while the alignment according to different channel qualities will increase the overall sector throughput. Their proposed scheme is therefore somewhat sim-

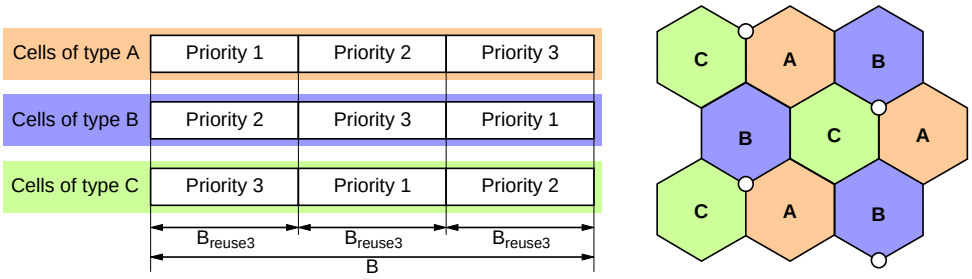



**Figure 4.7:** Implicit local interference coordination according to [R1-050833].

ilar to Autonomous Reuse Partitioning (ARP) proposed by Kanai for circuit switched systems [Kan92].

Similar to this, Kiani et al. propose *Power Matched Scheduling* (PMS) as an optimal scheduling policy, where terminals are ranked in decreasing order of channel gains in every cell sector, and the best $K$ users are assigned to the $K$ available transmission slots per frame [Kia06]. Fairness can be achieved by dividing the frame into as many slots as there are mobile terminals. This allows to trade off fairness against capacity, but it is more difficult to combine the scheme with arbitrary scheduling algorithms. The authors derive this scheduling policy based on a theoretical analysis in an *interference ideal* network, which they define as a network where, for every terminal, the total received interference is independent of the terminal's location. The authors claim that this is a good model for a reuse 1 network with a large number of cells. However, according to [Nec07e], this is questionable and only valid for a reuse 3 network. Nevertheless, the authors achieve capacity gains of more than 100% for large networks, with decreasing gains for smaller networks.

In [Kia07], Kiani et al. extend their work and formulate the scheduling and power allocation problem in each cell as an optimization problem that maximizes network capacity. The same interference ideal network model as in [Kia06] was assumed, and the authors derive a local interference coordination algorithm in which users are scheduled if their SIR exceeds a certain value. The authors achieve significant gains for very small numbers of mobile terminals per cell, with diminishing gains as the number of terminals increases. Fairness is not directly taken into account.

In [Kim07], Kim et al. classify subchannels according to how often they are used in the three cell sectors of a base station during one scheduling period. For a three-sectorized base station, this results in four classes with subchannels being used in no, in one, in two, or in all three cell sectors. The authors study the possibility of maximizing the number of subchannels in a particular class. However, only marginal performance gains are achieved. Moreover, the results in [Kim07] do not reveal the error margins and also do not consider fairness issues but only the aggregate system throughput.

#### *4.5.3.2 Frequency selective scheduling*

With frequency selective scheduling, every mobile terminal is served on its most favorable frequency range. This can be considered an implicit interference coordination mechanism. Since the fast fading processes from one mobile terminal to all its surrounding base stations are independent, and due to the immense signal variations of up to 30 dB due to fast fading, the SIR is greatly improved on average. This allows communication at the cell border even in a scenario with frequency reuse 1.

A further extension is given in [R1-070040], where Motorola proposes a scheme based on soft frequency reuse, in which first a frequency-selective scheduling takes place, followed by a reduction of transmit power for all terminals in high SINR conditions. This power is then redistributed among the rest of the scheduled terminals, and a link re-adaptation is performed. This reduces interference to neighboring cells, increasing the cell edge throughput by $15-25\%$ while providing even a small gain with respect to the overall spectral efficiency in some cases.

### 4.5.4 Explicit local interference coordination

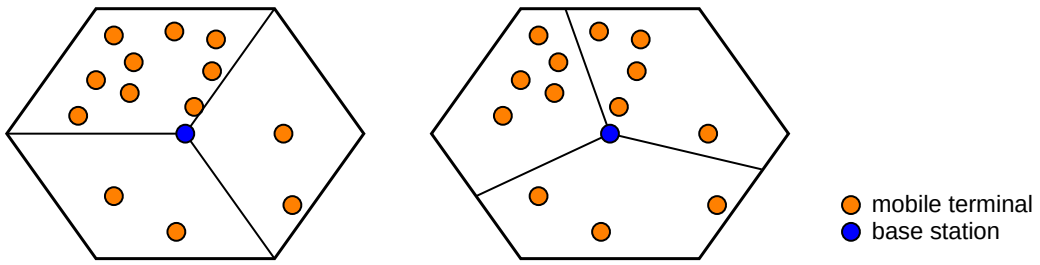
Explicit local interference coordination takes direct measures to avoid interference. Section 4.5.4.1 presents the possibility of adaptive cell sectorization, followed by a discussion of local IFCO schemes in Section 4.5.4.2. Subsequently, Section 4.5.4.3 treats local coordination schemes for the uplink direction.

#### *4.5.4.1 Smart and adaptive cell sectorization*

Conventional cell sectorization as described in Section 2.2 introduces fixed cell sectors in the system. The sector boundaries are static and determined during the network planning process based on the expected traffic load within a certain area. An adaptive cell sectorization technique that takes into account the load distribution in the cell was proposed by Saraydar et al. in [Sar01] for CDMA systems. This is done in a semi-static fashion for a nomadic scenario. The basic principle is illustrated in Fig. 4.8. In Fig. 4.8(a), a conventional three-sectorized cell with a non-uniform spatial traffic distribution is shown. In Fig. 4.8(b), an optimized cell sectorization for the same scenario is shown, which balances the load between the cell sectors up to a certain degree. The sector adaptation may be done based on different optimization criteria, such as the maximization or minimization of the received or transmitted power. Another proposal for dynamic cell sectorization based on sector rotation and sector resizing is given by Giuliano et al. in [Giu02] for a more dynamic scenario. Furthermore, Savazzi and Favalli [Sav07] employ clustering algorithms in order to achieve a dynamic cell sectorization. Note that in any case the adaptation of the sector borders may be subject to certain constraints, for example to physical constraints from the installed antenna systems.

#### *4.5.4.2 Local inter-cell sector coordination*

As it was discussed earlier, a global coordination of all transmissions in the cellular network is impossible to achieve on a per-frame level. However, it is well possible to coordinate the transmissions in the cell sectors belonging to a particular base station, since no communication among base station sites is required. This minimizes the interference among the transmissions in the cell sectors of one base station and thus potentially increases the system performance.




(a) static cell sectorization (b) adaptive cell sectorization

**Figure 4.8:** Illustration of adaptive cell sectorization compared to static cell sectorization.

The first such proposal was made by Sternad et al. in [Ste03a]. They coordinate the scheduling among cell sectors of the same base station in the following way. If a terminal receives a strong signal from an adjacent sector, that neighboring sector is prevented from using the same resources at the same time. While this allows to control the intra-BS interference, it does not allow to control the interference from neighboring base stations. Therefore, [Ste03a] additionally uses a reuse partitioning approach in order to enhance the performance in the border areas to neighboring base stations. The assignment of mobile terminals to the two resource partitions is done based on the terminal's distance from the base station. Altogether, the authors achieve an effective frequency reuse of $1.5-2$ with an average SIR of 16 dB within the cell. Further details can be found in the corresponding project report [Ste03b].

In [Nec07c], a comprehensive performance evaluation based on a related scheme was performed. Besides a distance-based assignment of terminals to the reuse partitions in FFR, an SINR-based scheme was evaluated. Moreover, the impact of the inter-cell sector coordination was studied, and the performance of the coordinated system was compared to that of a system with only reuse partitioning. It was shown by simulation that local inter-cell sector coordination can increase the aggregate throughput by almost 50%, while the cell edge performance cannot be improved, as expected.

#### *4.5.4.3 Local coordination for uplink transmissions*

In the uplink direction, the mobile terminals are the sources of interference. When looking at the cell edge terminals, the sources of interference are therefore directly located in the problematic cell edge area. This opens up new IFCO possibilities compared to the downlink direction. One example is the scheme proposed by Ericsson in [R1-050763] named *Muting*. In this scheme, the cell edge terminals listen to scheduling commands of neighboring base stations. If a collision with its own resource scheduling is detected, the scheduled transmission is "muted", i.e., the transmission is suppressed, thus instantaneously decreasing inter-cell interference. Results presented in [R1-050763] showed a significant improvement with respect to the 5% throughput quantile at a particular aggregate throughput.

### 4.5.5 Decentralized interference coordination

Alcatel-Lucent proposed a semi-static interference coordination scheme based on a request–grant mechanism in [R1-050407], which was further detailed and evaluated in [R1-073187, R1-081873]. Their proposed semi-static algorithm works on top of the static planning scheme previously discussed in Section 4.5.2.5. It is particularly used for cell load balancing, and for leveling out spatial concentrations of mobile terminals [R1-073187]. If a base station cannot serve all terminals in a cell sector at their required minimum bit rate, then it sends a *request* to one or more interfering base station. These base stations may respond with a *grant*, increasing the capacity in the requesting cell sector. The performance of this scheme was investigated in [R1-073187] in a scenario with one overloaded cell sector. The results showed a significant improvement with respect to the aggregate sector throughput and the cell edge throughput in an overloaded cell sector, while only a minor decrease for the performance of lightly loaded sectors was observed.

In [R1-060905], [R1-061852], and [R1-062150], Nortel proposes Adaptive Fractional Frequency Reuse, a decentralized interference coordination scheme based on FFR with soft frequency reuse. In Adaptive FFR, every cell sector can be served by using one of four different modes, which are illustrated in Fig. 4.9. Mode 1 corresponds to a frequency reuse of 1, i.e., the cell sector may use the full available frequency band at full power. On the other end, Mode 4 corresponds to a frequency reuse of 3. Modes 2 and 3 define a soft-reuse scheme with different power levels $P_{Tr,low2}$ and $P_{Tr,low3}$ for the soft-reuse areas, respectively, where $P_{Tr,low2} > P_{Tr,low3}$ as indicated in Fig. 4.9.

Every cell sector begins by using Mode 1 until a coverage problem is detected. In this case, the cell sector switches to Mode 2 and signals to all its neighbor cell sectors to not use Mode 1. The neighbor cell sectors are then restricted to Modes $2-4$. Similarly, if the coverage problem persists, the cell sector will switch to Modes 3 or 4 and signal to its neighbor cell sectors to not use a more aggressive transmit mode. In case the coverage problem no longer exists, the cell sector signals its desired new mode to all of its neighbor cell sectors, but can only switch to it if none of its neighbors has requested a less aggressive mode. In [R1-062150], it is proposed to switch to a less aggressive mode whenever the number of mobile terminals with a geometry factor below a threshold $G_{min}$ rises above a certain number $N_{max}$. The list of neighbors is generated dynamically depending on the strongest interfering cell sectors reported by mobile terminals with coverage problems [R1-074641]. The authors show by simulation that the distributed algorithm converges within a few time steps. Compared to a reuse 3 scenario, performance results show an increase in the aggregate sector throughput by more than 35% with a decrease in the cell edge throughput of about 30%. Compared to a reuse 1 scenario, the cell edge throughput can be doubled while sacrificing only $5-10\%$ on aggregate throughput performance.

In [R1-070442], Qualcomm Europe proposes a distributed interference coordination scheme for the uplink. Every cell periodically broadcasts a single-bit uplink load indicator [R1-063379], which indicates whether the cell is overloaded or not. Each mobile terminal then decodes the load indicator from at least one dominant interfering cell and subsequently increases or decreases its maximum transmit power when appropriate. Load indicators are broadcast frequently, namely once every 10 ms, and should be decodable even with a very low SINR (e.g., $-20$ dB as claimed by Motorola in [R1-070043]). [R1-070442] reports a significant increase in the cell edge throughput with a slight decrease of the overall spectral efficiency. Alternatively,

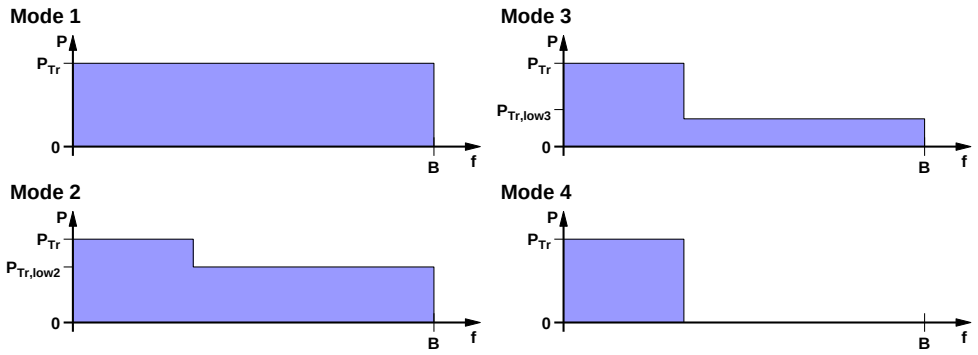



**Figure 4.9:** Illustration of transmission modes in Adaptive FFR.

in [R1-061716] Motorola proposes to exchange load and other information between the base stations via the fixed backhaul network. [R1-070043] further details this proposal by adapting the parameter $\alpha$ of the fractional power control proposed in [Xia06, R1-060401] based on these measurements every 50 ms. The authors report a similar performance compared to the load indicator based approach from [R1-070442] while claiming a much lower complexity for their approach. In response, Qualcomm Europe shows in [R1-070667] a significant dependency of the latter approach on signaling delays, revealing performance losses with respect to the overall spectral efficiency even at medium signaling delays up to 90 ms.

In [Ahm06], Ahmed et al. discuss another dynamic decentralized interference coordination scheme. However, the authors only describe the basic framework without mentioning any algorithmic details. The achieved performance gains are marginal with respect to the aggregate throughput, while a significant increase in packet delay was observed.

### 4.5.6 Distributed interference coordination

In [Rah08b], Rahman and Yanikomeroglu propose a dynamic scheme that uses a central coordinator for interference coordination among cells. The OFDMA frame is divided into resource units (chunks in [Rah08b]), constituting the resource assignment granularity for mobile terminals. First, every base station evaluates the interference situation for all of its served mobile terminals based on pilot measurements. Next, the algorithm iterates over all mobile terminals, evaluating the interference situation in every chunk. Up to two other cell sectors can then be selected for being blocked from transmitting on the same chunk, based on the expected throughput under the given SINR conditions without and with blocking. Subsequently, a utility value is computed for every mobile terminal and chunk, where higher data rates and fewer blocked chunks result in higher utility values. Based on the resulting utility matrix and the traffic demands, the base stations assign each chunk tentatively to a mobile terminal. This results in a restriction request list containing other cell sectors that should be blocked in every chunk. The base stations then forward this list to the central coordinator, which resolves the various blocking requests with a central optimization problem. Finally, the chunk allocations are signaled to the base stations. The communication interval with the central coordinator is correlated with the channel coherence time. In [Rah08b], the communication takes place every 10 chunks, corresponding to about 3.5 ms. Signaling was assumed to be done in zero time, i.e., the result of the central coordination is available at the base stations instantly [Rah08a]. In a full buffer scenario, the authors achieved a performance gain of up to 20% for cell edge terminals.

A distributed architecture was also proposed by Li and Liu in [Li03]. The foundation of their approach is the knowledge of each terminal's achievable data rate with and without strongest interferer. Based on this information, an optimization problem is solved in the central coordinator that assigns channels to all mobile terminals within its coordination realm. The objective function is the maximization of the aggregate system throughput. The channel assignment is communicated to the base stations, which may reassign channels of mobile terminals with empty buffers according to a second, local optimization problem. Local scheduling takes place for every slot, whereas global coordination takes place only for every superframe. The authors study the impact of the superframe duration depending on the channel's maximum Doppler shift, revealing a decrease in performance as the superframe duration or the maximum Doppler

shift increase. Furthermore, the authors study the aggregate system throughput depending on the loss probability of data bits in a finite input buffer of size 2 kBytes, with a bursty arrival of data chunks. The authors observe a significant gain in throughput for every considered loss probability at the input buffer. Unfortunately, it does not become clear how much of this performance gain is owed to an increase in offered traffic as the throughput results increase (keeping the input buffer loss probability constant). Finally, fairness issues are not considered. An approach that explicitly takes into account fairness based on a similar separation of optimization problems was presented by Koutsimanis and Fodor in [Kou08].

In [R1-051059, R1-060369] and related documents, Texas Instruments proposes a variation of FFR, where the size of the reuse partitions $B_{reuse3}$ varies between the different cells and is adapted by a central coordinator in a semi-static manner. Furthermore, it is proposed to extend the central allocation of frequency sub-bands to the allocation of resources in the time domain, thus offering a greater flexibility for resource allocation. With respect to the time scale of operation, [R1-060369] requires the possibility to reconfigure the resource partitions "at most every several seconds". A performance evaluation of the proposed algorithm is given in [R1-060368] showing significant performance gains with respect to both the cell edge and the aggregate performance.

A slightly enhanced concept was proposed by Mitsubishi Electric in [R1-081910]. The authors propose to combine a fixed and an adaptive soft frequency reuse scheme, where only the adaptive portion of the frequency spectrum may be traded among cell sectors. According to [R1-081910], this approach avoids greedy cells, which monopolize frequency spectrum that may be used at full power. Moreover, the approach reduces the signaling overhead and allows to more quickly react on traffic variations.

### 4.5.7 Global interference coordination

Das et al. discuss two different possibilities for globally coordinated scheduling in a time-slotted CDMA system [Das03]. Besides interference coordination, the authors also emphasize the aspect of load balancing that can be achieved by moving terminals from one cell to another. Both is achieved by dividing the network into clusters of base stations, where each cluster is served by a central scheduler. In the first discussed algorithm, the central scheduler performs per-frame scheduling decisions for all base stations in its cluster. This is done by solving an optimization problem that maximizes the achievable throughput under the mobile terminal's channel conditions towards all base stations in a cluster. The throughput is thereby weighted, which allows considering QoS constraints. As an example, a token-based approach is used to achieve fairness among all terminals. The optimization determines which base station is active in which time slot and which mobile terminal it serves. The resulting combinatorial optimization problem in general requires $2^{N_{cells}N}$ possible combinations of base stations and terminals to be considered, where $N_{cells}$ is the number of cell sectors in the cluster and $N$ the number of mobile terminals per cell sector. However, if sector antennas are used and if every mobile terminal is served with the same transmit power, this reduces to $2^N$ combinations. Still, all required channel information needs to be communicated to the central coordinator every frame. In order to reduce the signaling load, an alternative is considered in [Das03] that bases its decision on long-term average channel information. The central scheduler thereby determines

the active base station set for each time slot, and the base stations select appropriate terminals for the slots. This still allows any base station to serve any terminal in every time slot and therefore also allows load balancing among the cells.

The authors compare their algorithms to local scheduling and observed substantial performance improvements with respect to both the aggregate throughput and the cell edge throughput, especially for non-uniform terminal distributions. The given optimization problem directly maximizes the weighted throughput under the constraints given by the scheduling discipline. On the downside, the complexity of the proposed scheme is quite high, and only a small number of base stations per cluster is considered. Moreover, the optimization problem in [Das03] cannot be reduced in complexity as described above if beamforming antennas or variable transmit power is used. In the following Section 4.6, a global scheme based on an interference graph is proposed that overcomes these drawbacks by creating an interference graph with simple heuristic means.

## 4.6 Graph-based Global Interference Coordination

Graph theory has a long tradition in the application to scheduling and resource assignment problems. Examples include routing in wide area networks [Bea97], scheduling in ATM switches [And93], and frequency and channel assignment in cellular networks [MD79, Siv89, Nar02], but also scheduling in multiprocessor systems [Kwo99], or the problem of register assignment in modern compilers [Cha82]. Most examples base their solution on *graph coloring*, where colors are favorably mapped to resources in the real world. A summary of graph coloring problems together with examples of real world applications is given in [Mar03]. For a more basic introduction, the reader is referred to one of the many textbooks in the area of graph theory (e.g., [Kau71]).

Graph theory in general and graph coloring in particular can also advantageously be applied to interference coordination. This is the focus of the following Sections, in which a novel graph-based scheme for global interference coordination in the downlink of an OFDMA system will be introduced and discussed.

### 4.6.1 Basic principle

Graph-based interference coordination was introduced for cellular FDMA/TDMA networks in [Nec06b]. It is based on an interference graph, whose vertices correspond to the mobile terminals. The edges of the interference graph represent critical interference relations in-between the mobile terminals. If two mobile terminals have such a critical interference relation, they may not be served on the same set of resources, such as the same OFDMA slot. A sample interference graph is shown in Figure 4.10.

Due to the movement of the mobile terminals, the interference relations may constantly change. Therefore, the interference graph has to be updated frequently, ideally in every frame. The resource assignment process is therefore divided into the following three steps, which are performed on a per-frame basis:

1. *Creation of an interference graph:* In this first step, the interference graph is created based on the present signal and interference situation.

2. *Scheduling:* In the second step, scheduling tags are assigned to the individual traffic flows. These tags correspond to priorities and may be the result of an arbitrary scheduling discipline, for example a channel dependent scheduler or a QoS aware scheduler.

3. *Frame packing:* In this third step, MAC frame resources are assigned to the different mobile terminals while taking into account the constraints of the interference graph and the scheduling tags.

The creation of the interference graph and the frame packing can only be done with a global view of the system. For the creation of the interference graph, the amount of interference generated by any base station to any mobile terminal needs to be known. For the frame packing, assigned resources need to be immediately blocked for the use by other mobile terminals in other cells if they are connected in the interference graph. For these reasons, the proposed scheme can only be performed by a global omniscient device, which is able to perform the resource assignment on a per-frame basis. This makes it a fully dynamic global IFCO scheme.

Next, Section 4.6.2 formally describes the creation of the interference graph, followed by a discussion of its properties in Section 4.6.3. The process of frame packing and the integrated operation with an arbitrary scheduling scheme are studied in Sections 4.6.4 and 4.6.5, respectively. Finally, Section 4.6.6 presents an optimized method for the generation of the interference graph, and Section 4.6.7 discusses the extension of the algorithm to the uplink direction. The performance of the algorithm will be studied later in Chapter 7.

### 4.6.2 Creation of the interference graph

The interference graph can be created by many different algorithms. Different algorithms will use different criteria for their decision where to place an edge, and they may pursue different philosophies while creating the graph. In the following, an algorithm will be presented that generates an unweighted interference graph based on a hypothetical worst case analysis in the sense that it will assume every mobile terminal to be served in a MAC frame on the same resource and therefore interfere with all other terminals. This assumption does not reflect real life, as at most

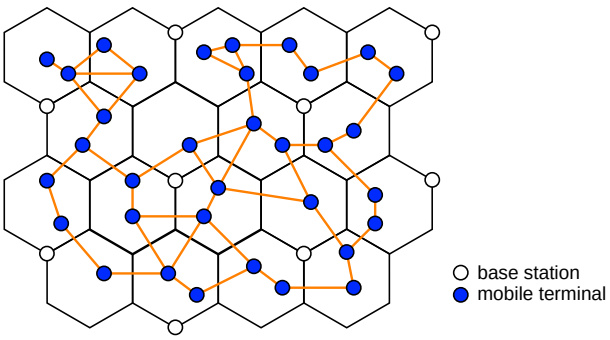



**Figure 4.10:** Example of an interference graph.

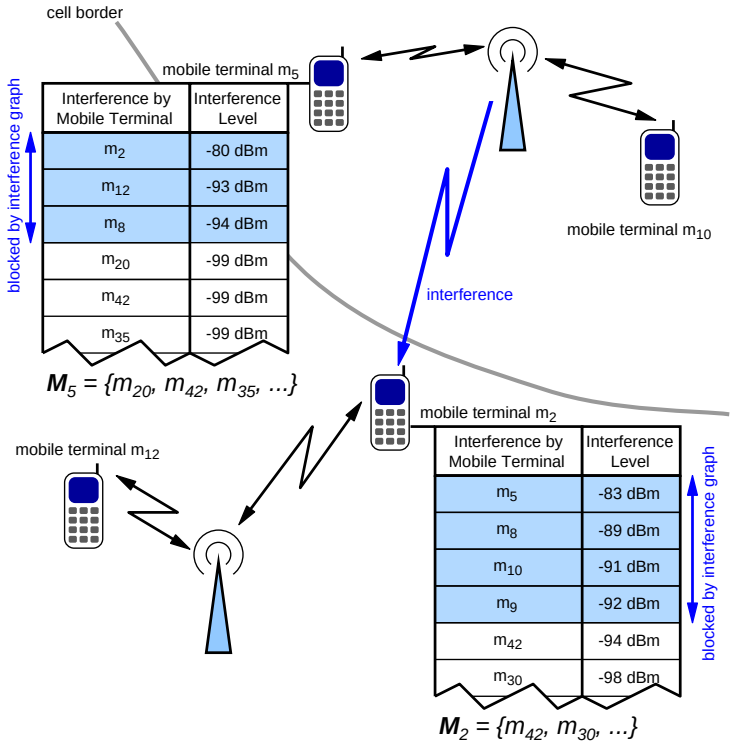



**Figure 4.11:** Illustration of the interference graph calculation with two adjacent cell sectors and four mobile terminals.

one terminal from every cell sector will interfere with another mobile terminal in a particular time/frequency slot. However, it is unknown until the completion of the frame packing step which of the mobile terminals will interfere with each other. The worst case approach therefore identifies potential interference relations before the frame packing is done and is thus one way to separate the interference graph creation from the frame packing as described in Section 4.6.1.

Without loss of generality it will be assumed in the following that there is exactly one traffic flow per mobile terminal. Multiple traffic flows per mobile terminal can easily be accounted for by replicating a mobile terminal at its position as many times as there are traffic flows. The basic principle of the interference graph creation is illustrated with Figure 4.11. For every mobile terminal $m_i$, a list of interferers is created. This list contains the amount of interference that a transmission to a particular other mobile terminal would generate for the respective mobile terminal $m_i$, as shown for $m_2$ and $m_5$ in Figure 4.11. All lists are sorted by the strength of the interferer. Subsequently, the strongest interferers are blocked by introducing an edge in the interference graph such that a desired minimum SIR $D_S$ is achieved.

Let $m_k$ and $m_l$ be two mobile terminals in different cell sectors, as illustrated in Fig. 4.12. $tr_i$ denotes the transceiver serving cell sector $i$. $p_{ik}$ describes the path loss from transceiver $tr_i$ to terminal $m_k$, including shadowing. Furthermore, the function $G_i(l,k)$ describes the gain of the sector $i$ beamforming antenna towards terminal $m_k$ when the array is directed towards terminal $m_l$. $e_{kl} \in \{0,1\}$ are the elements of the interference graph's adjacency matrix $\boldsymbol{E}$, indicating an interference relation between terminals $m_k$ and $m_l$ if $e_{kl} = 1$.

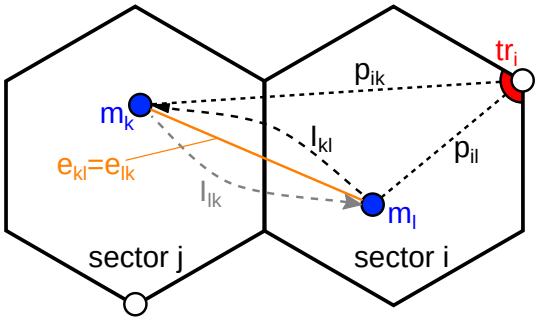



**Figure 4.12:** Path loss, interference, and desired signal.

The above definitions allow the calculation of the received signal strength $S_k$ of terminal $m_k$ if it is served:

$$S_k = p_{jk} G_j(k,k) P_k \ , \tag{4.6}$$

where $P_k$ is the transmit power of transceiver $j$ towards terminal $m_k$. Alike, the interference $I_{kl}$ that a transmission to mobile $m_l$ in sector $i$ would cause to mobile $m_k$ in sector $j$ is calculated as follows, where $i \neq j$:

$$I_{kl} = p_{ik} G_i(l,k) P_l \ . \tag{4.7}$$

For each terminal $m_k$, all interference relations are collected in the set $\boldsymbol{W}_k$:

$$\boldsymbol{W}_k = \{ I_{kl}, \ \forall \ l \neq k, |\vec{c}_l - \vec{c}_k| \leq d_{ic} \cdot d_{BS} \} \ . \tag{4.8}$$

$\vec{c}_l$ are the geographic coordinates of the transceiver which serves the cell sector where terminal $m_l$ is located in. The coordination diameter $d_{ic}$ then denotes the maximum distance that two base stations may have in order to still be coordinated, normalized to the distance $d_{BS}$ between two base stations. In the following, the term $n$-tier coordination will refer to the case of $d_{ic}$ being equal to $n$ times the distance between two base stations.

Subsequently, the largest interferers are being removed from $\boldsymbol{W}_k$ until the worst case SIR for terminal $m_k$ rises above a given desired SIR threshold $D_S$:

$$\mathrm{SIR}_k = \frac{S_k}{\sum\limits_{I_{kl} \in \boldsymbol{W}_k} I_{kl}} \geq D_S \ . \tag{4.9}$$

This results in a new set $\bar{\boldsymbol{W}}_k$ with non-critical interferers, and the edges $e_{kl}$ of the interference graph then follow as:

$$e_{kl} = \begin{cases} 0 & \text{if } I_{kl} \in \bar{\boldsymbol{W}}_k \wedge I_{lk} \in \bar{\boldsymbol{W}}_l \\ 1 & \text{otherwise} \end{cases} \ . \tag{4.10}$$

Equation (4.10) sets an interference relation $e_{kl}$ if terminal $m_l$ causes interference to terminal $m_k$, or vice versa. This is necessary since in both cases the usage of the same set of resources must be avoided. This results in a non-directional interference graph, i.e., $\boldsymbol{E}$ is symmetric.

Finally, all mobile terminals within a cell sector must be assigned disjoint resources. Hence, $e_{kl} = 1$ if $m_k$ and $m_l$ belong to the same cell sector, which leads to a modification of eq. (4.10):

$$e_{kl} = \begin{cases} 0 & \text{if } I_{kl} \in \bar{\boldsymbol{W}}_k \wedge I_{lk} \in \bar{\boldsymbol{W}}_l \wedge (tr(m_l) \neq tr(m_k)) \\ 1 & \text{otherwise} \end{cases} \ , \tag{4.11}$$

where $tr(m_l)$ denotes the transceiver serving mobile $m_l$.

### 4.6.3 Properties and degrees of freedom

For every mobile terminal $m_k$, the set $\boldsymbol{V}_k$ contains all mobile terminals that must not be assigned the same time / frequency resources:

$$\boldsymbol{V}_k = \{m_l | e_{kl} = 1, l \neq k\} \ . \tag{4.12}$$

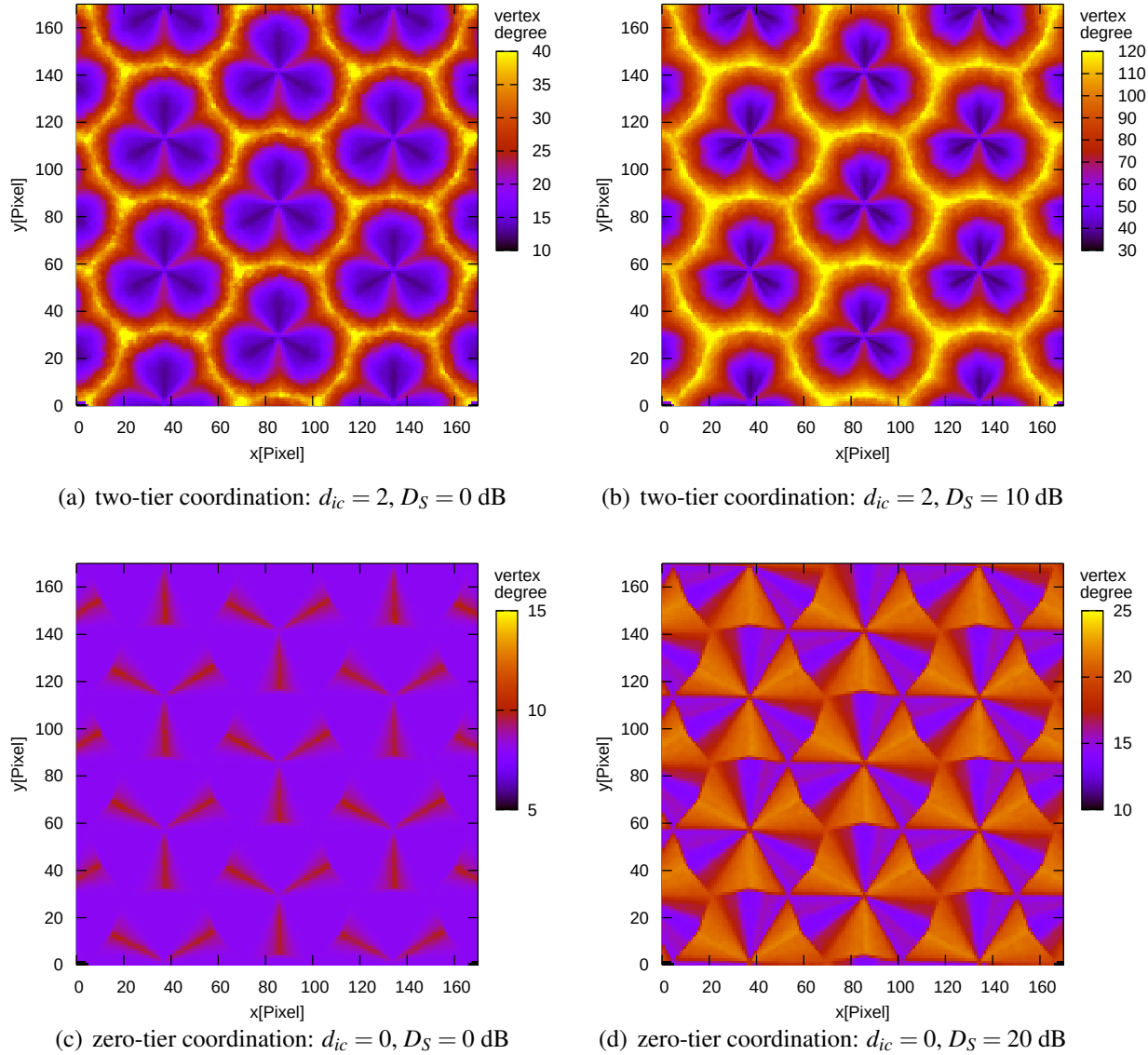
$\boldsymbol{V}_k$ contains those terminals that are in a critical interference relation to $m_k$ and are located in different cell sectors than $m_k$. It additionally contains all terminals within the same cell sector as $m_k$. While the former depend on the parameter choice of the algorithm and the present interference situation, the latter only represent a constant offset depending on the number of mobile terminals per cell sector.

#### *4.6.3.1 Vertex degree*

The cardinality $|\boldsymbol{V}_k|$ of the set $\boldsymbol{V}_k$ is equal to the vertex degree $\deg(m_k)$ of the node $m_k$ in the interference graph. A higher vertex degree therefore indicates a larger amount of critical interference relations, which will make the resource assignment more difficult. $|\boldsymbol{V}_k|$ depends on the position of a mobile terminal and is likely to be large in areas with a high amount of interference. High interference and thus a high vertex degree can especially be found in the cell border areas, i.e., between the sectors of a base station, and in the border areas towards cells of adjacent base station sites. Figure 4.13(a) illustrates a typical distribution of the mean vertex degree within the observation area according to Fig. 3.4(a). Plotted is the mean vertex degree of a mobile terminal if it is located at a particular position. The locations of the base station sites are clearly visible as areas with a low vertex degree. Alike, the three sectors of one site are well distinguished, as they are separated by areas with a particularly high vertex degree.

The remaining graphs of Fig. 4.13 show the impact of the two key parameters during the creation of the interference graph, namely the desired minimum SIR $D_S$ and the coordination diameter $d_{ic}$. The increase of $D_S$ leads to the addition of new edges and thus a higher vertex degree. This is intuitive, but can also be deducted from eq. (4.9). It is further clarified by comparing Fig. 4.13(a) with Fig. 4.13(b), and Fig. 4.13(c) with 4.13(d), showing an increase of the vertex degree at any geographic position.

The increase of the second key parameter $d_{ic}$ not only increases the vertex degree per node. Furthermore, it leads to a fundamental change of the topology of the interference graph. For $d_{ic} = 0$, only mobile terminals served by the same base station can be connected in the interference graph. Therefore, $d_{ic} = 0$ leads to an interference graph that is partitioned into at least as many partitions as there are base stations. In contrast, $d_{ic} > 0$ usually leads to a connected graph for reasonable values of $D_S$. $d_{ic} = 1$ allows edges between terminals served by neighboring base stations, $d_{ic} = 2$ allows edges over two tiers of base stations, and so on. A second effect of increasing $d_{ic}$ is an increased number of mobile terminals that are considered as potential interferers. Hence, the vertex degree increases with a larger coordination diameter $d_{ic}$. In particular, it increases in the border areas between cell sectors of different base stations due to the increased interference from other base station sites. These considerations are emphasized when comparing Fig. 4.13(c) with Fig. 4.13(a).




(a) two-tier coordination: $d_{ic} = 2$, $D_S = 0$ dB

(b) two-tier coordination: $d_{ic} = 2$, $D_S = 10$ dB

(c) zero-tier coordination: $d_{ic} = 0$, $D_S = 0$ dB

(d) zero-tier coordination: $d_{ic} = 0$, $D_S = 20$ dB

**Figure 4.13:** Mean vertex degree of a mobile terminal in the interference graph. Note the different scales.

### *4.6.3.2 Clique size and clustering coefficient*

Besides the vertex degree itself, the distribution of edges is an important property of the graph. This was already addressed in an illustrative way in the previous Section by the geographic distribution of the vertex degree in Fig. 4.13. In the following paragraphs, a more detailed discussion will be given. One of the most important characteristics with respect to the distribution of edges is the number of cliques and their sizes. The occurrence of cliques plays a major role during the resource assignment step, since all terminals that are part of a clique must be assigned disjoint resources.

First, eq. (4.11) makes it obvious that all terminals within one cell sector form a clique. Assuming an equal number of mobile terminals $N$ per cell sector, the minimum clique size is thus

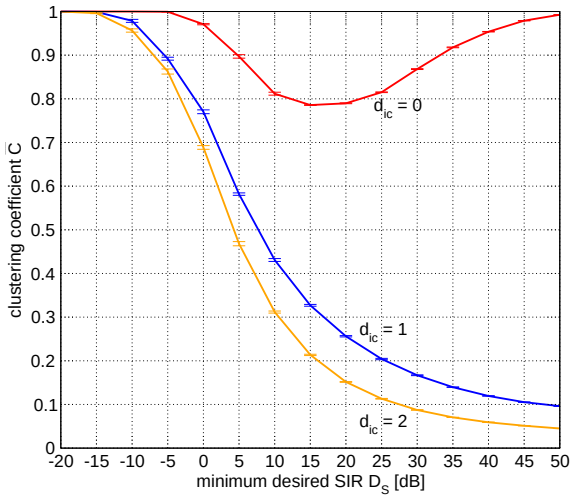



**Figure 4.14:** Clustering coefficient $\bar{C}$ depending on $D_S$ for different values of $d_{ic}$, $N = 8$.

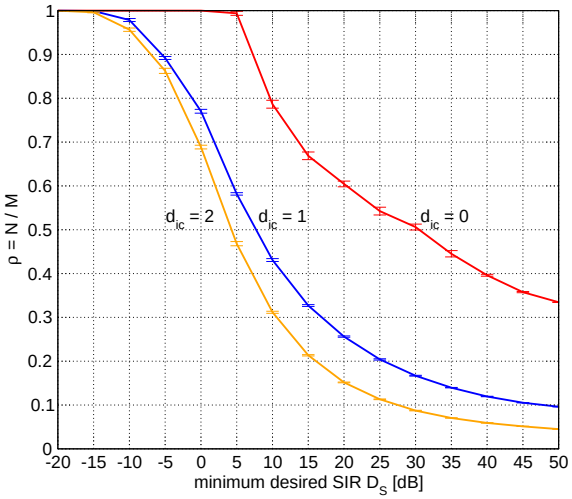



**Figure 4.15:** Resource utilization $\rho = N/M$ for different $d_{ic}$, graph coloring with Dsatur, $N = 8$.

equal to $N$. A graph where no other edges but the edges within the cliques exist is obtained for small enough values of $D_S$. Second, certain graphs may exhibit further cliques. For example, for $d_{ic} = 0$ and large values of $D_S$, all mobile terminals served by one base station will eventually be connected in the interference graph. Consequently, all terminals of one base station will form a clique of size $3N$. When increasing $D_S$ from very small values, more edges will be added thus moving away from a graph with only cliques. Large values of $D_S$ will lead to a graph consisting of only cliques again. The range of $D_S$ in-between can quantitatively be characterized by the clustering coefficient [Wat98]. For a certain vertex $m_k$, the clustering coefficient $C_k$ indicates how "close" the vertex $m_k$ and its neighborhood $\boldsymbol{N}_k$ are from being a clique. $C_k = 0$ means that there is no edge between any of the vertices in the neighborhood of $m_k$, $C_k = 1$ indicates that the neighborhood of $m_k$ is fully meshed and thus forms a clique. Formally, the clustering coefficient $C_k$ and the average clustering coefficient $\bar{C}$ of the whole graph are defined as follows:

$$C_k = \frac{2|\{e_{jk}\}|}{|\boldsymbol{N}_k| \cdot (|\boldsymbol{N}_k| - 1)}, \quad m_k, m_l \in \boldsymbol{N}_k \ , \tag{4.13}$$

$$\bar{C} = \frac{1}{N_{total}} \sum_{i=0}^{N_{total}-1} C_i \ . \tag{4.14}$$

Figure 4.14 plots $\bar{C}$ over $D_S$ for different $d_{ic}$. For $d_{ic} = 0$ it can readily be seen that $\bar{C} = 1$ for small or large enough $D_S$, just as discussed above. This also holds for $d_{ic} \geq 1$, though for much larger values of $D_S$.

### 4.6.4 Frame packing

After the creation of the interference graph, resources have to be assigned based on the constraints of the interference graph. As all terminals that are connected in the interference graph must not be served by using the same set of resources, the goal is to find the minimum set of

disjoint resources such that this requirement is met. This directly translates to the well-known graph coloring problem, in particular to the vertex coloring problem, in which the interference graph has to be colored such that no two adjacent vertices have the same color. The colors of the vertices correspond to particular resource partitions, for example to disjoint sets of slots in a permutation zone. Each partition can then be used for the transmission towards a mobile terminal.

The number of colors $M$ that are needed for the coloring determines the utilization of transmission resources in the system. If $M$ colors are needed, $M$ disjoint resource partitions have to be formed within a permutation zone. Since there are only $N$ mobile terminals in each cell sector, only $N$ of these resource partitions will eventually be used, which implies a resource utilization of $\rho = N/M$. As explained in Section 4.6.3.2, eq. (4.11) enforces the minimum clique size to be equal or greater than $N$. Consequently, at least $N$ colors are needed and $\rho \leq 1.0$. Most colorings will require more colors, leading to a resource utilization below 100%.

The graph coloring problem is NP complete. An optimum solution of the graph coloring problem uses the smallest possible number of colors. This number is known as the chromatic number $\chi(G)$ of the graph $G$. It is obvious that $M \geq \chi(G)$, where it depends on the coloring algorithm how closely $\chi(G)$ can be achieved. A large number of heuristics have been developed to efficiently determine near-optimal colorings, i.e., colorings whose number of colors are very close to $\chi(G)$. Among these heuristics are adaptations of well-known concepts such as simulated annealing [Joh91], genetic algorithms [Big74, Dav91] or tabu search [Her87]. Since these approaches are computationally expensive, other, much simpler heuristics have been developed. They do not approach the optimum as closely, but produce a valid solution very quickly and are thus more suitable for real-time applications. One example is the heuristic Dsatur proposed by Brélaz in [Bré79]. Dsatur is based on a greedy algorithm where the vertices are colored in the order of their *saturation degree*. The saturation degree of a vertex is defined as the number of different colors to which it is adjacent. Since Dsatur is very fast and still achieves acceptable results, it will be used as the basic coloring algorithm in the remainder of this monograph. As a reference, the original and most popular tabu search technique according to Hertz and de Werra [Her87] is used, which obtains near-optimal colorings, however with significant computational effort.

Figure 4.15 demonstrates the relation between $\rho$ and $D_S$ for different values of $d_{ic}$ when using Dsatur for the graph coloring. When the minimum desired SIR $D_S$ is increased, the average vertex degree of the interference graph increases, as already discussed in Section 4.6.3.1. This leads to an increase in the number of required colors and to a reduction of the resource utilization $\rho$. For the same reason, $\rho$ decreases as the coordination diameter $d_{ic}$ is increased.

Figure 4.16 furthermore illustrates the impact of the number of mobile terminals $N$ per cell sector and the influence of the graph coloring heuristic. Obviously, tabu search achieves a significantly better resource utilization, though at the already mentioned much higher computational cost. Despite these expected results, the relationship between the number of mobile terminals $N$ and the resource utilization $\rho$ is not obvious. If the number of mobile terminals $N$ is increased, an increase in the number of required colors $M$ can be expected. However, this does not allow any general conclusion for the ratio $\rho = N/M$, since the increase of $N$ may compensate for the increase of $M$. In fact, Fig. 4.16 shows that an increase in the number of mobile terminals from $N = 8$ to 12 or 16 leads to no significant changes in the resource utilization.

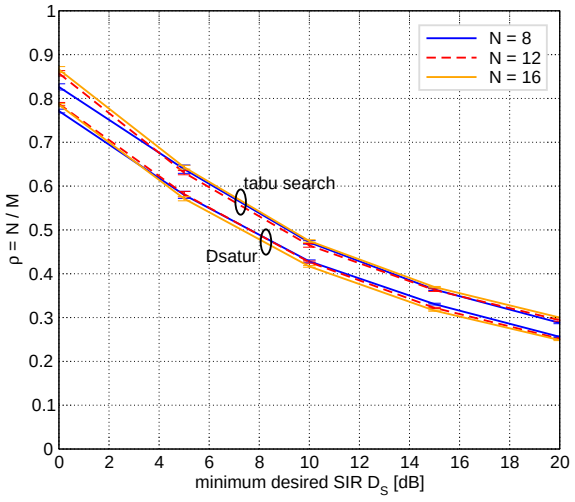



**Figure 4.16:** Resource utilization $\rho = N/M$ for different number of terminals $N$.

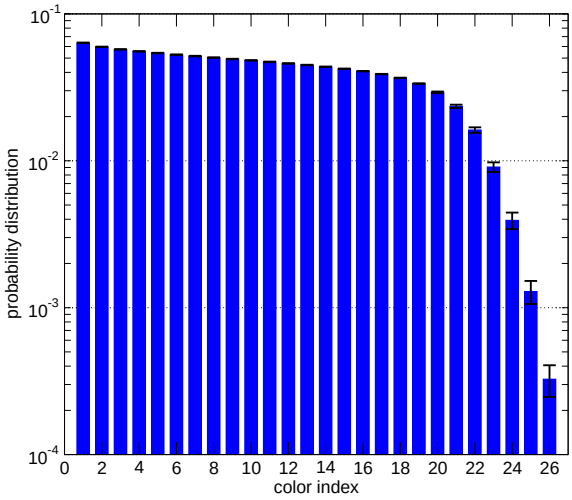



**Figure 4.17:** probability distribution of used color indices, $N = 8$, $D_S = 15$ dB.

As a conclusion, in the considered parameter range, the resource utilization is independent of the number of mobile terminals if graph based resource assignment is used. Moreover, Dsatur delivers no fundamentally different results than tabu search.

### 4.6.5 Resource assignment and scheduling

The assignment of resources by vertex coloring as described in the previous Section has several drawbacks. First, the permutation zone needs to be divided into $M$ resource partitions. Since $M = N/\rho$, it can be seen from Fig. 4.16 that this may lead to a relatively large number of resource partitions and thus to a fragmentation of the permutation zone. This increases the overhead caused by map control information messages, signaling messages in the reverse link direction, and padding. Also, resource partitions have a minimum size since they must be able to host at least one ARQ PDU plus header. Consequently, there is an upper limit for the number of resource partitions, which may be smaller than the required number $M$.

A second drawback becomes obvious from Fig. 4.17. It plots a typical example of the probability distribution of the color indices that are assigned to the vertices in average during the graph coloring. For each individual coloring, the indices are sorted such that lower indices indicate more frequently used colors. In the given example, about 6.4% of all vertices are colored with color index 1, while only about 0.033% of all vertices are colored with color index 26. This demonstrates that the different colors are distributed unequally across the terminals. In particular, a small number of mobile terminals require extra colors in order to solve the coloring problem. This decreases the resource utilization while at the same time it only supplies a small number of additional terminals.

Last but not least, a conventional vertex coloring does not incorporate a service discipline, for example to provide QoS. In fact, it treats all terminals equally such that each terminal is served with an equal amount of time/frequency resources. This eventually corresponds to a

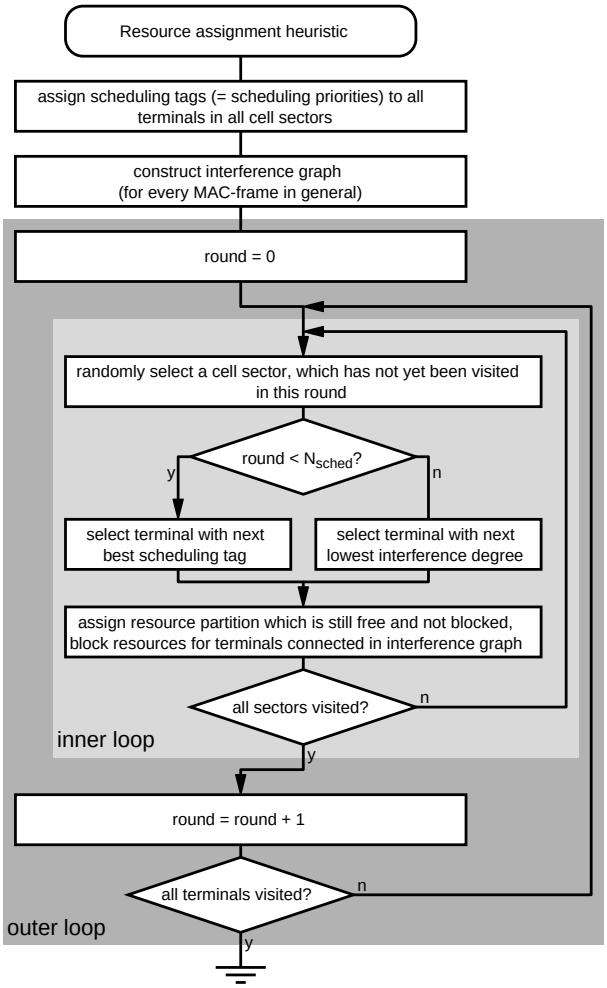



**Figure 4.18:** Heuristic for scheduling and resource assignment in every frame.

Round Robin scheduling of all terminals, and the equal treatment of terminals even in the cell border areas results in a rather low resource utilization. Instead, it is desirable to determine the amount of assigned resources by an arbitrary service discipline, giving way to serving cell edge terminals only as often as needed.

To overcome these problems, Fig. 4.18 presents the flowchart of a resource assignment heuristic, which combines an arbitrary scheduling algorithm with the constraints imposed by the interference graph. In order to avoid frame fragmentation, only a certain number $N_{srvd}$ of mobile terminals will be served in each frame. This allows choosing $N_{srvd}$ terminals out of the $N$ terminals in each cell sector according to a certain criterion. This may either be determined by a particular scheduling algorithm, but it may also be based on which terminals are best suitable to solve the graph coloring problem with a minimum number of colors. One heuristic approach is to serve the terminals with the lowest vertex degree, since they produce the fewest conflicts.

To trade off the constraints from the service discipline and the resource utilization, $N_{sched}$ terminals will be served based on the scheduler's decision, and the remaining $N_{srvd} - N_{sched}$ terminals based on their vertex degree. As illustrated in the flowchart of Fig. 4.18, the algorithm is divided into an inner loop and an outer loop. The inner loop traverses all cell sectors in a random order

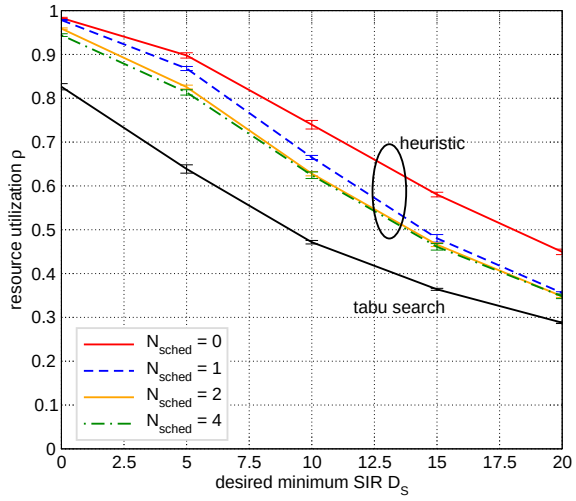



**Figure 4.19:** Resource utilization ρ achieved using the resource assignment heuristic of Fig. 4.18 with different values of $N_{sched}$, and using the graph coloring with tabu search, $d_{ic} = 1$, $N = 8$, $N_{srvd} = 4$.

and assigns resources to one terminal within every sector. The outer loop counts the number of rounds, where one round corresponds to the assignment of resources to one terminal in each cell sector. In the inner loop, the terminal to be served is selected based on the index of the current round. This is done such that in the first $N_{sched}$ rounds the terminal is selected according to the scheduler's decision, and in all remaining rounds the terminal is selected based on its vertex degree. Whenever resources are assigned to a mobile terminal, these resources are blocked for all other mobile terminals that are connected to it in the interference graph. Resources are assigned by dividing the permutation zone into $N_{srvd}$ equally sized resource partitions formed by disjoint sets of slots and by assigning the first free and non-blocked partition to the current terminal. The algorithm terminates once it has tried to assign resources to all mobile terminals.

Figure 4.19 plots the resource utilization ρ over the desired minimum SIR $D_S$ for different values of $N_{sched}$. As an example, a random scheduling discipline as described in Section 6.8 was used. If $N_{sched}$ is set to 0, the service discipline has no influence. Instead, terminals with favorable interference properties will be served preferably. In particular, the system will give preference to terminals with a low vertex degree in the interference graph, which increases the resource utilization at the cost of fairness. As $N_{sched}$ is increased, more terminals are served based on the scheduler's decision. This leads to a decrease in the resource utilization and an increased fairness. In any case, the resource utilization is significantly improved compared to the completely fair full coloring with tabu search, since the latter suffers from the uneven distribution of colors discussed previously in Fig. 4.17. Further performance studies of this resource assignment heuristic including throughput results can be found in [Nec07b].

As mentioned before, the resource assignment needs to be done by a global omniscient entity, where the above resource assignment heuristic is well suitable for. However, note that this only holds if $d_{ic} \geq 1$. In case $d_{ic} = 0$, the presented graph based interference coordination can be done locally in each base station, since only local state information is required for the creation

of the interference graph, and scheduling and resource assignment decisions are restricted to the respective base station. This is similar to the local IFCO scheme proposed by Sternad [Ste03a].

### 4.6.6 Advanced interference graph creation

In order to obtain interference graphs that suit specific requirements, it is possible to merge separately generated interference graphs that resemble different properties. Let $G_1$ and $G_2$ be two interference graphs with adjacency matrices $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$. $G_1$ and $G_2$ can be merged to $G_M$ with adjacency matrix $\boldsymbol{E}_M$ by a logical element-wise or-operation on the elements of $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$:

$$\boldsymbol{E}_M = \boldsymbol{E}_1 \vee \boldsymbol{E}_2 \ . \tag{4.15}$$

This procedure was proposed in [Nec07a] at the example of combining a zero-tier interference graph ($d_{ic} = 0$) and a one- or two-tier interference graph ($d_{ic} = 1$, $d_{ic} = 2$), demonstrating performance improvements over a globally coordinated system based on a single interference graph. These performance aspects will be discussed later in Section 7.2.2.

### 4.6.7 Extension to uplink direction

The interference graph as presented in the previous Sections is tailored to the downlink direction. It is straightforward to adapt it to the uplink direction, as it was done in [Bra07]. However, it should be noted that there are other powerful techniques to improve the system performance in the uplink direction, which are difficult to combine with a graph based approach to interference coordination. Above all, advanced antenna systems based on optimal combiners can efficiently be used in the uplink direction. Interference coordination in combination with such a technique requires a fundamentally different approach, where one example can be found in [Bra07].

## 4.7 Related Work in the Area of Ad Hoc and Mesh Networks

Besides cellular networks, IFCO has been an active research area in multi-hop and mobile ad hoc networking for many years. In [Vil05], Vilzmann et al. consider the possibility of beamforming in a multi-hop wireless network. They first evaluate the general improvement with respect to the interference conditions compared to omnidirectional antennas. Subsequently, they study the requirements for a MAC protocol that is capable of blocking the transmissions of the strongest interferers. In [Ram89], Ramaswami and Parhi coordinate broadcasts in a multi-hop wireless network by means of a sequential graph coloring heuristic.

In [Jai03], Jain et al. consider the coordination of transmissions in a wireless ad-hoc network. The interference conditions are evaluated by an omnipotent central entity with full system state information, which is able to schedule the data transmissions of the individual nodes on the MAC-frame level. This is done based on a *conflict graph*, which represents critical interference relations in-between the network nodes. This conflict graph has the same semantics as the previously introduced interference graph. Hence, the approach in [Jai03] is similar to the graph-based global IFCO scheme of Section 4.6 with respect to the basic idea. However, the conflict

graph is created based on a very simple distance-based interference model (also known as disk model).

The problem of resource assignment based on the conflict graph was traced back to the graph coloring problem for example by Wu et al. in [Wu05]. In [Stu06], Stuedi and Alonso calculate the throughput capacity of a wireless multi-hop network with the help of a very similar schedule graph, which is derived from physical layer properties of the network.

Wei et al. propose an interference-aware routing scheme for wireless mesh networks according to IEEE 802.16 in [Wei05]. The authors define a blocking metric describing the number of network nodes that are affected by a transmission of a certain network node. Packets are then routed through the network based on this blocking metric while trying to achieve as many concurrent transmissions as possible. While this routing scheme is based on a blocking metric determined for the static case, a dynamic interference-aware routing scheme was discussed by in [Tor06]. This dynamic routing scheme is based on an interference-aware Dijkstra algorithm. The edges in the network graph are weighted by the number of required frames to finish a particular transmission under the present interference conditions. Then, a Dijkstra algorithm dynamically determines the optimal route for every packet. The performance was shown to be significantly better than that of a regular static routing scheme and that of the routing scheme based on the blocking metric.

# 5 Distributed Graph-Based Interference Coordination

The previous Chapter introduced the concept of an interference graph for the purpose of interference coordination. This was done based on the prerequisite of global system knowledge and the ability to instantly take global scheduling actions. The interference coordination algorithm presented in Chapter 4.6 is therefore not implementable in a real system. However, it allows for an estimate of the upper performance bound and is thus well suitable as a reference. Moreover, the fundamental considerations on interference graphs in the previous Chapter can provide the basis for further practically implementable interference coordination schemes. In this Chapter, a novel distributed dynamic interference coordination algorithm based on a central coordinator will be developed. Its distributed nature is one aspect that makes it well implementable in actual systems. The algorithm is primarily intended for the downlink direction, but an adaption to the uplink direction is straightforward similar to the discussion in Section 4.6.7. A summary of the proposed algorithm has been published by the author of this monograph in [Nec08b].

Section 5.1 introduces the basic principle of the distributed interference coordination approach and formulates the involved optimization problems. Sections 5.2 and 5.3 present solution approaches to these optimization problems. Subsequently, Section 5.4 discusses how real-time and other QoS requirements can be handled by the algorithm. The coordination of large networks is treated in Section 5.5. Finally, Sections 5.6 and 5.7 discuss aspects of measurement and signaling aspects, respectively.

## 5.1 Basic principle

Section 4.6.5 discussed the possibility of restricting the interference graph to only those mobile terminals served by a particular base station, i.e., $d_{ic} = 0$. In this case, the graph based interference coordination is well implementable, since only state information available locally at every base station is required. Unfortunately, such local coordination has no control over the interference caused by neighboring base station sites, which is why the performance in the border areas towards neighboring base station sites is quite bad (see Section 7.3 and [Nec07c]). To solve this problem, the proposed distributed interference coordination scheme enhances the cell edge performance with the help of global coordination information delivered by a central coordinator. While this central coordination takes place on a relatively long time scale due to the involved signaling and synchronization delays, the local coordination in the base stations can be done frequently on a per-frame basis. Consequently, the proposed scheme is not only

implementable in a real system, but it is also suitable for real-time traffic which requires instant scheduling decisions.

The local coordination is based on solving a local, *inner* optimization problem in every base station taking into account a local interference graph with $d_{ic} = 0$. This inner optimization problem is subject to constraints delivered by the global, *outer* optimization problem, which is solved in the central coordinator based on a global interference graph with $d_{ic} \geq 1$. This is motivated by the advanced interference graph creation described in Section 4.6.6. Since the communication with the coordinator may take place in time intervals in the order of seconds, this is a true distributed scheme with a practical application. The corresponding system architecture is shown in Fig. 5.1. The following Section 5.1.1 formulates the outer optimization problem, which is solved by the central coordinator. Subsequently, Section 5.1.2 formulates the inner optimization problem solved by every base station. Solution approaches to these optimization problems are presented in Sections 5.2 and 5.3.

### 5.1.1 Formulation of outer optimization problem

The purpose of the central coordinator is to achieve coordination among cell sectors belonging to different base station sites, specifically to achieve cell edge coordination. In doing so, the most important constraint is the limited communication with the base stations. It is therefore not feasible to have the central coordinator directly assign resources to the mobile terminals. Instead, it is beneficial that the central coordinator assigns a set of resources to the different mobile terminals, which can then be allocated by the respective base stations during the inner, local optimization on demand. In order to be able to easily communicate the assigned resources from the central coordinator to the base stations, the available resource space is divided into resource partitions, which will later be associated with colors as it was also done in Section 4.6.4. This is illustrated in Fig. 5.2, where the permutation zone of a single frame is divided into $M$ equally large resource partitions. If $M$ is large, this leads to a fragmentation of the permutation zone. This is problematic, since a resource partition must have a certain minimum size due to the minimum size of a data burst in IEEE 802.16e and any other wireless technology. Moreover, smaller resource partitions imply smaller transmission bursts, which comes along with more overhead due to header information and padding, thus reducing the system efficiency.

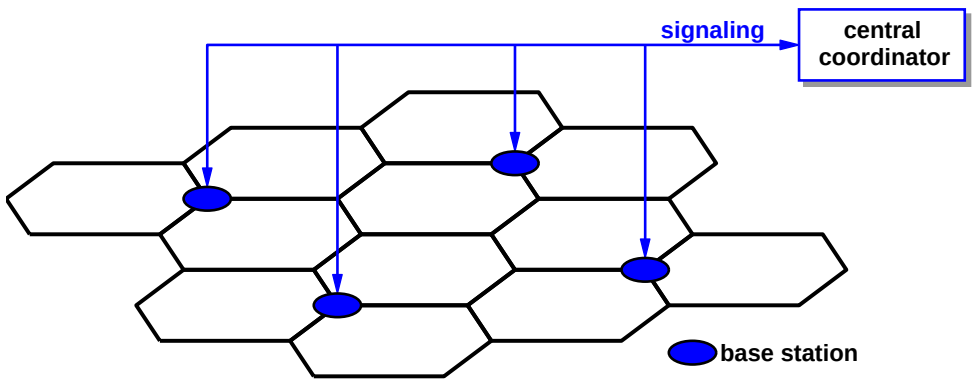



**Figure 5.1:** System architecture for distributed graph-based interference coordination.

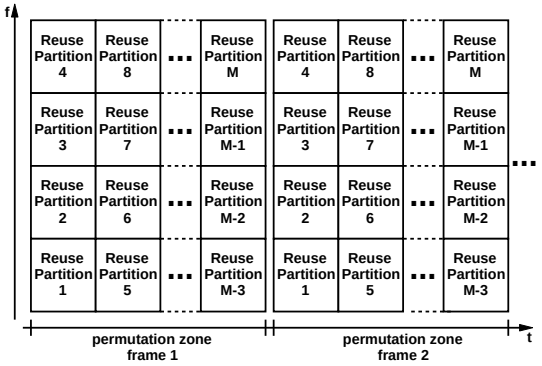



**Figure 5.2:** Resource partitioning within one frame.

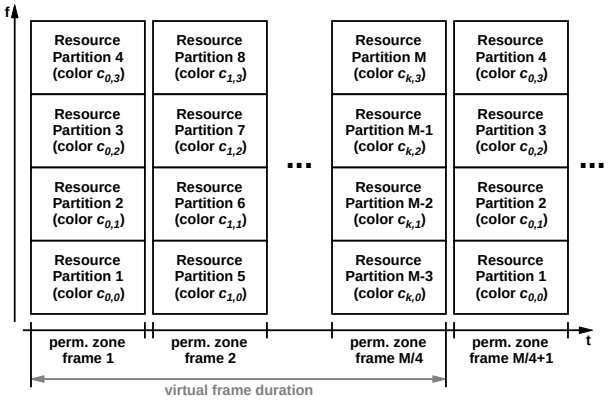



**Figure 5.3:** Resource partitioning with $N_F = 4$ over multiple frames.

In order to avoid this fragmentation, it is advantageous to introduce a minimum partition size and spread the resource partitions over multiple frames, when necessary. This is shown in Fig. 5.3. The number of required frames defines one virtual frame, whose duration may be significantly larger than that of a single frame. This needs to be taken into account for real-time applications, and will be discussed later in Section 5.4. A simplification of the resource assignment process can be achieved by defining a constant resource partition size. This implies a constant number $N_F$ of resource partitions per frame, which may be of equal size for easier handling. The length of one virtual frame then yields to $L_{VF} = \lceil M/N_F \rceil$.

After these general remarks, the following paragraphs discuss the specific coordination problem that needs to be solved in the central coordinator. As known from the previous Chapter 4, cell edge coordination can be achieved by building and coloring an interference graph with coordination diameter $d_{ic} \geq 1$. The graph may be created in the central coordinator based on measurements obtained from the mobile terminals and communicated by the individual base stations, which will be further discussed in Section 5.6. Based on this graph, a set of resource partitions needs to be assigned to each mobile terminal. In other words, the central coordinator needs to find a set of colors $\boldsymbol{C}_i \subseteq \boldsymbol{C}$ (where each color corresponds to a resource partition) for every mobile terminal $m_i$ such that color sets of mobile terminals with critical interference relations are disjoint. With the edges $e_{ij}$ of the interference graph, this condition can be formulated as follows:

$$\forall \left\{ (i,j) \mid e_{ij} = 1, i \neq j \right\} : \boldsymbol{C}_i \cap \boldsymbol{C}_j = \varnothing \ . \tag{5.1}$$

An example for a possible coloring is shown in Fig. 5.4. All resource partitions are indexed with their corresponding color $c_{kl}$, where $k = 0, \ldots, \lceil M/N_F \rceil - 1$ is the index of the physical frame within the virtual frame, and $l = 0, \ldots, N_F - 1$ is the index of the resource partition within the $k$-th physical frame. The colors $c_{kl}$ are also annotated in Fig. 5.3 and define the set $\boldsymbol{C}$ as:

$$\boldsymbol{C} = \{ c_{kl} \mid k \in \{0, \ldots, \lceil M/N_F \rceil - 1\}, \ l \in \{0, \ldots, N_F - 1\} \} \ . \tag{5.2}$$

Note that $\boldsymbol{C}$ dynamically varies over time since it is constantly adapted to the solution of the outer optimization problem depending on the number of required colors $M$. The number of colors per vertex $|\boldsymbol{C}_i|$ and the total number of colors $|\boldsymbol{C}|$ directly depend on each other, since an increase in the total number of colors also allows for more colors per vertex. The choice of

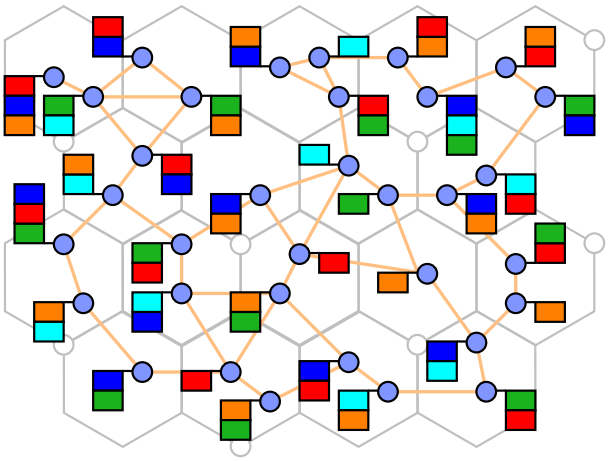


**Figure 5.4:** Graph coloring of outer optimization problem. Flags indicate assigned colors.

the total number of colors $|\boldsymbol{C}|$ is therefore of secondary importance for the overall achievable resource utilization in the system. However, there is another tradeoff when choosing $|\boldsymbol{C}|$. On the one hand, large values of $|\boldsymbol{C}_i|$ are advantageous since they facilitate the inner optimization. On the other hand, large values of $|\boldsymbol{C}|$ and consequently large values of the virtual frame duration $L_{VF}$ make the inner optimization more complex. Moreover, longer virtual frames make it more difficult to handle real-time traffic (see Section 5.4).

The choice of $L_{VF}$ is therefore not obvious. It is useful to try to find a good balance between the number of colors per vertex $|\boldsymbol{C}_i|$ and the virtual frame duration $L_{VF}$. In particular, for a specific virtual frame duration $L_{VF}$, the number of colors per vertex $|\boldsymbol{C}_i|$ should be maximized for every vertex. This makes it difficult to formally specify the objective function and the related constraints. One possibility would be to preset the virtual frame duration $L_{VF}$ and thus the total number of available colors $|\boldsymbol{C}|$ and then try to maximize the overall number of assigned colors:

$$\max \sum_i |\boldsymbol{C}_i| \ . \tag{5.3}$$

This needs to be subject to a constraint, which foresees every node to receive at least $b$ colors (where $b$ usually needs to be much smaller than $|\boldsymbol{C}|$ for the problem to be solvable):

$$\forall i: \ |\boldsymbol{C}_i| \geq b \ . \tag{5.4}$$

An alternative approach to eq. (5.3) is to maximize the minimum of the number of colors assigned to any vertex:

$$\max \min_i |\boldsymbol{C}_i| \ . \tag{5.5}$$

These are only two of the many possibilities. In Section 5.2, a problem specific approach will be presented to find good color sets $|\boldsymbol{C}_i|$ while at the same time adapting the virtual frame duration $L_{VF}$ to the current situation.

### 5.1.2 Formulation of inner optimization problem

The inner optimization problem is solved locally by every base station and coordinates the transmission among the cell sectors of the respective base station. It usually has to be solved once for every virtual frame with the possibility of a corrective action at the beginning of every physical frame if high priority traffic arrives. If the offered traffic does not change from one virtual frame to the next, and if the positions of the mobile terminals do not change significantly, the solution of the inner optimization problem may even be valid for several virtual frames.

The inner optimization problem needs to be based on local state information. Its goal is to assign every mobile terminal to one or more resource partitions $c_{kl}$ of the respective cell sector. This means that every mobile terminal is assigned a set of colors $\boldsymbol{R}_i \subseteq \boldsymbol{C}$ on which it is served. One possible approach is a graph-based coordination with a coordination diameter of $d_{ic} = 0$ as discussed in Section 4.6.5. In contrast to the there discussed solution, the inner optimization problem additionally needs to account for the constraints delivered by the outer optimization problem. That is, $\boldsymbol{R}_i$ must be chosen from the color set $\boldsymbol{C}_i$ assigned to mobile terminal $m_i$ by the central coordinator, hence $\boldsymbol{R}_i \subseteq \boldsymbol{C}_i$. To formulate the optimization problem, a binary matrix is introduced for every mobile terminal that describes the resource allocation for a particular cell sector. For mobile terminal $m_i$, the matrix elements will be denoted as $x_{ikl}$ and defined as follows:

$$x_{ikl} = \begin{cases} 1 & \text{if mobile } m_i \text{ is served in resource partition } c_{kl} \\ 0 & \text{if mobile } m_i \text{ is not served in resource partition } c_{kl} \end{cases} . \tag{5.6}$$

From this follows the definition of $\mathbf{R}_i$:

$$\mathbf{R}_i = \{c_{kl} \mid x_{ikl} = 1\} \ . \tag{5.7}$$

Furthermore, a utility $u_i$ is defined for every mobile terminal $m_i$. $u_i$ is a real number and denotes the utility if the mobile terminal is scheduled in a virtual frame. It is therefore better to schedule mobile terminals with a higher utility $u_i$ more often. $u_i$ can thus be used to prioritize certain mobile terminals.

The goal in every base station is to maximize the sum of the utility of all served mobile terminals under certain constraints, such as fairness. Considering the utility $u_i$, the objective function of the inner optimization problem for base station $b$ then is

$$\max \left( \sum_{m_i \in \boldsymbol{M}_b} \sum_k \sum_l u_i x_{ikl} \right) \ , \tag{5.8}$$

where $\boldsymbol{M}_b$ contains all mobiles $m_i$ which are served by any of the three transceivers of base station $b$. Note that eq. (5.8) maximizes the utility sum for one virtual frame. Hence, for optimal results, the resource allocation problem has to be solved at the beginning of every virtual frame.

The inner optimization problem is subject to a number of constraints.

1. Every mobile $m_i$ has to be served using one of the colors in $\boldsymbol{C}_i$ assigned by the central coordinator:

$$\forall \{x_{ikl} \mid x_{ikl} = 1\} : \ c_{kl} \in \boldsymbol{C}_i \ . \tag{5.9a}$$

This constraint can be formulated directly in terms of the variables $x_{ikl}$:

$$\forall i \forall \{(k,l) \mid c_{kl} \notin \boldsymbol{C}_i\} : x_{ikl} = 0 \ . \tag{5.9b}$$

2. Every mobile terminal has to be served at least once in every virtual frame:

$$\forall i : \sum_k \sum_l x_{ikl} \geq 1 \ . \tag{5.10}$$

3. Every mobile terminal must not be served more than once per MAC frame:

$$\forall i \forall k : \sum_l x_{ikl} \leq 1 \ . \tag{5.11}$$

4. The constraints of the local interference graph have to be met:

$$\forall \{(i,j) \mid e_{ij} = 1\} : \boldsymbol{R}_i \cap \boldsymbol{R}_j = \varnothing \tag{5.12a}$$

This constraint can also be formulated directly in terms of the variables $x_{ikl}$:

$$\forall \{(i,j) \mid e_{ij} = 1\} \forall k \forall l : x_{ikl} + x_{jkl} \leq 1 \tag{5.12b}$$

It is interesting to note that the above optimization problem is related to a Knapsack problem [Mar90]. In particular, the objective function in equation (5.8) has the same form as the objective function of a 0-1 Multiple Knapsack Problem, where the resource partitions are the knapsacks, and the mobile terminals are the items to be placed in the knapsacks. However, the constraints given by equations (5.9)–(5.12) differ from the standard 0-1 Multiple Knapsack Problem. Most important, the items can only be placed in certain knapsacks. Furthermore, an item may be placed in several knapsacks at the same time. Together with the remaining constraints 2 and 4, the inner optimization problem therefore has to be treated as a standard BILP. It was already mentioned in Section 3.6.1 that BILPs are NP-complete to solve. Section 5.3 will present a solution approach for the inner optimization problem based on genetic algorithms. This approach is particularly suitable due to the complex constraints, and it allows a flexible prioritization of constraints by choosing an according fitness function.

### 5.1.3 Solvability and optimal solution

The local interference graph and the available colors per mobile terminal impose the most stringent constraints on the inner optimization. If the local interference graph is a sub-graph of the outer interference graph that was used to obtain the color sets $\boldsymbol{C}_i$ of the mobile terminals, the inner optimization problem is solvable. However, depending on the parametrization, the interference situation, and the signaling delay, this condition may not hold. As a consequence, it may happen that the inner optimization problem is not solvable, even though a proper parametrization will avoid such a situation most of the times. If the problem is not solvable, a sub-optimal

solution has to be used. There exist two different ways for obtaining such a solution. First, constraints 1 and 4 in equations (5.9) and (5.12) may be violated, leading to more or even all mobile terminals being served at the expense of the SINR. Alternatively, constraint 2 in equation (5.10) can be violated, thus maintaining the required SINR at the expense of unserved mobile terminals. The latter possibility will then require a compensation for the unserved terminals in the subsequent virtual frame. This issue will further be discussed in combination with the genetic solution approach in sub-section 5.3.7.

## 5.2 Solution of Outer Optimization Problem

The outer optimization problem from Section 5.1.1 is related to *fractional graph coloring* [Sch97]. In fractional graph coloring, each vertex is assigned a set of $b$ colors such that adjacent vertices are assigned disjoint sets. The graph is *M:b-colorable* if there exists a b-fold coloring where the colors are taken from a palette of $M$ colors in total. Such a coloring is referred to as *M:b-coloring*. A more general variant is *graph multicoloring*, also referred to as *weighted graph coloring*, where each vertex $m_i$ in the graph additionally is assigned a non-negative integer weight $w_i$. Given the weight vector $\vec{w}$ with elements $w_i$, a weighted coloring is characterized by the assignment of $w_i$ colors to every node $m_i$ such that the colors assigned to adjacent nodes are pairwise disjoint. This implies that fractional graph coloring is identical to weighted graph coloring with $w_i = b$ for every $i$. Similar to regular graph coloring, there exists a minimum number of colors required in total for all nodes given a particular weight vector $\vec{w}$. This number is known as the *weighted chromatic number* $\chi_{\vec{w}}(G)$.

The multicoloring problem can be transformed to a regular graph coloring problem in the following way (see also [McD00] or [Lim05]). Given a graph $G$ and a weight vector $\vec{w}$, every vertex $m_i$ is replaced with a complete graph of $w_i$ vertices such that all edges ending at the replaced vertex are replicated $w_i - 1$ times, once for each of the new vertices. This transformation results in the graph $G'$ and is illustrated in Fig. 5.5. When applying regular graph coloring to the new graph $G'$, every vertex in the substituted complete graphs receives one of the colors which are subsequently assigned to the original vertices in $G$. Note that these considerations also imply that multicoloring is NP complete, just like regular graph coloring.

Multicoloring has been applied to various practical problems. In the area of mobile communications, frequency assignment in cellular networks is one of the most prominent examples.

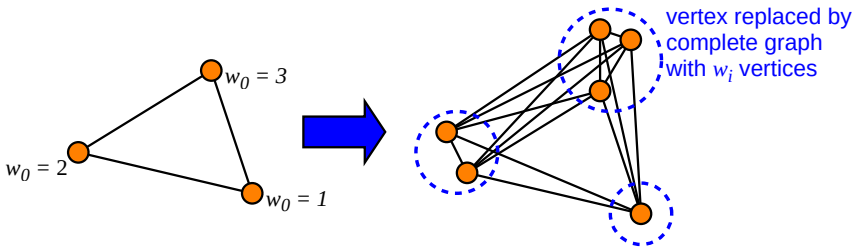



**Figure 5.5:** Transformation of graph with vertex weights $w_i$ for multicoloring by regular graph coloring algorithms.

In this case, the colors correspond to frequencies, and the weights $w_i$ to the demands in terms of number of frequencies in the cell sectors. A good overview of this topic can be found in [Nar02]. The same problem was investigated by McDiarmid and Reed in [McD00], where the authors apply a simple heuristic based on the transformed graph $G'$ to solve the multicoloring problem. A more complex algorithm is proposed by Lim et. al in [Lim05]. The authors combine a *Squeaky Wheel Optimization (SWO)* strategy [Jos99] with a tabu search technique achieving good results.

It is not straightforward to map the multicoloring to the outer optimization problem as described in Section 5.1.1. First, it is difficult to choose the weights $w_i$ appropriately. Of course, it would be possible to choose all weights equal and set them to a predefined number. Alternatively, the weights could be chosen according to the traffic demand of the associated mobile terminal. This would ease the realization of QoS differentiation during the inner optimization. However, choosing the weights in any of these ways does not necessarily lead to an optimal solution of the outer optimization process. The reason for this is the inhomogeneity of the interference graph. Some mobile terminals are in favorable positions, with their corresponding vertices having a small vertex degree. They can easily be assigned a large number of colors $|\boldsymbol{C}_i|$. In contrast, mobile terminals in unfavorable conditions have a large vertex degree, and it is therefore difficult to assign a large number of colors. Consequently, it is advisable to adapt the cardinality of $\boldsymbol{C}_i$ to the position of the associated mobile terminal. In the following, a problem specific solution heuristic based on a standard graph coloring algorithm will be presented.

The proposed solution algorithm tries to find a "natural" length of the virtual frame duration $L_{VF}$ in that it adapts $L_{VF}$ to the present system state. Subsequently, the colors are distributed to the mobile terminals such that terminals in favorable locations are more likely to receive more colors than terminals in unfavorable locations. In order to allow for enough colors per vertex, the adaptation of $L_{VF}$ is based on a dense interference graph, while the following distribution of colors to the vertices is based on a sparser graph. Figure 5.6 shows the flow chart of the proposed algorithm. It consists of the following three phases:

- **Graph preparation**

  During this phase, three different interference graphs are constructed. First, a graph $G_1$ with a coordination diameter of $d_{ic,1} = 0$ and a desired minimum SIR $D_{S,1}$ is created. Afterwards, a second graph $G_2$ with a coordination diameter $d_{ic,2} \geq 1$ and a desired minimum SIR $D_{S,2}$ is created. Finally, both graphs are merged yielding the graph $G_M$. The last step is motivated by the advanced interference graph creation discussed in Section 4.6.6 and produces a graph with a relatively high meshing degree.

- **Initial coloring based on $G_M$**

  This step generates an initial coloring which will be used to fix the virtual frame duration $L_{VF}$. This is done by setting $\boldsymbol{C}$ to the set of colors obtained from this initial coloring. Every identified color is also added to $\boldsymbol{C}_i$ of its corresponding vertex. In order to allow for sufficient extra colors per node, the initial coloring is based on the graph with the highest meshing degree, namely $G_M$, leading to a large set $\boldsymbol{C}$. For this initial coloring, any standard graph coloring algorithm can be used. As discussed earlier, it is not important to reach the minimum number of colors $|\boldsymbol{C}|$. Hence, the relatively simple coloring heuristic Dsatur [Bré79] is well suitable, which is already known from Section 4.6.4. Note that the

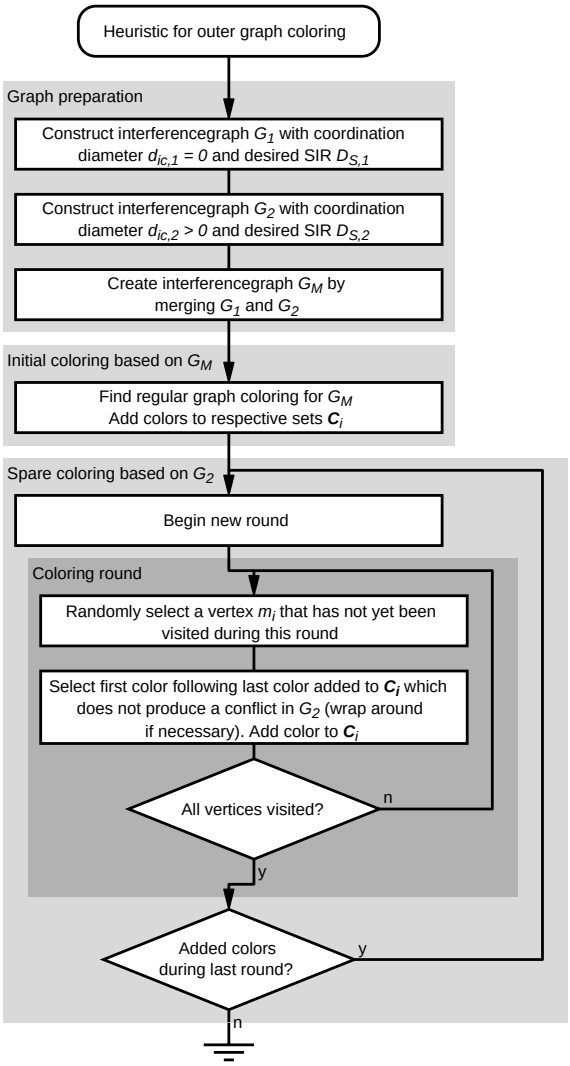



**Figure 5.6:** Heuristic solution approach to outer optimization problem.

number of colors and hence the virtual frame duration is proportional to the number of mobile terminals in the considered parameter range, as it was discussed in Section 4.6.4.

- **Spare coloring based on $G_o$**

  After the initial coloring, extra (spare) colors are assigned to the vertices. All colors in $\boldsymbol{C}$ that were obtained during the initial coloring may be used. To allow for sufficient extra colors per node, the spare coloring is based on a graph with a lower meshing degree compared to the initial coloring. A suitable graph is $G_2$, which is sparser than $G_M$ but still takes care of the coordination among different base stations due to its coordination diameter of $d_{ic,2} \geq 1$. Spare colors are assigned in a greedy manner by randomly selecting vertices and adding the next non-conflicting color starting from the last color added to $\boldsymbol{C}_i$, which may require wrapping around the set $\boldsymbol{C}$. In order to not favor certain vertices, the procedure is organized in rounds during which all vertices are selected in a random order and assigned at most one additional spare color. The algorithm terminates when there was one round during which no more spare colors could be assigned.

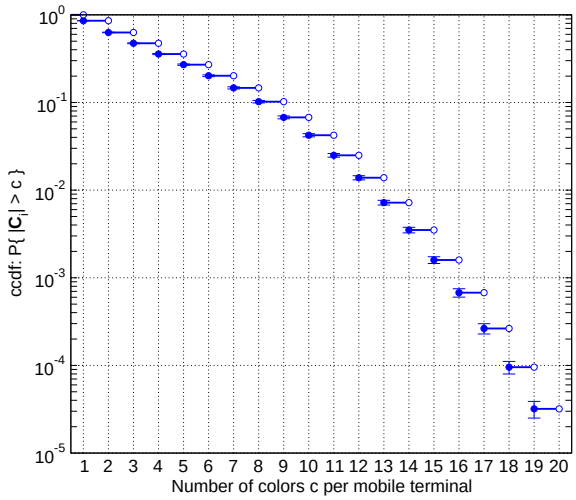



**Figure 5.7:** ccdf of the number of colors per mobile terminal, $d_{ic} = 2$, $D_{S,o} = 5$ dB, $D_{S,i} = 20$ dB.

It is important that every vertex receives a sufficient amount of colors. Fig. 5.7 shows the complementary cumulative distribution function (ccdf) of the number of colors per mobile terminal for a typical configuration as it will be evaluated later on in Chapter 7. In this scenario, only about 14% of all mobile terminals are assigned only one color, whereas almost half of all mobile terminals receive at least 4 colors. 10 or more colors are assigned to almost 7% of all terminals. This gives sufficient degrees of freedom for the solution of the inner optimization problem.

The described algorithm has three adjustable parameters: the desired minimum SIR $D_{S,1}$ of the graph $G_1$, the desired minimum SIR $D_{S,2}$ of the graph $G_2$, and its coordination diameter $d_{ic,2}$. All three together are the adjustable parameters of the merged graph $G_M$ and are responsible for the total number of colors $|\boldsymbol{C}|$ and hence the virtual frame duration $L_{VF}$. The cell edge performance is mainly determined by the parameters of $G_2$, namely $D_{S,2}$ and $d_{ic,2} \geq 1$. In the remainder of this monograph, $D_{S,1}$ will be chosen equal to $D_{S,2}$ and labeled as the desired minimum SIR $D_{S,o}$ of the outer interference graph. Likewise, $d_{ic,2}$ will be labeled $d_{ic,o}$.

The algorithm's complexity is very low. Dsatur itself operates in polynomial time with complexity $O(N^2)$. The same applies to each round of the spare coloring phase, where the number of rounds depends on the number of assignable spare colors. Altogether, the heuristic operates in polynomial time and poses no problems for the implementation in the central coordinator.

## 5.3 Solution of Inner Optimization Problem

The inner optimization problem is a BILP, which is particularly difficult to solve (cmp. Section 3.6.1). This Section discusses the possibility of solving the inner optimization problem by means of genetic algorithms. A main advantage of this approach is the flexibility to extend the algorithm to handle additional constraints simply by modifying the fitness function and/or the genetic operators. Moreover, GAs can easily be scaled with respect to their computational com-

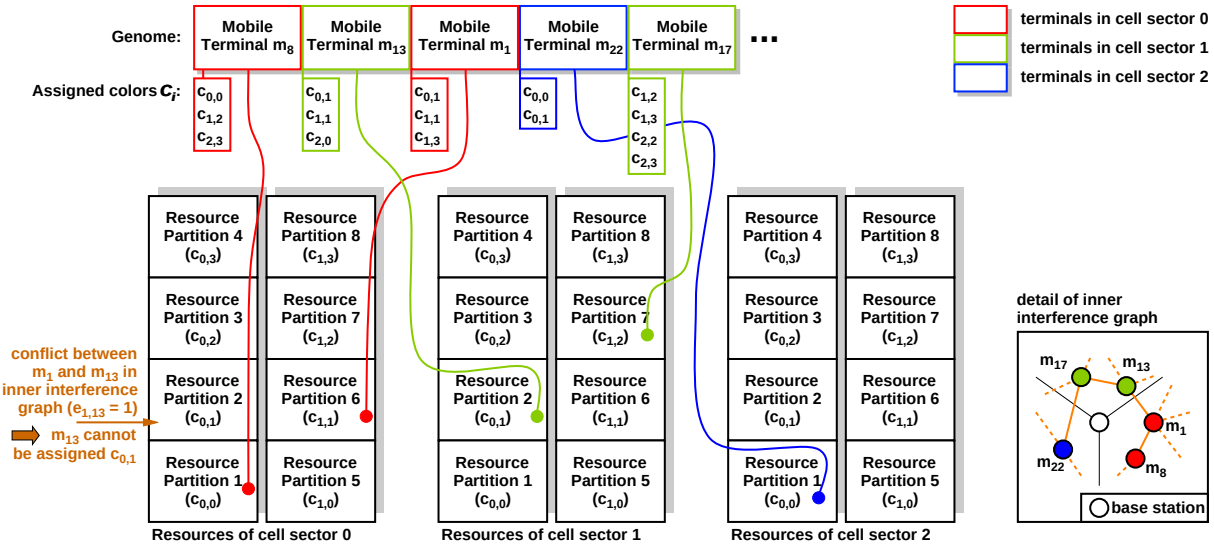



**Figure 5.8:** Representation of genome as a list of terminals, and sequential assignment of terminals in this list to resources in cell sectors. This example contains a conflict in the inner graph between mobile $m_1$ and $m_{13}$ (i.e., $e_{1,13} = 1$), which is why $m_1$ cannot be assigned $c_{0,1}$.

plexity by modifying the number of generations or the population size. Even if the optimization problem is not solvable, the result of the GA can be used for the resource assignment process.

First, Section 5.3.1 introduces the genetic representation and modeling for the inner optimization problem. Subsequently, Sections 5.3.2, 5.3.3, 5.3.4, and 5.3.5 introduce the applied fitness calculation, selection operator, mutation operators, and crossover operators, respectively. Section 5.3.6 describes the underlying genetic algorithm, and Section 5.3.7 discusses compensations for unserved mobile terminals. Section 5.3.8 finally elaborates on implementation aspects of the algorithm.

### 5.3.1 Genetic representation and modeling

The representation of a solution in a genetic algorithm is problem-specific and often not obvious. The challenge lies in finding a compact representation for which it is possible to obtain suitable mutation and crossover operators. These operators must be able to perform the genetic operation quickly and efficiently, and it is beneficial if they produce a valid solution. Moreover, the mutation of a genome or the combination of two genomes must lead to a solution that still sustains some properties of the original genome(s). Last but not least, it must be possible to determine a fitness value for all genomes.

For the inner optimization problem, a list representation was found to be adequate. This is illustrated in Fig. 5.8. The list contains references to the mobile terminals along with the colors $\mathbf{C}_i$ that were assigned during the outer optimization in the central coordinator. The order of the mobile terminals in the list determines the assignment of resources to the mobile terminals. To assign resources, a placement algorithm traverses the list and assigns the first possible and free resource partition to each encountered mobile terminal. The resource partition must not

yet be occupied, and the assignment must not be in conflict with the inner interference graph. For example, in Fig. 5.8, there is a conflict between mobile $m_1$ and $m_{13}$ in the inner interference graph (i.e., $e_{1,13} = 1$), which is why $m_1$ cannot be assigned to $c_{0,1}$. Instead, it is assigned to $c_{1,1}$, which is the next free resource partition available to $m_1$. The placement algorithm immediately takes care that constraints (5.9), (5.11), and (5.12) are fulfilled. This does not apply to constraint (5.10), since not all mobile terminals will necessarily be assigned resources during the placement process due to conflicts in the inner interference graph. Constraint (5.10) is therefore taken into account during the calculation of a genome's fitness value, which is described in the next Section 5.3.2.

Several possibilities to construct the list and work on it are possible. In general, it is desirable to keep the list as short as possible since this allows an easier handling by the GA. In the following, three different list strategies will be considered.

- **Random Genomes**

  In this strategy, genomes have a constant length of

  $$L = k_{OLF} \sum_i N_i \ . \tag{5.13}$$

  $N_i$ is the number of mobile terminals to be served in cell sector $i$, and $\sum_i N_i$ is the total number of mobile terminals to be served by the considered base station. $k_{OLF}$ is an integer overlength factor, where every mobile terminal occurs in the genome exactly $k_{OLF}$ times. Some mobile terminals may occur in the genome more often than they have assigned colors $|\boldsymbol{C}_i|$. In this case, the additional list entries will not cause any performance improvement for the respective terminal. They will also not harm the overall performance, since they will simply be ignored during the resource assignment step. Obviously, $k_{OLF}$ needs to be chosen large enough so that all resource partitions can be allocated to appropriate mobile terminals. If it is chosen too small, not all feasible solutions may be reachable.

- **Weighted Genomes**

  In this strategy, every mobile terminal occurs in a genome as many times as colors were assigned to it by the central coordinator. Consequently, the length of a genome depends on the color sets $\boldsymbol{C}_i$ in the following way:

  $$L = \sum_i |\boldsymbol{C}_i| \ . \tag{5.14}$$

  Compared to Random Genomes, every terminal occurs in the list exactly as often as it potentially could be served. In general, this will lead to relatively long genomes. The choice of an overlength factor is not necessary, and all feasible solutions are principally reachable.

- **Exact Genomes**

  While the previous strategies generally work with over- or undersized genomes, this strategy uses genomes of the same length as there are resource partitions available:

  $$L = k_s |\boldsymbol{C}| \ , \tag{5.15}$$

with $k_s$ being the number of cell sectors served by the considered base station. As a constraint, every mobile terminal has to occur at least once in the list, and no mobile terminal $m_i$ must occur more than $|\boldsymbol{C}_i|$ times. The resulting genomes are relatively short while all feasible solutions are still reachable if the adapted mutation operator described in Section 5.3.4 is used.

In all cases, the genomes are initialized such that the sequence of mobile terminals is random. Different search operators are required depending on the strategy. This will be detailed in Sections 5.3.4 and 5.3.5. Before, the following Sections will deal with fitness calculation and selection.

### 5.3.2 Fitness function

The fitness value of a solution has to take into account several factors. First, the number $n_o$ of occupied resource partitions is directly related to the resource utilization, which should be maximized. Second, constraint (5.10) is taken into account by counting the number $n_u$ of mobile terminals that have not been assigned resources. A penalty is imposed on the fitness value of every solution that does not serve all mobile terminals. Finally, $u_i = 1$ for all mobile terminals $m_i$. Hence, the number of scheduled mobile terminals will be maximized, i.e., it will be attempted to schedule a mobile terminal in every resource partition. The overall fitness $F$ of a genome is then calculated as:

$$F = \xi + n_o - \kappa \cdot n_u \ . \tag{5.16}$$

$\kappa$ is a scaling factor that allows to trade off unserved terminals with the resource utilization. Choosing a small $\kappa$ will emphasize resource utilization at the expense of unserved terminals. With a large $\kappa$, the GA tries to avoid unserved terminals at the expense of the overall resource utilization. By choosing $\kappa > |\boldsymbol{C}|$, it is assured that unserved mobile terminals are always more penalized than any possible gain in resource utilization. $\xi$ is a constant that needs to be chosen large enough such that the fitness value of any solution never falls below zero.

### 5.3.3 Selection operator

From the fitness definition according to equation (5.16) it can be seen that the fitness values show a highly non-continuous behavior as the quality of the solutions improves. Moreover, the constant $\xi$ introduces an offset, leading to a very small relative difference in fitness between very different solutions. Fitness proportional selection therefore requires scaling in order to achieve the desired selective pressure. It is therefore much more practical to use tournament selection, which is not sensitive to the absolute fitness values.

### 5.3.4 Mutation operators

Two mutation operators will be considered. The first is a standard swap operator as described in Section 3.6.3.4.1. It is used in combination with Random and Weighted Genomes. For Exact Genomes, a different mutator was designed. The new mutator swaps genes like the

swap mutator does. Additionally, it may replace a gene with a randomly chosen terminal. The exact procedure is as follows. First, the genome is traversed, and every gene is mutated with probability $R_M$. A second parameter is the *swap-in probability* $R_{SI}$, which determines whether a gene is swapped with another gene, or whether it is replaced. Note that genes are only replaced if the corresponding mobile terminal would not vanish from the genome, and if the newly inserted gene would not lead to the number of occurrences of the corresponding mobile terminal $m_i$ exceeding $|\boldsymbol{C}_i|$.

### 5.3.5 Crossover operators

All considered crossover operators are based on the Standard PMX operator as described in Section 3.6.3.4.2. Variations of the Standard PMX operator were developed and applied depending on the list strategy, taking into account the specific properties of the genome. Moreover, variations were developed aiming at possible performance improvements. In the following, all used crossover operators will be described.

#### *5.3.5.1 Standard PMX for Random and Weighted Genomes*

The Standard PMX operator can be used for Random and Weighted Genomes without any modifications. An efficient implementation of the Standard PMX takes advantage of the fact that every allele occurs only once within every genome. It can realize the PMX operation by simply shuffling the genes within one parent as illustrated in Fig. 5.9. The corresponding flow chart can be found in Appendix C. Such an implementation is equivalent to the PMX operator as described in Section 3.6.3.4.2 and is for example implemented by the GALib [GAlib] used later on. The algorithm searches for genes matching those in the matching region starting from the leftmost position in the genome and exchanges them with the corresponding genes in the matching region. This allows for a very efficient implementation and is perfectly valid if every allele occurs only once in a genome. Obviously, if an allele occurs more than once in the genome as it is the case for $k_{OLF} > 1$, such an implementation of the Standard PMX causes a positional bias. This can be avoided by using the randomized version of the PMX described next.

#### *5.3.5.2 Random PMX for Random and Weighted Genomes*

An example for the operation of the Random PMX is depicted in Fig. 5.10. Instead of searching from the leftmost position for a matching gene during the shuffling process, the Random PMX first searches within the matching region. In the example, the first gene A within the matching region of parent 2 is therefore simply exchanged with the second gene D in the matching region. If there is no appropriate gene within the matching region, a suitable gene is randomly chosen outside the matching region, as it is for example the case for gene pairs D–C and B–D in Fig. 5.10. The Random PMX avoids positional bias, albeit at the expense of a significantly higher computational complexity. A flow chart of the Random PMX is given in Appendix C.

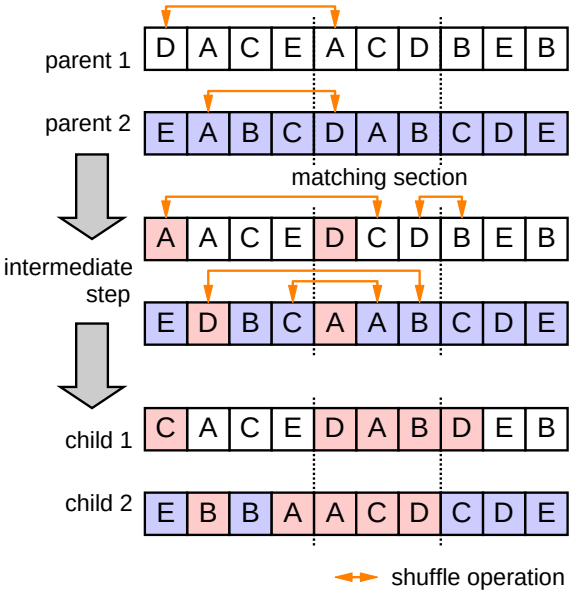



**Figure 5.9:** Example for operation of Standard PMX operator.

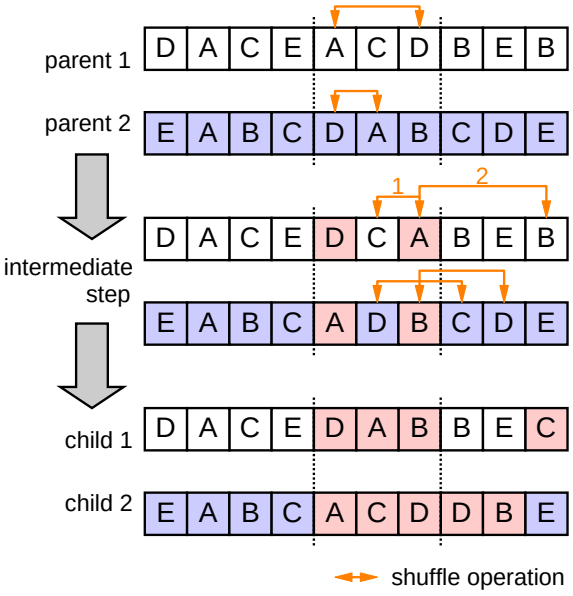



**Figure 5.10:** Example for operation of Random PMX operator.

#### *5.3.5.3 Counting PMX for Exact Genomes*

When using Random or Weighted Genomes, every solution has the same amount of individual mobile terminals $m_i$. With Exact Genomes, the number of occurrences of a particular mobile terminal may differ in every genome. This makes it necessary to adapt the original PMX such that it never erases a certain mobile terminal from a genome. Additionally, it has to take care that a mobile terminal $m_i$ never occurs more than $|\boldsymbol{C}_i|$ times in any genome. The modified PMX achieves this in the following way.

1. Swap matching sections of parents, generating an initial version of the two childs.
2. In every child: Identify terminals $m_i$ that do not occur in the child or which occur too often in the child (more than $|\boldsymbol{C}_i|$ times).
3. In every child: Substitute terminals that occur too often with terminals that do not occur. If this does not suffice to bring the count of all mobile terminals below $|\boldsymbol{C}_i|$, replace terminals that occur too often with randomly chosen terminals that have not reached their maximum count yet.
4. In every child: If there are terminals left that do not occur in the genome, randomly select genes and substitute them with these terminals such that no other terminal vanishes from the genome.

While there exist efficient implementations of the Standard and Random PMX, the Counting PMX exhibits a much larger implementation complexity. This has to be taken into account when comparing the performance of the different list strategies after a certain number of generations.

### 5.3.6 Genetic algorithm

A steady state GA with overlapping populations according to Fig. 3.9 was used. When moving from one generation to the next, $R_R = 50\%$ of the population is replaced by the genetic operators. Since the used variant of the steady state GA first amends the population by the new offsprings and afterwards deletes the worst genomes, the replacement rate $R_R$ has a much smaller impact compared to a variant where the population is first shrinked and then amended by the new offsprings. The complexity of the algorithm is mainly determined by the number of offsprings that have to be generated, that is, by the product $N_{gen} \cdot R_R \cdot |\boldsymbol{P}|$. Consequently, after the same amount of computation, the same amount of offsprings will have been created regardless of $R_R$. While, naturally, the explored portion of the search space is different, the parameter $R_R$ will not be studied further due to its secondary impact regarding the overall complexity of the GA.

### 5.3.7 Optimal solution and compensations

As already discussed in Section 5.1.3, the optimization problem may not be solvable. In this case, the genetic algorithm automatically delivers a sub-optimal solution since it can violate constraint (5.10). Moreover, sub-optimal solutions will be obtained in many cases due to the heuristic nature of the genetic algorithm. Up to a certain degree, non-optimal solutions can be handled without a big performance loss. For example, if the resource utilization of a solution is lower than possible, this will automatically increase the SINR for transmissions in the neighbor cells, thus counteracting the lower resource utilization. Nevertheless, a high resource utilization is desirable.

If a mobile terminal cannot be served during a virtual frame, a compensation may be required in the subsequent virtual frame (see Section 5.1.3). This can easily be reached with the genetic algorithm by modifying the fitness value in such a way that an extra penalty is imposed on any solution that does not serve the mobile terminal requiring compensation. It will be shown in Section 7.4.1.1 that compensation is only required in very few cases. Therefore, and due to the fact that there exists a straightforward realization, this issue will not be treated further.

### 5.3.8 Implementation issues

Genetic algorithms are known for their relatively long computation times. In contrast to that, the inner optimization problem is a real-time problem. An efficient implementation that meets these real-time requirements is therefore inevitable. Often, sequential implementations are used, as it was also done for the performance evaluation in Chapter 7. In contrast to that, GAs exhibit an inherent parallelism, since all mutations, crossover operations, and fitness calculations of one generational step are independent of one another. In order to decrease computation times, this parallelism of GAs can be exploited. This can further be supported by population structures based on demes (cmp. Section 3.6.3.1), which allow the parallel evolution of several sub-populations.

The inherent parallelism also makes genetic algorithms well suitable for a massively parallel hardware implementation on a Field Programmable Gate Array (FPGA) or an Application Specific Integrated Circuit (ASIC). An introduction to this topic is given by Higuchi and Manderick

in [Hig00]. Recently, the genetic algorithm proposed by Necker et al. for the OFDMA frame packing ([Nec08c], see also Section 2.8.4) was implemented in hardware. This GA has many similarities with the GA proposed for the inner optimization problem, in particular with respect to the genome representation. Its FPGA hardware implementation in [Mas08] achieves computation times of about 1 $\mu$s per generation for reasonable parametrizations of the optimization problem. It allows for a parallel execution of all mutations in one generational step as well as for a parallel execution of all fitness calculations. The final evaluation of all solutions has to be done sequentially, just as the crossover operations, which cannot be executed in parallel due to access conflicts to the original genomes and the newly created genomes. Consequently, the crossover operations use up the most processing time in [Mas08]. As it will be seen later in Section 7.4, the inner optimization performs excellent if crossovers are deactivated (i.e., $R_C = 0$). Together with the small number of required generations of only $N_{gen} = 20$, it is expected that an efficient hardware implementation of the inner optimization with a total computation time in the order of several tens of $\mu$-seconds is well feasible.

## 5.4 Quality of Service and Traffic Considerations

In Section 5.2 the assumption was made that all mobile terminals require the same amount of resources. An unequal amount of resources was assigned only due to the interference situation of a particular mobile terminal. Section 5.3 considered unequal traffic demands by means of a utility $u_i$. Further traffic differentiation can be included in these algorithms by only minor modifications. In the following Sections, these modifications will be discussed separately for the outer and the inner optimization problem. Nevertheless, only a brief outline will be given, since traffic differentiation will not be considered further in this monograph.

### 5.4.1 Modification to outer optimization problem

The central coordinator has to provide a sufficient amount of resource partitions such that the traffic demand of a mobile terminal can be satisfied. It was already discussed in Section 5.1.1 that there exists a large number of possibilities to assign color sets $\boldsymbol{C}_i$ to the mobile terminals. One possibility was presented in Section 5.2, where no constraints with respect to the traffic requirements were taken into account. Traffic and QoS demands deliver important constraints to restrict the color assignment process. The easiest way to realize this is to modify the heuristic from Section 5.2 such that a certain minimum number of colors is assigned to particular mobile terminals, which depends on their traffic requirements. This can be done for example by modifying the algorithm in Fig. 5.6 in such a way that it does not randomly select mobile terminals during the coloring round, but instead chooses mobile terminals based on their traffic requirements. This should also take into account delay requirements of the traffic flows. For example, it may be useful to assign one color in every MAC frame for a delay sensitive traffic flow instead of assigning colors belonging to only one MAC frame. This gives the inner optimization the freedom to transmit data in every MAC frame. Instead of a modified heuristic, a more systematic approach could be taken by applying metaheuristics such as genetic algorithms to optimize the color distribution with respect to the traffic demand and the overall number of colors assigned to each mobile terminal.

### 5.4.2 Modification to inner optimization problem

The objective function in equation (5.8) contains a utility $u_i$ for every mobile terminal. This allows for a relative differentiation of traffic flows. For absolute or delay guarantees, an extension of the optimization problem is required. This can be achieved by editing existing or adding further constraints depending on the QoS requirements. For example, a bandwidth guarantee can be realized by modifying constraint 2 in equation (5.10) as follows:

$$\forall i : \sum_k \sum_l x_{ikl} \geq v_i \ . \tag{5.17}$$

$v_i$ is the number of resource partitions which need to be assigned to mobile terminal $m_i$. This number not only depends on the required bandwidth, but also on the current modulation and coding scheme. Similarly, a constraint can be added which requires a color to be assigned to a particular mobile terminal every so and so many MAC frames.

The universal nature of genetic algorithms makes it easy to take into account these additional constraints. In particular, it is sufficient to modify the fitness function in equation (5.16) accordingly, for example by adding a penalty depending on the packet delay of a delay-sensitive traffic flow. This automatically leads to the GA converging towards a solution that fulfills most or all of the required constraints. Further modifications of the GA are only necessary if the Exact Genomes list strategy is used, for which it is necessary to modify the mutator and crossover to take into account the additional constraints. If not all constraints can be fulfilled by the GA (or if no solution exists at all), a compensation in the subsequent virtual frame may be required as mentioned earlier, for example to make up for the assignment of too little bandwidth.

### 5.4.3 Dealing with empty buffers

The so far discussed coordination scheme works well if the traffic requirements can be predicted until the next communication instant with the central coordinator. This is the case for many applications, such as VoIP, video streaming, or file transfers. They either lead to a full input buffer in the base station for a longer period of time, or they exhibit a predictable traffic pattern. This makes it easy to include them in the coordination process. In contrast to that, it is more difficult to handle interactive traffic, such as telnet sessions or web browsing, since the central coordinator may not have assigned resources to the respective mobile terminals when they want to transmit data.

Several possibilities exist to solve this problem. First, the central coordinator could assign colors even to mobile terminals that currently have an empty buffer. Since a color is assigned to several mobile terminals, the inner optimization can compensate for colors “wasted” to terminals with an empty buffer. An advanced central coordination would try to assign every color at least to one mobile terminal with a full buffer. A further enhancement is a probabilistic approach taking into account the probability of every traffic flow to transmit data within the next outer coordination round.

Second, a virtual frame can be divided into a coordinated and an uncoordinated region, allowing for spontaneous data transmissions in the uncoordinated region. This allows the initial transmission of reduced data amounts, such as the first few packets of a web page. If larger data amounts

are then anticipated, the mobile terminal will be included in the color assignment of the central coordinator, allowing a coordinated transmission in the next outer coordination round.

Third, a similar approach can be taken without the need for an uncoordinated region. A convenient approach is to perform the coordination with all mobile terminals that have data to transmit, and then assign random resources within the up to now fully coordinated virtual frame as soon as data arrives for a previously idle mobile terminal. This mobile terminal will then automatically be considered by the central coordinator in the next coordination round. Naturally, the random resource assignment will cause uncontrolled interference. Moreover, the data rate of a newly transmitting mobile terminal will be lower until it is considered by the central coordinator for the first time. This approach is very easy to realize and will be applied later in the performance evaluation in Section 7.6.

### 5.4.4 Further considerations

Up to now, the resource assignment granularity was one resource partition in every MAC frame. It may be necessary to assign more or fewer resources at a finer granularity level per MAC frame, depending on the QoS requirements. A finer resource assignment granularity can easily be achieved by the inner optimization by sub-dividing resource partitions during the resource assignment process. This does not necessarily require any actions by the central coordinator. However, the central coordinator can support this by assigning adjacent colors to the respective mobile terminals. The inner optimization can then occupy adjacent resource partition. Both requires a modification of the corresponding optimization problems, which adds more complexity and complicates the solution algorithms.

## 5.5 Partitioning of Cellular Network into Coordination Areas

It is obvious that a countrywide cellular network cannot be coordinated by one central coordinator. Instead, the network has to be partitioned into smaller *coordination areas* with $N_{cells}$ cell sectors, which are associated with one coordinator. Moreover, the effort of interference coordination may not be justifiable everywhere, for example in rural areas with low traffic demand. This may lead to islands of one or several contiguous coordination areas within the overall network coverage area. The individual coordination areas thereby need to be small enough to be manageable by one central coordinator, and they need to be large enough to achieve a significant performance gain. Especially the regions close to the borders of a coordination area are subject to interference from adjacent coordination areas or cells with no coordination. It is therefore favorable to align coordination areas with the spatial traffic distribution. Often, natural terrain features, such as forests or mountain ridges, can serve as boundaries of coordination areas.

## 5.6 Measurement of Relevant Parameters

In order to be able to build the interference graph, the following parameters need to be determined. First, the matrix $\boldsymbol{I}$ with elements $I_{kl}$ describing the interference to mobile $m_k$ when

transmitting to mobile $m_l$ needs to be identified. Second, the received signal power $S_k$ needs to be determined. While the received signal power is usually known at the base station, for example in terms of the Received Signal Strength Indicator (RSSI) in IEEE 802.16, the elements of $\boldsymbol{I}$ need to be measured separately. The procedure for the measurement and reporting of $\boldsymbol{I}$ strongly depends on the deployed antenna configuration. In the following, this will be outlined for sector antennas and for phased arrays with beam steering.

In the case of sector antennas, the matrix $\boldsymbol{I}$ reduces to a vector if all mobile terminals are served with the same power level, since in this case $I_{kl} = I_{kj} \ \forall l, j$. The elements of $\boldsymbol{I}$ can thus be determined along with the regular scanning of neighboring cells, for example by the measurement of neighboring cell's pilots. This was also proposed in [R1-050896] with a reporting period of $50-100$ ms [R1-060864]. If not all mobile terminals are served with the same power level, the power offset to measurements obtained from pilots can be considered offline in a separate correction step within the central coordinator by collecting the power levels from all base stations.

If beamforming antennas are used, the measurement of $\boldsymbol{I}$ is more complicated, since the antenna's radiation pattern changes depending on the served mobile terminal. There are two basic approaches to take this effect into account. First, the data obtained from the above mentioned scanning procedures can be corrected by precalculated antenna radiation patterns based on the positions of the mobile terminals within the network. These radiation patterns can for example be calculated during the network planning process, or they may be determined from measurements once the network has been deployed. The advantage of this approach is the very low overhead, as all corrections can be done in the base stations or the central coordinator. Its disadvantage is the coarse accuracy.

In contrast, an approach based on active measurements is more precise. By inserting measurement MAC frames with well-known measurement patterns in regular intervals, the mobile terminals are able to measure the elements of $\boldsymbol{I}$ and feed them back to the base stations. A sample structure of a measurement MAC frame is shown in Fig. 5.11. Every cell sector is exclusively allocated a certain amount of data symbols, where the allocated pattern must be known to all base stations and mobile terminals. Within its allocated data symbols, the cell sector then broadcasts known pilot symbols, thereby shaping the antenna beam on every data symbol towards one of its mobile terminals. The transmission pattern within every cell sector needs to be known to all mobile terminals. Then, the mobile terminals can measure the elements of $\boldsymbol{I}$ from the transmitted pilot symbols and feed them back to their serving base station. Note that such an approach based on dedicated measurement signals was also proposed in [IEEE80216h].

The drawback of this approach is the increased overhead. First, measurement MAC frames need to be allocated. Assuming for example two measurement frames per second, the overhead adds up to only 1% of the air interface resources, which is negligible compared to the possible performance gain due to IFCO. More overhead is caused in the uplink direction by the feedback information from mobile terminals. For every measurement frame, each terminal has to transmit $N \cdot (N_{cells} - 1)$ elements of $\boldsymbol{I}$. In a scenario with $N_{cells} = 57$ cell sectors and $N = 10$ active mobile terminals per cell sector, this yields 560 values per mobile terminal, and 5600 values per cell sector. This signaling overhead can be reduced if the outer interference graph is generated with a smaller coordination diameter. In the case of $d_{ic} = 1$, only 2000 values have to be fed back per cell sector. It is well feasible to transmit this amount of data within the uplink portion of a couple of MAC frames. Further optimizations can be done by transmitting only those elements

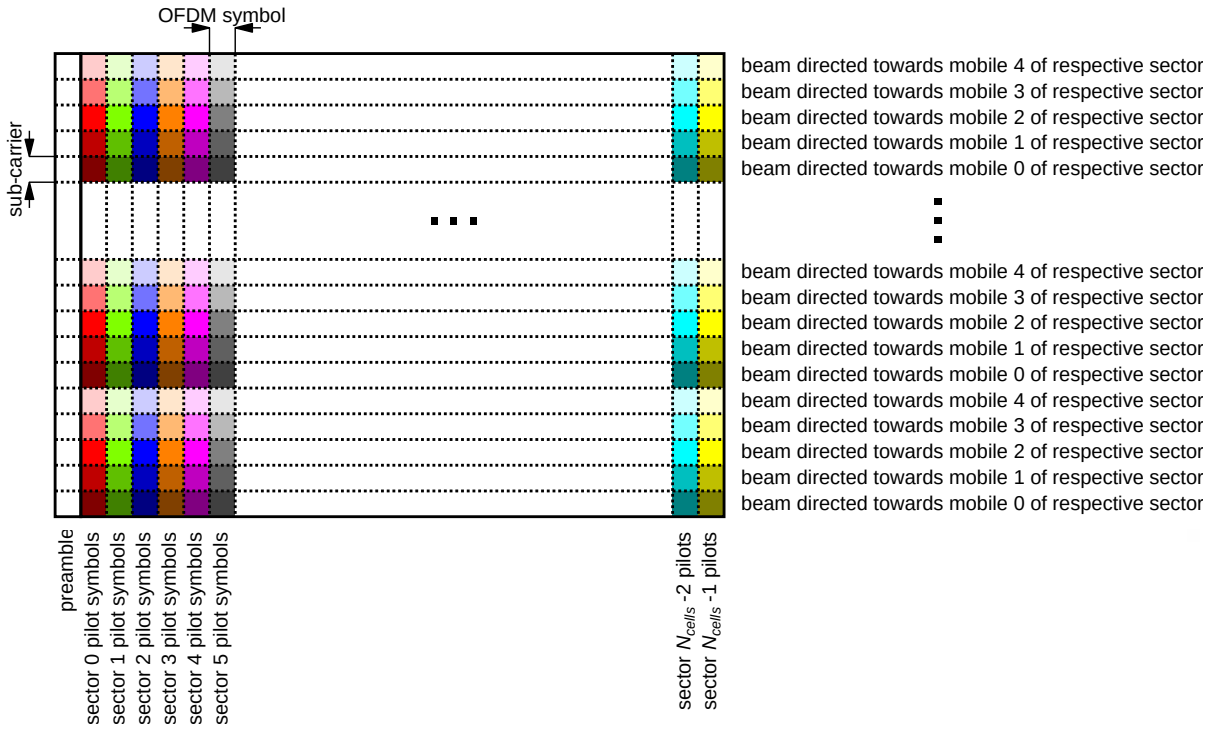



**Figure 5.11:** Structure of a measurement frame with $N = 5$ mobile terminals per cell sector.

of $\boldsymbol{I}$ that have changed since the last feedback report. Especially in a scenario with nomadic users, $\boldsymbol{I}$ will remain static over longer time periods, giving rise to significant savings regarding the signaling and measurement overhead.

## 5.7 Signaling and Timing Constraints

In a distributed system, signaling plays a key role. Two factors are important when considering the system performance, namely the additional load caused by signaling and the delay that signaling messages experience. The signaling load on the air interface was already discussed in the previous Section. In the following, the focus will be on the signaling procedures within the fixed portion of the network. With respect to the delay, the overall delay from measurement until application of the information from the central coordinator in the base stations will be considered.

### 5.7.1 Signaling load between base stations and central coordinator

As discussed earlier, the matrix $\boldsymbol{I}$ with elements $I_{kl}$ needs to be known by the central coordinator. In a scenario with $N \cdot N_{cells}$ mobile terminals in total, this results in the transmission of $N^2 \cdot N_{cells} \cdot (N_{cells} - 1)$ matrix elements in total. Additionally, the received signal power $S_k$ for every mobile terminal needs to be communicated. Hence, approximately $N^2 \cdot N_{cells}$ values have to be transmitted per base station. The communication overhead per base station increases quadratically with the number of mobile terminals per cell sector, but only linearly with the number of cell sectors. In a scenario with $N_{cells} = 57$ cell sectors and $N = 10$ active mobile

terminals per cell sectors, this results in a total number of about 5700 values per base station, which constitutes a well manageable amount of data if communication with the central coordinator does not take place too frequently (depending on the dimensioning of the backhaul network). In return, the central coordinator distributes far less information to the base stations. First, the set of colors $\boldsymbol{C}_i$ per mobile terminal has to be distributed. Second, the duration of the virtual frame has to be communicated, together with synchronization information determining the first frame in which the new information is valid.

### 5.7.2 Transport and delay of signaling messages

Figure 5.12 illustrates the flow of signaling messages between the base stations and the central coordinator. Local state information is transmitted from the base stations to the central coordinator with an update period of $t_{C,up}$, which is also the frequency of measurement MAC frames if a direct measurement of $\boldsymbol{I}$ is chosen. A new set of colors is then applied after a total delay of $t_{C,delay}$, which includes all signaling delays, processing delays, and the synchronization delay. In other words, a new color set becomes active with a delay of $t_{C,delay}$ after the base station has measured local state information.

In the case of IEEE 802.16e or 3GPP LTE, the transport of signaling messages between base stations and central coordinator takes place in an IP network. Therefore, it is likely that there is no guarantee with respect to message loss or message delay. In the event of a message loss, the system performance will decrease gradually, since only a part of the information in the central coordinator will be obsolete or only single base stations are not updated with a new color set. Nevertheless, it is advisable to ensure reliable message transport, for example by means of the Transmission Control Protocol (TCP) or the Stream Control Transmission Protocol (SCTP).

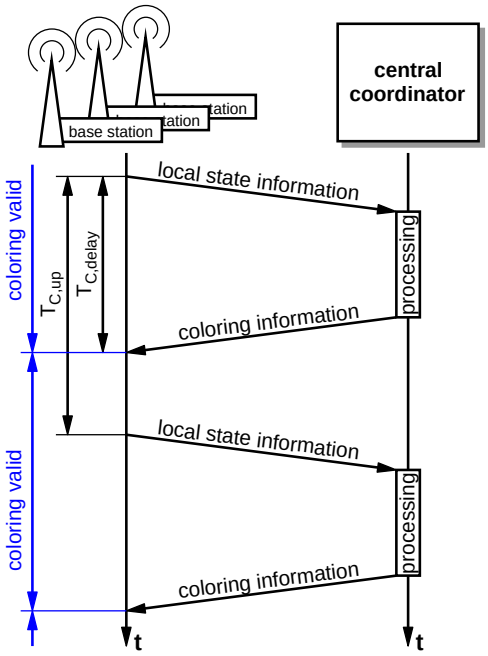



**Figure 5.12:** Timing diagram for the distributed interference coordination.

Possible retransmissions cause additional delay, which needs to be accounted for by the total delay $t_{C,delay}$. Note that, in the case of SCTP, this additional delay is usually not significant. According to measurements by Kiesel and Scharf in [Kie07], a response time below 100 ms can be achieved for 99.9% of all transactions assuming an IP network with a Round Trip Time (RTT) of 20 ms and a packet loss probability of 1%. Considering additional processing and synchronization delay, an overall delay $t_{C,delay}$ of several hundred milliseconds is well feasible.

# 6 Scenario and System Model

This Chapter introduces the object-oriented simulation model and its parametrization, which was used for the performance evaluation in Chapter 7. The considered IEEE 802.16e system was modeled at the system and the network level, based on the methodology introduced in Chapter 3. Its parametrization follows the already mentioned specifications [TS 25.814] (3GPP), [WiM07b] (WiMAX Forum), and [NGMN2007] (NGMN) in many aspects. References will be given were appropriate. The described model was implemented using the event-driven simulation library IKR SimLib [Koc94, SimLib] in a way that allows both event-driven and Monte-Carlo simulations.

The system model emphasizes the downlink direction, since the distributed IFCO algorithm presented in Chapter 5 was specifically designed for the downlink. For the performance evaluation in Chapter 7, user data traffic is therefore carried only in the downlink direction, while the uplink merely carries ARQ/HARQ feedback information. Where appropriate, the uplink was modeled in less detail than the downlink, in particular with respect to physical layer aspects.

After briefly listing the overall system parameters of the considered cellular IEEE 802.16e network in Section 6.1, the Chapter outlines the cellular scenario and mobility model in Section 6.2, the antenna radiation patterns in Section 6.3, and the propagation modeling and system's link budget in Section 6.4. Subsequently, the considered traffic models are outlined in Section 6.5, and the end-to-end data path model is discussed in Section 6.6. This is followed by Section 6.7 with a description of the burst profile management and by Section 6.8 with the discussion of the applied resource and scheduling algorithm. Finally, Section 6.9 concludes the Chapter with the parametrization of the simulation environment.

## 6.1 Overall System Parameters

The overall system configuration was chosen according to band class index 2 as defined by the WiMAX profile [WiM07b]. Hence, an IEEE 802.16e system operating in TDD mode was considered. The system bandwidth was set to 10 MHz, and the frame length to 5 ms. The carrier frequency was assumed to be 2350 MHz. The length of the cyclic prefix was set to $T_G = 1/8 T_U$, which implies an overhead of 12.5% due to the guard band.

## 6.2 Cellular Scenario and Terminal Mobility

A cellular layout according to Fig. 3.4 with wrap-around was used as the underlying geographic scenario. The scenario comprises 19 base station sites with three cell sectors each, resulting in a total number of $N_{cells} = 57$ cell sectors. The distance between base stations was set to $d_{BS} = 1400$ m. Every cell sector is served by a linear phased array with 4 array elements.

Terminals move according to a random direction model (see Section 3.3.2) with a mean free path length of 50 m and a maximum turning angle of 25°. The terminal velocity was chosen to $v = 30$ km/h for all mobile scenarios. All terminals were restricted to their respective cell sectors [Nec06b]. Terminals that were to leave a cell bounce off the cell borders, where the cell borders were approximately determined based on the path loss without shadowing. This model avoids handovers and allows to parametrize the amount of cell load asymmetry by a well-directed placement of mobile terminals. This is advantageous, since both asymmetric and symmetric conditions can be studied. It is important to note that the movement restriction to the respective cells does not alter the uniform spatial terminal distribution of the random direction mobility model. However, this model also has drawbacks, and a closer look is taken in Appendix D.

Most studies were done with a symmetric placement of $N = 9$ mobile terminals per cell sector. In contrast, if mobile terminals are placed randomly in the whole area, a terminal falls into a certain cell sector with probability $1/N_{cells}$. Consequently, the number of mobile terminals per cell sector approximately follows a Poisson distribution. For studies that investigate the performance of the interference coordination algorithms in unsymmetrical load conditions, the number of mobile terminals per cell sector is drawn from a Poisson distribution with mean value $N = 9$. In order to obtain statistically significant results, 10 independent simulation runs are performed where in each simulation run the number of mobile terminals in every cell sector is drawn independently from the Poisson distribution. The histogram of the number of mobile terminals per cell sector of those simulation runs is plotted in Fig. 6.1. An average number of

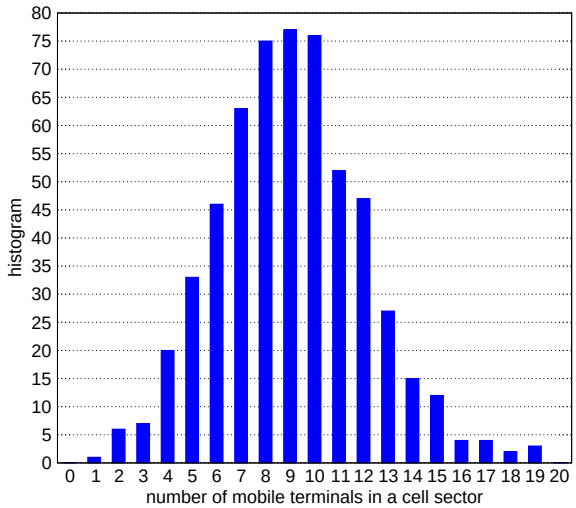



**Figure 6.1:** Histogram of the number of mobile terminals per cell sector for all simulation runs with Poisson terminal distribution.

mobile terminals of $N = 9.037$ is obtained, with the maximum number of mobile terminals in a cell sector being 19.

## 6.3 Antenna Radiation Patterns

The antenna radiation pattern was taken from a typical 120° sector antenna, the SEC35-120ANVH from Radio Frequency Systems [RFS-SEC]. Accordingly, at ±60°, the antenna gain falls by 3 dB compared to its main direction. The SEC35-120ANVH is designed for frequency ranges between 3.4 and 3.6 GHz. Its gain $G_S$ in main direction varies between 15.1 dBi and 15.6 dBi depending on the frequency band. The front-to-back ratio is given as 28 dB. Figure 6.2 shows the horizontal radiation pattern of the antenna at 3.5 GHz normalized to $G_S$. The radiation characteristics of an antenna mostly depend on lengths normalized to the wavelength, allowing for the construction of antennas in the 2 GHz band with similar radiation characteristics. The SEC35-120ANVH radiation pattern was therefore taken as a basis for the performance evaluation. The gain in main direction was adapted to $G_S = 16$ dBi in order to meet NGMN and WiMAX requirements [NGMN2007, WiM07b].

For all beamforming scenarios, a linear phased array with $N_{ant} = 4$ array elements of type SEC35-120ANVH was used. The element spacing was chosen to half a wavelength, and element tapering was applied with $\vec{w}_{taper} = [\sqrt{1/2}\ ,\ 1\ ,\ 1\ ,\ \sqrt{1/2}]$. The weighting factors $w_k$ were calculated as described in Section 2.7.3. The weights were updated for the transmission of every data burst, and the beam was assumed to be steerable towards the mobile terminals with an accuracy of 1°. Together with the tapering weights, the power gain $G_{array}$ on top of the sector antenna is therefore ($\alpha = \varphi = 0$):

$$G_{array} = 20\log \sum_{k=0}^{N_{ant}-1} w_{taper,k} e^{jk\frac{2\pi d}{\lambda}(\sin\varphi - \sin\alpha)} \overset{\alpha=\varphi=0}{=} 20\log\left(2\sqrt{\frac{1}{2}}+2\right) = 10.67 \text{ dBi} \ . \quad (6.1)$$

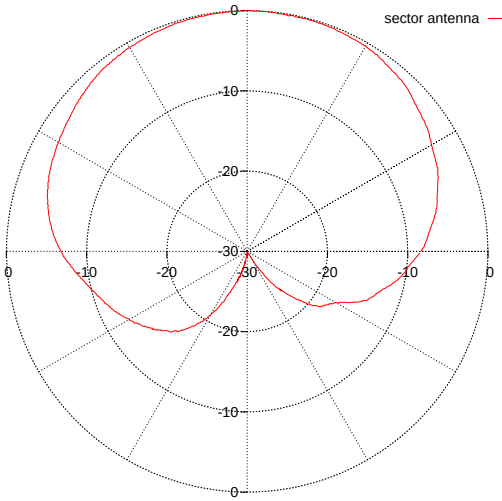



**Figure 6.2:** Radiation pattern of sector antenna RFS SEC35-120ANVH [dBi].

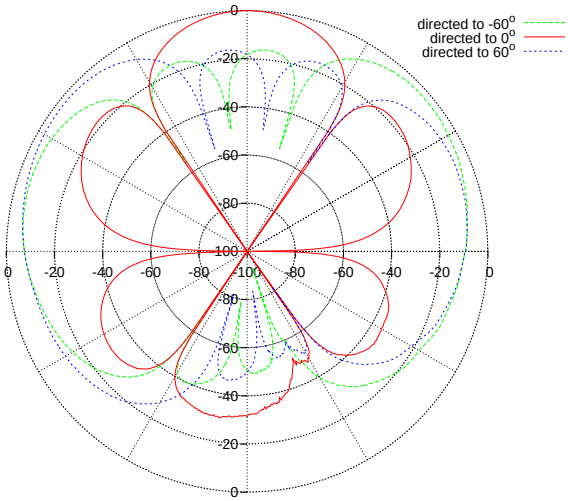



**Figure 6.3:** Radiation patterns of beamformer directed to 0°, −60°, and 60° [dBi].

Examples of the resulting normalized radiation patterns are plotted in Fig. 6.3 for beam steering directions of $\alpha = -60°$, $0°$, and $60°$. For $\alpha = 0°$, a well-shaped beam is obtained, whereas for $\alpha = \pm 60°$ the graph reveals significant side lobes especially in the steering angle's reverse direction $-\alpha$.

The mobile terminal's antenna was modeled as omnidirectional with a gain of 0 dBi [NGMN2007, WiM07b].

## 6.4 Propagation Modeling and Link Budget

The WiMAX forum system evaluation methodology [WiM07b] requires the use of either the COST231-Hata model, or the model according to Erceg et al. (see Section 3.2.1). Accordingly, the Erceg-model will be used in the following, parametrized for terrain category B and with the following minor modifications. First, a frequency correction was applied in order to account for the higher path loss at 2350 MHz. The correction factor is given as [Erc00]:

$$C_f = 6\log\left(\frac{f_C}{1900\ \text{MHz}}\right) = 6\log\left(\frac{2350\text{MHz}}{1900\ \text{MHz}}\right) \approx 0.54\ \text{dB}\ . \tag{6.2}$$

Second, all cells were chosen to be equal, hence $\sigma_\gamma$ was set to 0. Finally, $\mu_\sigma$ and $\sigma_\sigma$ were set to 0. Instead, shadowing was modeled separately based on the model proposed by Gudmundson (see Section 3.2.2). The standard deviation of the lognormal shadowing process was chosen to $\sigma = 8$ dB [NGMN2007]. The decorrelation distance was set to $d_{corr} = 20$ m, which is in the order of the values proposed in [TS 25.814]. Nevertheless, it is only $2/5$ of the value given in [NGMN2007]. This introduces more variations, thus imposing more challenges on the investigated IFCO algorithms. The shadowing correlation of the different channels from one mobile terminal to all base stations was set to 0.5 [NGMN2007, WiM07b]. Fast fading was disregarded for all simulations.

The base station transmit power of a single antenna element was set to 39 dBm. For a linear array with 4 transmit antennas, this corresponds to a total transmit power of 45 dBm if all array elements transmit with the same power level, which is slightly below the value of 46 dBm given by [NGMN2007] for a 10 MHz bandwidth. Due to element tapering, the actual total transmit power is about 1.2 dBm smaller. Furthermore, a cable loss of 2 dB [NGMN2007, WiM07b] and a receiver noise figure of 5 dB were assumed. This results in the link budget components given by Table 6.1.

It should be noted that the considered cellular system is heavily interference limited. Small changes to the noise figures can be neglected compared to the interference. Moreover, changes to the antenna gain, the transmit power, or the cable loss will affect the received signal as well as the received interference in the same way. Therefore, changes to the values given in Table 6.1 will not significantly affect the results presented in Chapter 7.

| Parameter | Value |
|---|---|
| transmit power of array element | 39 dBm |
| base station single antenna gain $G_S$ | 16 dBi |
| cable loss | 2 dB |
| mobile terminal antenna gain $G_{MT}$ | 0 dBi |
| receiver noise figure | 5 dB |
| thermal noise density | $-174.3$ dBm per Hertz |

**Table 6.1:** Link budget components and noise.

## 6.5 Traffic Model

Two different models were used for the user data traffic in the downlink direction. Greedy traffic sources that generate IP packets of length 1500 Bytes were used for most simulation studies. A greedy traffic source implies that there is always enough data to be transmitted within every connection. This implies a full buffer scenario as recommended by [NGMN2007] and [WiM07b] for the measurement of the spectral efficiency. A full buffer model is adequate for many applications that have bulk data to transmit, such as video and FTP downloads, file sharing, or currently emerging interactive map services.

Additionally, an on-off traffic model as illustrated in Fig. 6.4 was used to supplement the performance evaluation. During the on-periods, the input buffer of a connection is well filled. During the off-periods, the input buffer is drained and it remains empty until the beginning of the next on-period. $T_{on}$ and $T_{off}$ are random variables for the duration of the on- and off-period, respectively. $T_P$ is the sum of both random variables and represents one period in the on-off model. The *activity factor* describes the time fraction during which the input buffer is non-empty and is defined as follows:

$$A_F = \frac{E[T_{on}]}{E[T_{on}] + E[T_{off}]} = \frac{E[T_{on}]}{E[T_P]} \quad . \tag{6.3}$$

For the performance evaluation, the mean period duration is set to $E[T_P] = 20$ s. The on-period is chosen constant with $T_{on} = A_F \cdot E[T_P]$. The off-period follows a negative exponential distribution whose mean value is directly determined by the above choices of $E[T_P]$ and $A_F$.

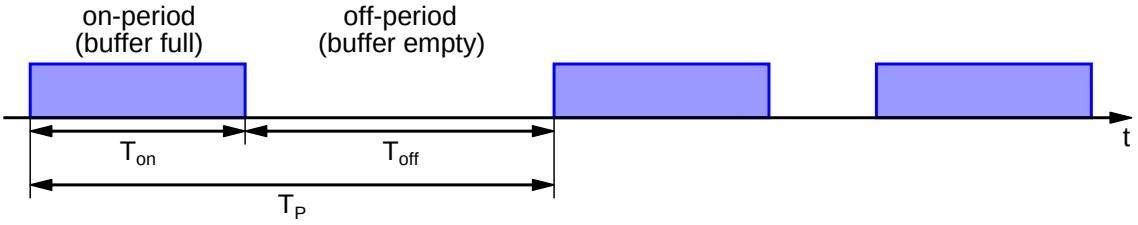



**Figure 6.4:** Illustration of on-off traffic model.

## 6.6 End-to-End Data Path Model of the Considered IEEE 802.16e System

This Section describes the end-to-end data path model. After an overview in Section 6.6.1, the MAC frame structure is described in Section 6.6.2. Subsequently, Section 6.6.3 details the interface between physical and MAC layer, followed by Section 6.6.4 with the description of the data path configuration for ARQ-enabled connections.

### 6.6.1 Overview

Figure 6.5 depicts a simplified block diagram of the downlink portion of the considered IEEE 802.16e system model. The model extends to the uplink direction in a similar way. IP packets are generated by a data source connected via a public IP network. When arriving at the base station, these packets are first stored in an input buffer. Segmentation and concatenation takes place in order to obtain constant size ARQ blocks. Subsequently, the ARQ blocks are concatenated to form a data burst, which is additionally protected by the HARQ. The bursts are then

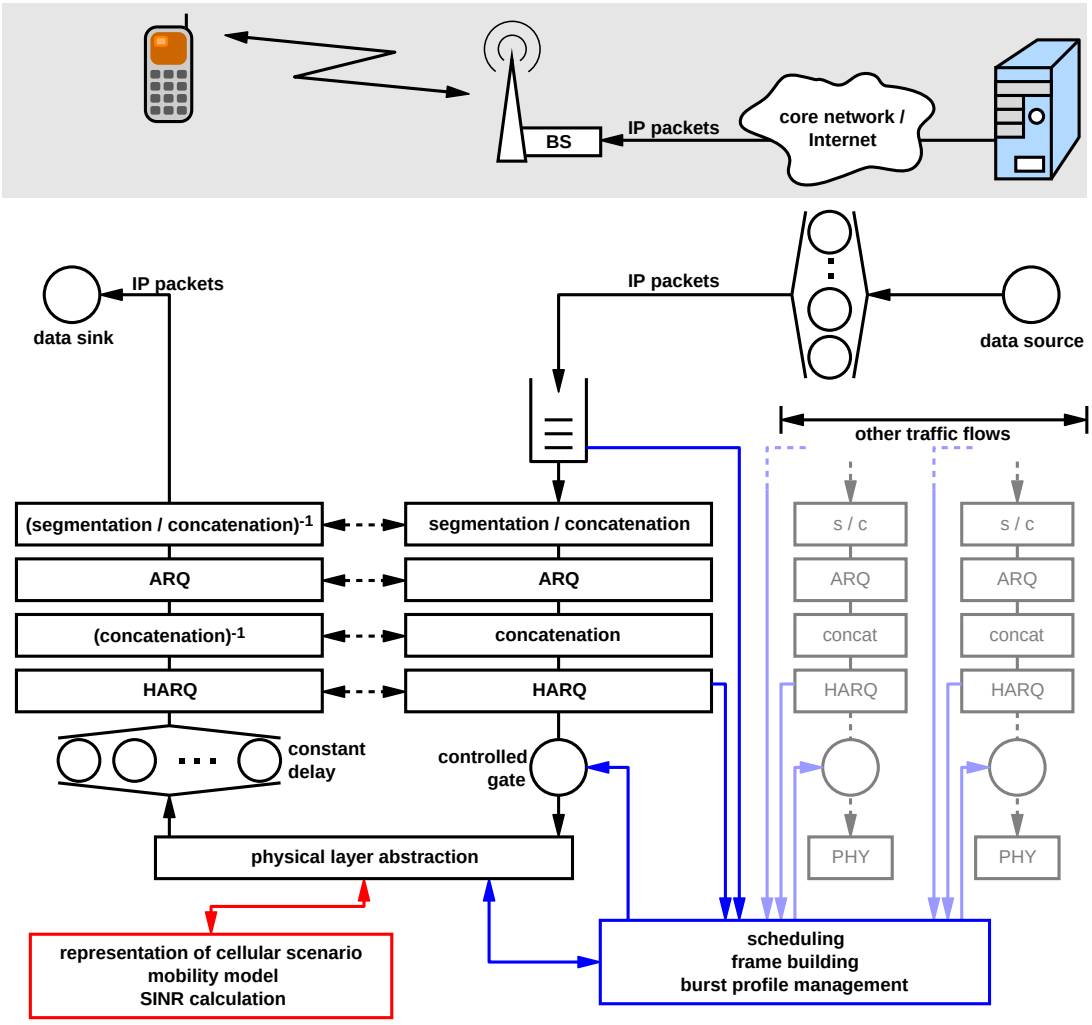



**Figure 6.5:** Simplified view of the downlink portion of the IEEE 802.16e system model.

passed through a physical layer abstraction block, which annotates every burst with its corresponding SINR at the receiver side. After an infinite server with a constant delay that models all processing, propagation, and interleaving delay, the data burst is passed through the receiver side HARQ block, the inverse of the concatenation, and the receiver side ARQ. Finally, the ARQ SDUs are concatenated or segmented to yield the originally transmitted IP packets, which are delivered in-order to higher layers in the mobile terminal.

Multiple parallel data connections towards the mobile terminals are realized by copying of the just described functional blocks. That is, every data connection has its private chain of functional blocks that pass the respective IP packets from the data source to the data sink in the mobile terminal. All data connections are synchronized to the frame level by a separate block, which is responsible for the scheduling, the frame building, and the burst profile management. This block controls a gate in every data connection, thus regulating the number of data bursts that are transmitted per connection in every frame. Moreover, it controls the amount of data per burst and the burst's resource allocation in the respective permutation zone. It is obvious that this model represents a synchronized network, in which all base stations use the same reference clock, with synchronized frame durations and frame starting times.

A separate block models the geographic layout of the cellular scenario, including terminal mobility. It is also responsible for calculating the path loss of all physical channels in the system. Based on this information, the transmit power of the data bursts, and the resource allocation in the permutation zones, it finally calculates the SINR for a particular data burst.

After having presented a high-level overview of the IEEE 802.16e model, the following Sections give more detail on the most important model aspects and implemented algorithms.

### 6.6.2 Frame structure

Interference coordination in IEEE 802.16e systems can either be used within the band AMC zone, or it can be used within the PUSC zone if all base stations are configured with the same permutation base. When doing so, it must be taken care that the respective zones in all cell sectors fully overlap, and that there is no overlap with other zones. The IEEE 802.16e simulation model represents all zones as a two-dimensional resource field, where every element corresponds to a slot. This is sketched in Fig. 6.6, which shows a permutation zone consisting of $N_{slots,t}$ slots in the time direction by $N_{slots,f}$ slots in the frequency direction. This kind of representation is applicable to all types of permutation zones. The number of slots in time and frequency direction depends on the bandwidth and the permutation zone type. Furthermore, the number of slots in the time direction depends on the length of the permutation zone. Given a bandwidth of 10 MHz, the number of slots in frequency direction is $N_{slots,f} = 48$ in band AMC mode, and $N_{slots,f} = 30$ in PUSC mode. In time direction, one slot corresponds to three OFDM-symbols in band AMC mode, and to two OFDM-symbols in PUSC mode.

Only one permutation zone was considered in every MAC frame, and all cells were assumed to use the same permutation mode with equally sized and fully overlapping permutation zones. Frame headers including the downlink maps were assumed to be protected by strong enough error codes (e.g., by means of repetition coding) in order to be received correctly in any case.

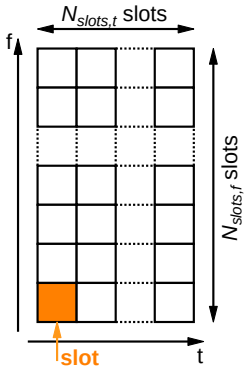



**Figure 6.6:** Slot-level model of a permutation zone.

The size of the resource field was set to $3 \times 48$ slots, which results in a maximum possible data rate of about 6.2 MBit/s if 64QAM 3/4 is applied.

### 6.6.3 Interface between physical and MAC-layer

The interface between the physical and the MAC layer in the downlink direction was based on lookup tables with an average value interface. Lookup tables were generated for all applied modulation and coding schemes. They map an SINR value to the loss probability of a data block of a certain size. This size was chosen to match the length of an FEC block for the respective MCS. The loss probability of a data burst, which consists of a certain number of FEC blocks, can thus easily be calculated from a single table lookup. It should be noted that the size of an FEC block is constant depending on the MCS, except for the last two blocks of a data burst. Since a data burst seldom is an integer multiple of the FEC block size, the last two FEC blocks are often chosen smaller in order to avoid a particularly small and thus error prone FEC block at the very end of the data burst. This effect was neglected, assuming a data burst to consist of a certain integer number of constant size FEC blocks. This greatly simplifies the model and has only a small impact on the overall system performance.

The SINR of a data block was calculated based on the assumption that a particular slot in the resource field of Fig. 6.6 causes interference to the same slot in all other cell sectors. The SINR was calculated based on propagation loss and shadowing without taking into account the effects of fast fading. Instead, the lookup tables accounted for fast fading based on the well-known Vehicular A channel model defined in [ITU-R M.1225]. All lookup tables were generated assuming PUSC permutation and perfect channel knowledge at the receiver.

In the uplink direction, a much simpler statistical model was used. In particular, every data burst was independently lost with constant loss probability $P_L = 0.1$. As an exception, HARQ feedback information was assumed to be received correctly in any case. Obviously, this statistical model effectively disables HARQ for the uplink direction. Nevertheless, the effect on the overall downlink performance is marginal, since the downlink ARQ and HARQ processes can operate very well with the described model.

### 6.6.4 Data path configuration for ARQ-enabled connections

All connections were assumed to be ARQ-enabled. For this configuration, Fig. 6.7 shows the way of IP packets down to the physical layer with respect to the protocol encapsulation. First, the IP packets are segmented or concatenated into smaller size ARQ SDUs, whose size is determined by the ARQ_BLOCK_SIZE system parameter. The fragments of the IP packets are amended by packing subheaders and fragmentation subheaders, as illustrated in Fig. 6.7, depending on whether the fragment begins a new IP packet or is a continuation of an IP packet, respectively. The size of the packing subheader is 3 Bytes, whereas the fragmentation subheader is one Byte smaller. In case there is not enough IP data available to fill up the last ARQ PDU, padding is necessary due to the constant size of the ARQ PDUs, leading to overhead. As a small model simplification, packing and fragmentation subheaders were assumed to have an equal size of 3 Bytes.

The MAC PDUs are then concatenated to form a data burst, which is mapped to the slots of a permutation zone. The burst additionally contains the Generic MAC Header of size 6 Bytes, and a Cyclic Redundancy Check (CRC) checksum of 4 Bytes length. Padding may occur since only an integer number of slots can be allocated. Moreover, the allocation restrictions in the downlink discussed in Section 2.8.4 may cause additional padding due to necessary overprovisioning of resources.

The parameter ARQ_BLOCK_SIZE is limited to powers of two in the IEEE 802.16e standard. Its choice leads to a tradeoff between overhead due to packing and fragmentation subheaders, and overhead due to padding. The optimal choice depends on the applied MCS, and on the size of the data burst. The size was eventually chosen to 64 Bytes, which proved to be a good compromise for all considered scenarios [Nec07f].

ARQ was modeled by a partial implementation of the IEEE 802.16 ARQ mechanism capturing all relevant effects for the performance evaluation. Receiver status reports are generated every 150 ms (status report interval). Irrespective of that, ARQ PDUs are retransmitted after 300 ms if no positive or negative acknowledgment has been received (retransmission timer). The maximum number of retransmissions was set to 10, and the maximum window size to 1024. The latter two parameters are the most important ARQ parameters in a full buffer scenario, and they must not be chosen too small. HARQ was modeled as described in Section 3.4.2, with the described simplifications from [Nec07e]. The number of HARQ channels, each of which

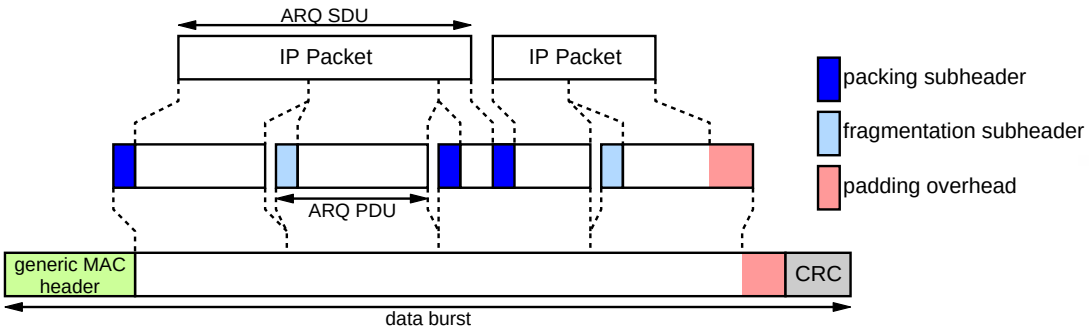



**Figure 6.7:** Packing of IP data into a data burst on the physical layer.

implements a stop-and-wait HARQ process, was set to 8, and the maximum number of retransmissions was set to 4. Feedback information of the HARQ was assumed to be received error-free after three frames in the downlink.

## 6.7 Burst Profile Management

Adaptive modulation and coding was applied in the downlink direction. In the uplink, all data bursts were transmitted with 16QAM 1/2. Due to only feedback information being carried in the uplink direction, and due to the simple physical layer model in the uplink, this circumstance has no impact on the downlink performance. In the following, the operation in the downlink direction is detailed.

The burst profile management takes into account a given target loss probability $P_{L,target}$ of the data burst, which is also referred to as *target Block Error Probability (target BLER)*. In all simulations, $P_{L,target}$ was set to 10%. Furthermore, an estimate of the SINR for the upcoming transmission is required. Based on this information, the MCS is selected to meet the target BLER requirements as good as possible, and a power correction is performed in order to transmit at the lowest possible power. During this process, the burst profile management selects among the modulation and coding schemes QPSK 1/2, 16QAM 1/2, 64QAM 1/2, 64QAM 2/3, and 64QAM 3/4.

AMC requires the prediction of the SINR of future transmissions, which is a challenging issue. It was shown in [Nec07f] that the interference level may vary heavily in the order of 10 dB from one transmission to the next. This translates directly to SINR variations, since the transmission power is kept at an almost constant level. The variations are caused by the beamforming antennas in the considered scenario. As a consequence, the prediction error of the SINR increases, leading to inaccurate selections of the burst profile and thus a decrease of the system performance. A similar effect was observed by Berger et al. in a HSDPA system with two transmit antennas per cell sector [Ber04a]. The authors found the effect to be relatively small, in particular compared to inaccuracies introduced by channel feedback delay. However, it needs to be noted that the effect will increase when moving from two to four transmit antennas. Moreover, since all terminals within a cell are served on the whole frequency spectrum in a UMTS/HSDPA system, the variations due to beamforming are more likely to even out as compared to an OFDMA system, where only one terminal is served within a time/frequency resource block. Variations in the SINR may also be caused by other sources, such as frequency selective scheduling. A specialized scheduling algorithm has been developed in [R1-070099] that aims at minimizing the inter-cell interference fluctuations, thus increasing the performance of AMC.

In the following, the SINR estimation is based on the exponential average of the previous transmissions within a connection:

$$SINR_{avg} = a \cdot SINR_{avg} + (1-a) SINR_{inst} \quad . \tag{6.4}$$

$a$ is the averaging factor, which was set to 0.75 in all simulation studies. $SINR_{inst}$ is the instantaneous SINR of the last transmission as reported by the mobile terminal. It is assumed that this reporting takes place in zero time, i.e., ideal signaling conditions are assumed. The value

$SINR_{avg}$ is then used in order to determine the MCS and the transmit power of the upcoming transmission. This is done according to the algorithm outlined in Fig. 6.8. The algorithm initializes by setting the power correction to 0 dB and the MCS to the least robust MCS 64QAM 3/4. Subsequently, the BLER of the considered data burst is calculated, taking into account the modulation and coding scheme and the estimated SINR. Furthermore, the size of the data burst in terms of number of FEC blocks is taken into account, which may vary greatly when moving from 64QAM 3/4 to more robust modulation and coding schemes. If the BLER is below the target BLER, the MCS selection is terminated, otherwise a more robust MCS is selected, and the BLER is determined again. This process ends once the BLER falls below the target BLER, or once the most robust MCS QPSK 1/2 has been reached. After the MCS selection, the SINR is reduced in 0.1 dB-steps until the BLER moves above the target BLER. This power correction value is then applied to the maximum transmit power for the transmission of the data burst. Note that strong power corrections only take place if the mobile terminal experiences extremely good SINR conditions. In this case, the transmit level is lowered such that the base station can still transmit at 64QAM 3/4, but does not cause unnecessary interference.

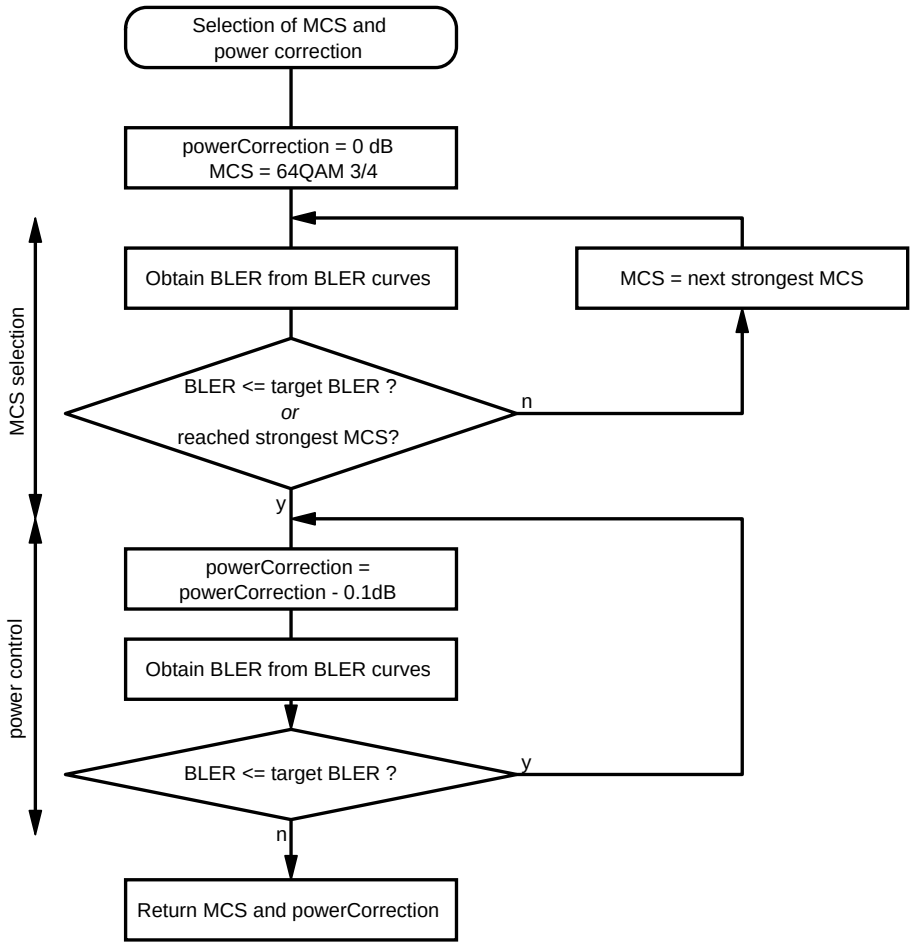



**Figure 6.8:** Power control and selection of modulation and coding scheme.

## 6.8 Resource Assignment and Scheduling

The resource assignment and scheduling process was optimized to handle a full buffer scenario. Every permutation zone is subdivided into $N_F$ equally sized resource partitions, which each can be allocated to a data burst of one connection. In the reuse 1 scenarios, $N_F$ was set to 4. In the reuse 3 scenarios, $N_F$ was set to 6, but only 2 of the partitions were used in a cell sector.

In a full buffer scenario with no traffic differentiation, a Round Robin scheduler is an obvious choice. However, a Round Robin scheduler produces a very predictable traffic pattern, especially if every cell sector contains an equal number of mobile terminals. This leads to a rather deterministic SINR on small time scales for all data bursts within a connection, which is an unwanted side effect for the burst profile management and which would not occur in a real system [Nec07f].

In order to avoid this issue, a Random scheduler was applied. The Random scheduler was designed such that it resembles the important Round Robin property of serving a mobile terminal at least once within a scheduling period. Figure 6.9 shows an example of a scheduling sequence with 5 mobile terminals. At the beginning of every scheduling period, the scheduling sequence of the mobile terminals is randomly determined. It is then fixed until the end of the scheduling cycle is reached. This behavior reproduces a second important property of the Round Robin scheduler, namely that the mobile terminal with the $n$-th highest scheduling priority becomes the terminal with the $(n-1)$-th highest priority in the subsequent scheduling cycle (unless $n = 1$).

## 6.9 Simulation Environment

The simulation environment was implemented based on the event-driven simulation library IKR SimLib [Koc94, SimLib]. Support for genetic algorithms was added using the GAlib [GAlib]. The following Sections discuss the parametrization of the simulation environment, and the application of VRTs.

### 6.9.1 Parametrization

Two types of simulations were performed. First, time-continuous event-driven simulations were performed to evaluate the throughput performance and spectral efficiency of the system. Second, Monte-Carlo simulations were performed in order to parametrize the distributed IFCO scheme of Chapter 5.

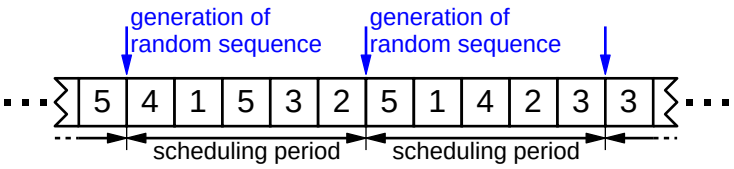



**Figure 6.9:** Illustration of Random scheduler.

All time-continuous simulations with terminal mobility were simulated for a total real time of $T_R = 1000$ seconds. Owing to the wrap-around technique and the large number of base station sites, this proved to be sufficient, since measurements can be taken in every cell sector. The time $T_R$ was divided into 10 batches of length $T_B = 100$ seconds, and an additional transient phase at the beginning ensured that the system reached steady state until the beginning of the first batch.

Exceptions are those simulations that underlay results plotting a certain metric within the observation area defined in Fig. 3.4(a). In order to obtain smooth graphs without measurement gaps, the total real time of the simulation was increased to $T_R = 4000$ seconds, resulting in a batch duration of $T_B = 400$ seconds.

Monte-Carlo simulations were performed with a fixed number of drops $N_{drops} = 1000$ in order to parametrize the genetic algorithm involved in the distributed graph-based interference coordination algorithm. Drops were grouped into 10 batches in order to be able to obtain confidence intervals. In every drop, the mobile terminals are placed randomly in their cells, followed by the solution of the outer and inner optimization problem formulated in Section 5.1 based on this terminal distribution. Indirect performance metrics are recorded afterwards and then a new drop is initiated.

### 6.9.2 Metrics

The objective metrics introduced in Section 3.5 will be used for the evaluation of the different systems. The aggregate cell sector throughput represents the average aggregate throughput realized by all mobile terminals in the network divided by the number of cell sectors $N_{cells}$. The 5% throughput quantile is measured from the cdf of all short-term throughput values produced by all mobile terminals in the network over the course of the simulation run. It is additionally multiplied by the number of mobile terminals per cell sector $N$ for normalization. The short-term period was thereby set to $T_{STP} = 4$ s, which proved to be a good compromise for the tradeoff discussed in Section 3.5.

### 6.9.3 Variance-Reduction Technique

A VRT was applied by using separate RNGs for the mobility and shadowing model, the genetic algorithm, and the graph coloring. For the Monte-Carlo simulations, a separate RNG was used for the placement of the mobile terminals in every drop. This setup ensures identical constellations of mobile terminals and identical channel conditions for every simulation run. Moreover, the genetic algorithms operate with an independent stream of random numbers in order to provide every genetic algorithm with the same initial conditions. This also avoids side effects on other parts of the simulation resulting from a varying amount of random numbers drawn within the genetic algorithm. A separate RNG for the graph coloring ensures identical graph colorings as input for the graph-based interference coordination independent of other parts of the simulation environment.

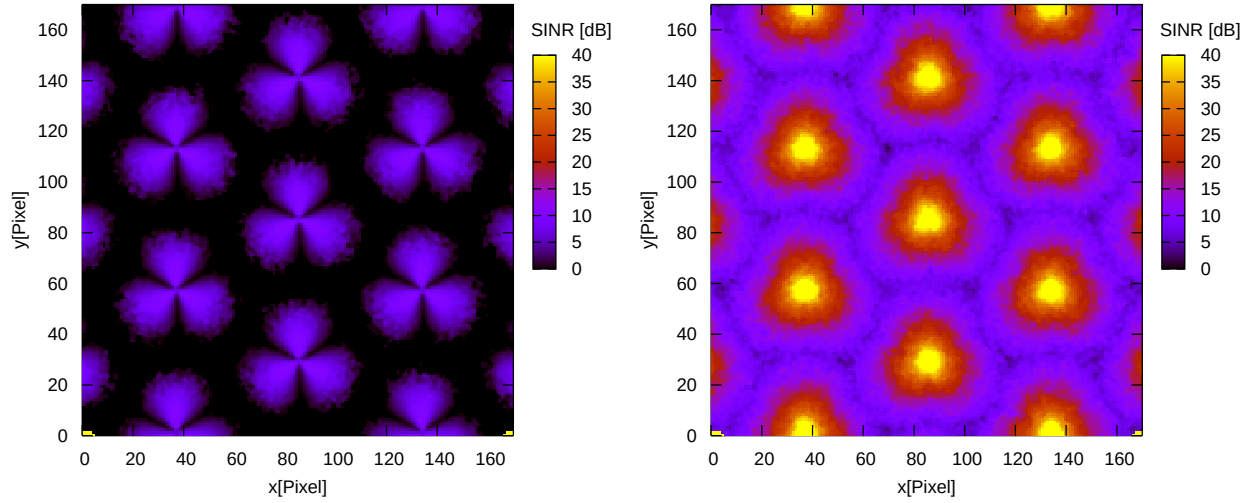
# 7 Performance Evaluation

This Chapter evaluates the performance of a cellular IEEE 802.16e network under different resource assignment schemes. First, Section 7.1 discusses the performance of uncoordinated reference systems with sector and beamforming antennas. Subsequently, Section 7.2 presents performance results for a globally coordinated system based on the graph-based global coordination scheme introduced in Section 4.6, followed by a discussion of locally coordinated systems in Section 7.3. The performance of the distributed graph-based interference coordination proposed in Chapter 5 is then studied in Section 7.4. Finally, Section 7.5 investigates the impact of a non-uniform distribution of mobile terminals within the cellular scenario, and Section 7.6 evaluates the performance with on-off traffic sources.

## 7.1 Reference Systems

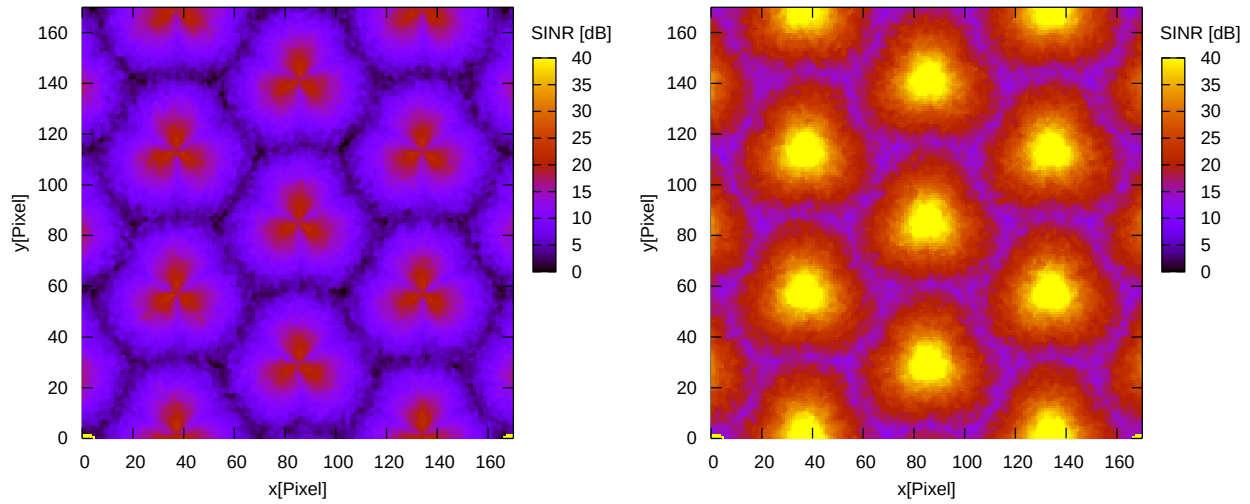
The performance of the interference coordinated systems will be compared to uncoordinated reference systems with a frequency reuse factor of $r = 1$ and $r = 3$, both with sector antennas and with beamforming antennas. The position-dependent SINR-performance of the resulting four reference systems in the observation area as defined in Fig. 3.4(a) is plotted in Fig. 7.1 and 7.2. It is intuitive that the SINR performance is better with a frequency reuse factor of $r = 3$ compared to $r = 1$ due to the larger distance of interfering cell sectors. Alike, the average SINR performance of a system with beamforming antennas is better than that of a system with sector antennas. In all scenarios, the worst performance can be observed in the edge areas of the cell sectors, particularly at the "border triangle" where the borderlines of three cell sectors meet. Moreover, in a reuse 1 scenario, the cell sectors of a base station can be well distinguished since they produce a significant amount of interference on each other. This is not the case in a reuse 3 scenario, where the cell sectors of a base station are only vaguely distinguishable.

From Fig. 7.1 it can be seen that a system with sector antennas achieves a decent average SINR performance only with a frequency reuse factor of $r = 3$, whereas the average SINR performance with $r = 1$ is particularly bad. Due to shadowing, the instantaneous SINR may be significantly better or worse than the average SINR performance at any position. Hence, data transmission may be possible even in areas with low average SINR values. This is reflected by the average area-dependent throughput plotted in Fig. 7.3, which shows data rates above 0 kBit/s even in the cell edge areas of the reuse 1 scenario. When comparing the throughput performance of a reuse 1 and a reuse 3 system, a tradeoff between a reduced resource utilization ($\rho = 1$ for the reuse 1 system and $\rho = 1/3$ for the reuse 3 system) and the achieved SINR improvement occurs. In the case of sector antennas, Fig. 7.3 shows a clear advantage of the reuse 3 system.




(a) frequency reuse 1 (b) frequency reuse 3

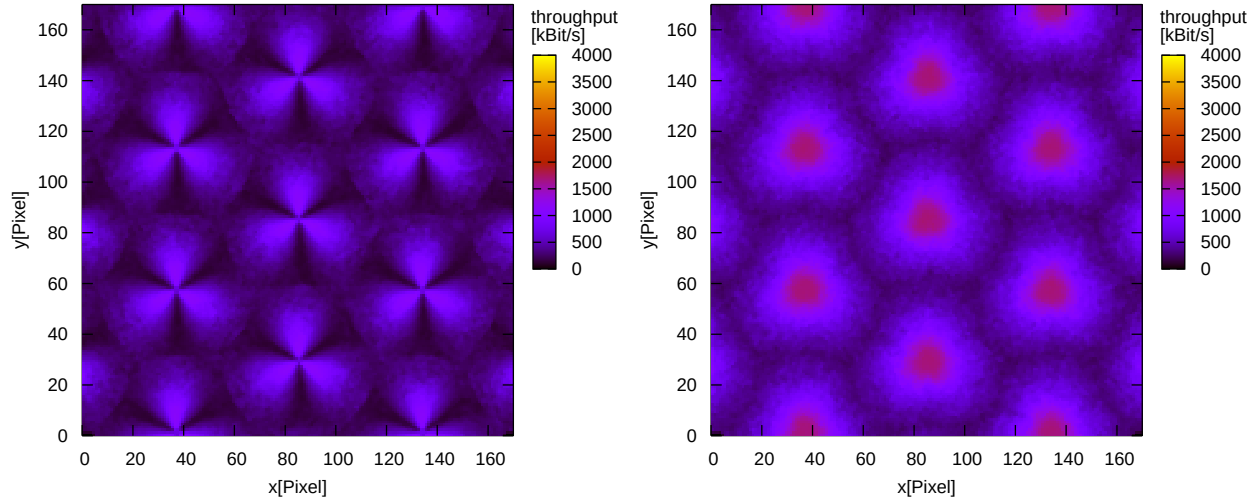
**Figure 7.1:** Area-dependent mean of SINR for uncoordinated frequency reuse 1 and frequency reuse 3 systems with sector antennas.




(a) frequency reuse 1 (b) frequency reuse 3

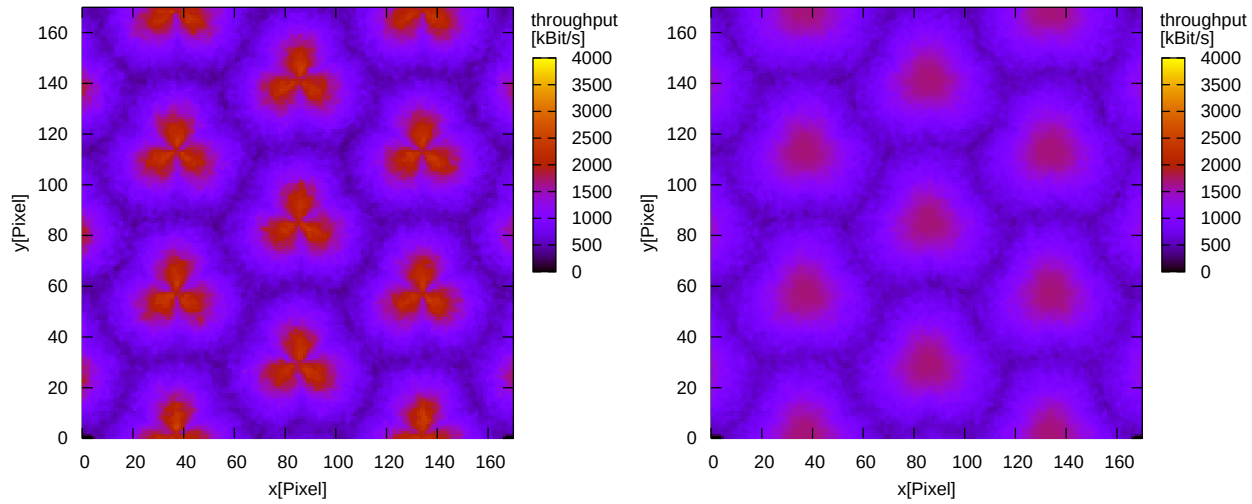
**Figure 7.2:** Area-dependent mean of SINR for uncoordinated frequency reuse 1 and frequency reuse 3 systems with beamforming antennas.

Regarding the average SINR performance of a network with beamforming antennas, Fig. 7.2(a) and 7.2(b) show a decent SINR performance for both a system with $r = 1$ and $r = 3$. Alike, the throughput performance plotted in Fig. 7.4 is at an acceptable level everywhere. In contrast to the system with sector antennas, it is now no longer obvious which reuse factor achieves the better performance. Instead, when comparing the throughput results in Fig. 7.4, a tradeoff between the performance at the cell edge and the overall throughput performance can be observed. For the reuse 1 system, the aggregate throughput per cell sector $D_{aggr,reuse1} = 959$ kBit/s and




(a) frequency reuse 1 (b) frequency reuse 3

**Figure 7.3:** Area-dependent throughput mean for uncoordinated frequency reuse 1 and frequency reuse 3 systems with sector antennas.




(a) frequency reuse 1 (b) frequency reuse 3

**Figure 7.4:** Area-dependent throughput mean for uncoordinated frequency reuse 1 and frequency reuse 3 systems with beamforming antennas.

the 5% throughput quantile $D_{edge,reuse1} = 156$ kBit/s. On the other hand, for the reuse 3 system $D_{aggr,reuse3} = 889$ kBit/s and $D_{edge,reuse3} = 297$ kBit/s. Hence, the reuse 1 system achieves a better overall throughput performance but a worse cell edge performance compared to the reuse 3 system. Note that the given values of the cell edge performance do not correspond to the values in Fig. 7.3 and Fig. 7.4 in the cell edge areas, since the latter charts reflect average values at the respective geographic positions, which are pushed by favorable constellations of mobile terminals and temporally improved SINR conditions due to shadowing.

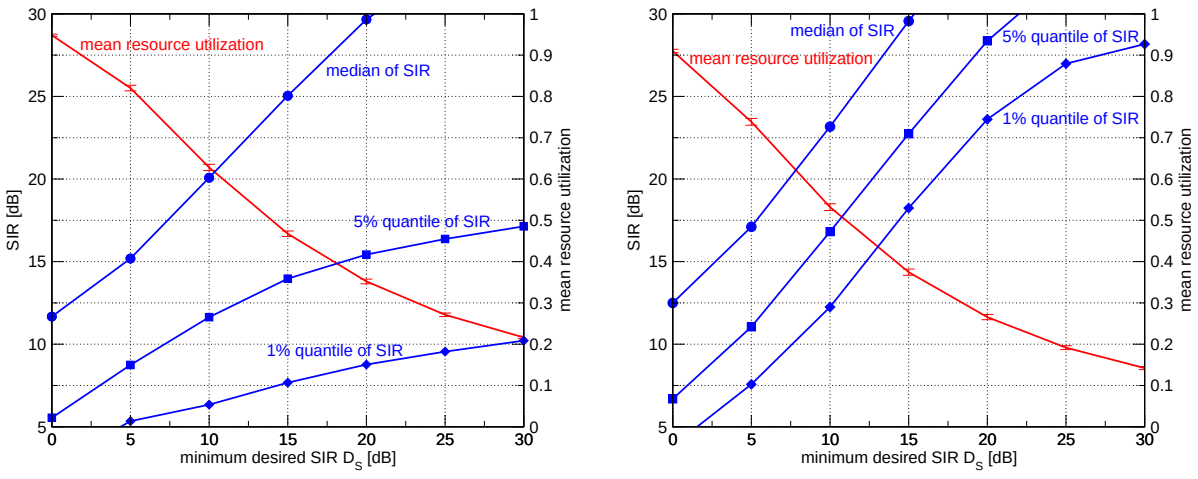
Since all investigated interference coordination schemes are evaluated in combination with beamforming antennas, these values will be used as reference in the following.

## 7.2 Graph-based Global Coordination

This Section evaluates the graph-based global coordination algorithm from Section 4.6 in a frequency reuse 1 network. While the interference coordination algorithm basically has access to the complete frequency range in every cell, the resource utilization will generally be smaller than 1, thus resulting in an effective frequency reuse larger than 1. In the following, Section 7.2.1 treats the coordination based on a single interference graph. Subsequently, Section 7.2.2 evaluates the performance with advanced interference graphs as introduced in Section 4.6.6.

### 7.2.1 Graph-based global coordination with single interference graph

In Sections 4.6.4 and 4.6.5 it was shown that the resource utilization ρ decreases as the desired minimum SIR $D_S$ increases. This stems from an increased number of conflicts in the interference graph and thus an increased difficulty to assign resources to the mobile terminals as $D_S$ increases. On the other hand, the SIR increases as $D_S$ increases, leading to a tradeoff between the SIR and ρ. This is again illustrated in Fig. 7.5 for two different coordination diameters. The SIR is thereby illustrated by means of the median, the 5% quantile, and the 1% quantile. This is more appropriate than the mean value, which is dominated by high SIR values of mobile terminals close to a base station. Both the SIR and ρ behave as expected. When $D_S$ is increased, ρ decreases as was already observed in Section 4.6.5. At the same time, the considered SIR metrics increase almost linearly before going into saturation. When the coordination diameter $d_{ic}$




(a) $d_{ic} = 1$ (b) $d_{ic} = 2$

**Figure 7.5:** SIR and resource utilization for graph-based global IFCO.

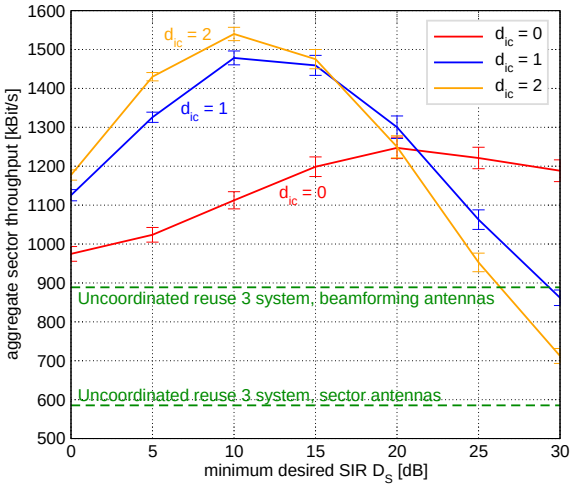



**Figure 7.6:** Aggregate throughput for global graph-based coordination depending on $D_S$.

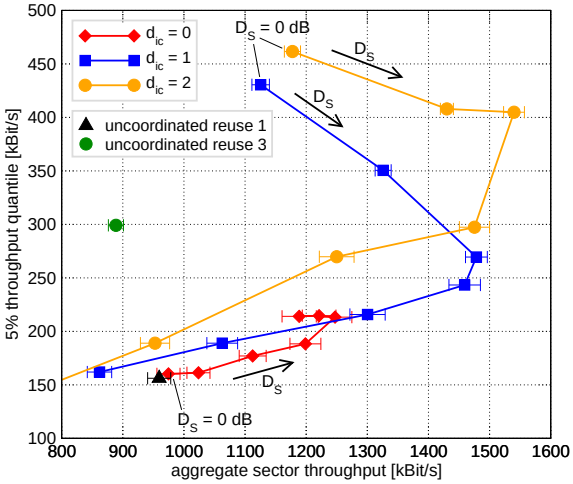



**Figure 7.7:** 5% throughput quantile versus aggregate throughput for global graph-based coordination and different values of $D_S$ and $d_{ic}$. Values of $D_S$ correspond to values used in Fig. 7.6.

is increased, ρ decreases due to the increased number of conflicts in the interference graph. On the other hand, the SIR increases, as can be seen from the comparison of Fig. 7.5(a) and 7.5(b).

The optimal configuration of $d_{ic}$ and the desired minimum SIR $D_S$ cannot be concluded from these considerations. Instead, the throughput performance needs to be evaluated directly. The aggregate throughput per cell sector is plotted in Fig. 7.6 for different values of $d_{ic}$ and $D_S$. The graph shows the anticipated maximum throughput for a particular desired SIR $D_S$. With respect to the coordination diameter, the total sector throughput decreases as $d_{ic}$ is decreased. For smaller $d_{ic}$, it is more difficult to control the interference situation in the border areas of the cell sectors. Consequently, larger values of $D_S$ are required to compensate this effect and achieve the maximum sector throughput. In all cases, the achievable aggregate sector throughput is higher than in the uncoordinated frequency reuse 3 system.

Besides the aggregate sector throughput, fairness is an important issue. Fig. 7.7 plots the 5% throughput quantile over the aggregate sector throughput. Such a representation immediately reveals the tradeoff between the aggregate throughput performance and the cell edge throughput performance, and it allows to easily identify interesting configurations depending on the given requirements. The data points in Fig. 7.7 represent the results of simulation runs for specific coordination diameters $d_{ic}$ and desired minimum SIR values $D_S$. For every value of $d_{ic}$, they are spaced 5 dB apart and correspond to the values of $D_S$ in Fig. 7.6. For a zero-tier coordination, the maximum sector throughput automatically delivers the best cell edge performance. For a larger coordination diameter the cell border performance can be traded off against the aggregate throughput by varying $D_S$. This is particularly the case for the one-tier coordination. In contrast, the two-tier coordination has even more control over the SIR in the cell border areas and achieves an almost maximum throughput quantile and maximum aggregate throughput at the same time. After a certain threshold, the throughput quantile decreases as the minimum desired

SIR $D_S$ increases, since more conflicts in the interference graph are introduced especially for mobile terminals in the cell border areas.

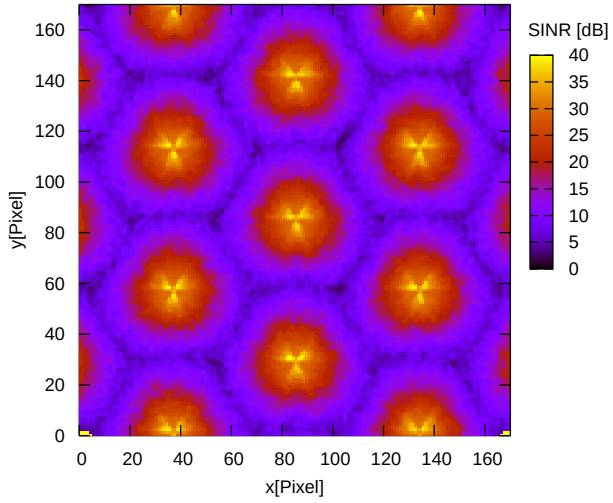
Compared to the performance of the uncoordinated reference systems with frequency reuse 1 and 3, Fig. 7.7 shows a significant improvement both with respect to the aggregate performance and the cell edge performance when global coordination is applied. A locally coordinated system ($d_{ic} = 0$) approaches the performance of the uncoordinated reuse 1 system in the border case, since for very small values of $D_S$ the interference graph contains no edges, thus resulting in a system similar to the uncoordinated reuse 1 system. Note that a coordination of only neighboring base stations achieves an almost as high aggregate throughput as a coordination with a larger coordination diameter. Even the zero-tier coordination, which takes place within a base station and therefore is well feasible, achieves a performance gain of approximately 30% over the reuse 3 system. However, the zero-tier coordinated system suffers from degradation in the cell border areas and cannot match the aggregate performance of the systems with a larger coordination diameter.

Figures 7.8(a), 7.8(b), and 7.8(c) give more insight into the distribution of the mean SINR within the observation area. For $d_{ic} = 0$, the SINR distribution is rather heterogeneous, with good SINR averages close to the base stations, and bad SINR values in the border areas between base stations. Note that the local coordination achieves good SINR values in the border areas between cell sectors of the same base station, in contrast to the uncoordinated reuse 1 system in Fig. 7.2(a). As $d_{ic}$ is increased, the SINR distribution becomes much more uniform. It is interesting to note that $d_{ic} = 1$ still has an obvious SINR fall-off in the border areas between base stations. For $d_{ic} = 2$, the SINR distribution is almost uniform in the observation area, since the interference of all base stations is controlled.

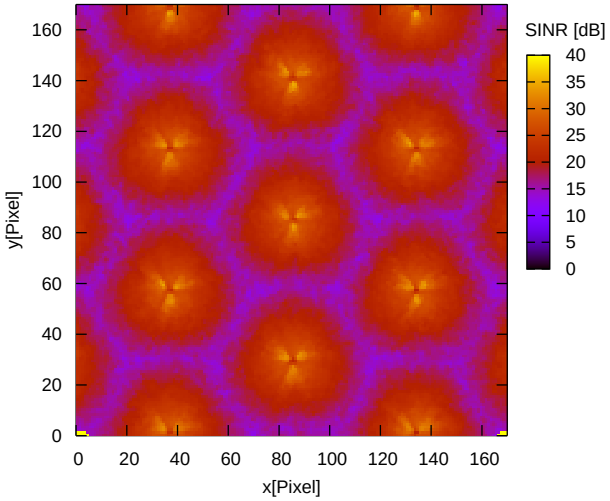
In contrast to the SINR, the throughput will not have a uniform distribution within the observation area. Figures 7.9(a), 7.9(b), and 7.9(c) plot the throughput for the zero-tier, the one-tier, and the two-tier coordinated systems. The throughput improvement is mainly observed in the inner portion of the cell area, especially when comparing the results to the uncoordinated systems in Fig. 7.4. The graphs also reveal the cell border areas where the throughput is particularly low. The throughput in the border areas can be improved by increasing $d_{ic}$, or by sacrificing aggregate sector throughput with a different choice of $D_S$. Figure 7.9(a) also shows that the zero-tier coordinated system suffers from degradation in the cell border areas towards other base stations and cannot match the aggregate performance of the systems with a larger coordination diameter. One approach to solve this problem while still avoiding coordination in-between base stations is the usage of FFR, which will be evaluated in Section 7.3.

When comparing the SINR distribution in Fig. 7.8(a) close to the base station with the corresponding average throughput values in Fig. 7.9(a), it can be seen that the maximum throughput values can be found in the boresight direction of the respective cell sectors, while the maximum SINR is found in the border areas of cell sectors of the same base stations. The reason for this is the increased number of conflicts in the interference graph in the cell border areas, leading to a decreased number of scheduled mobile terminals and an increased SINR.

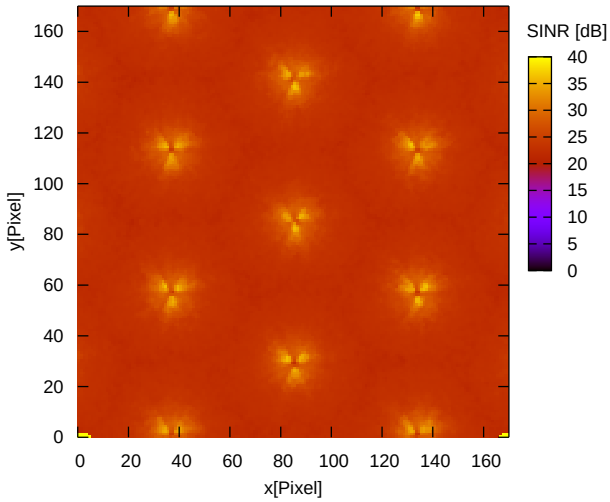
It is interesting to note that several other publications only consider the strongest interferer for interference coordination, such as done by Li and Liu in [Li03]. Xiao et al. [Xia06] claim that the strongest interferer matters the most when interference coordination is done. On the other




(a) $d_{ic} = 0$, $D_S = 20$ dB




(b) $d_{ic} = 1$, $D_S = 10$ dB



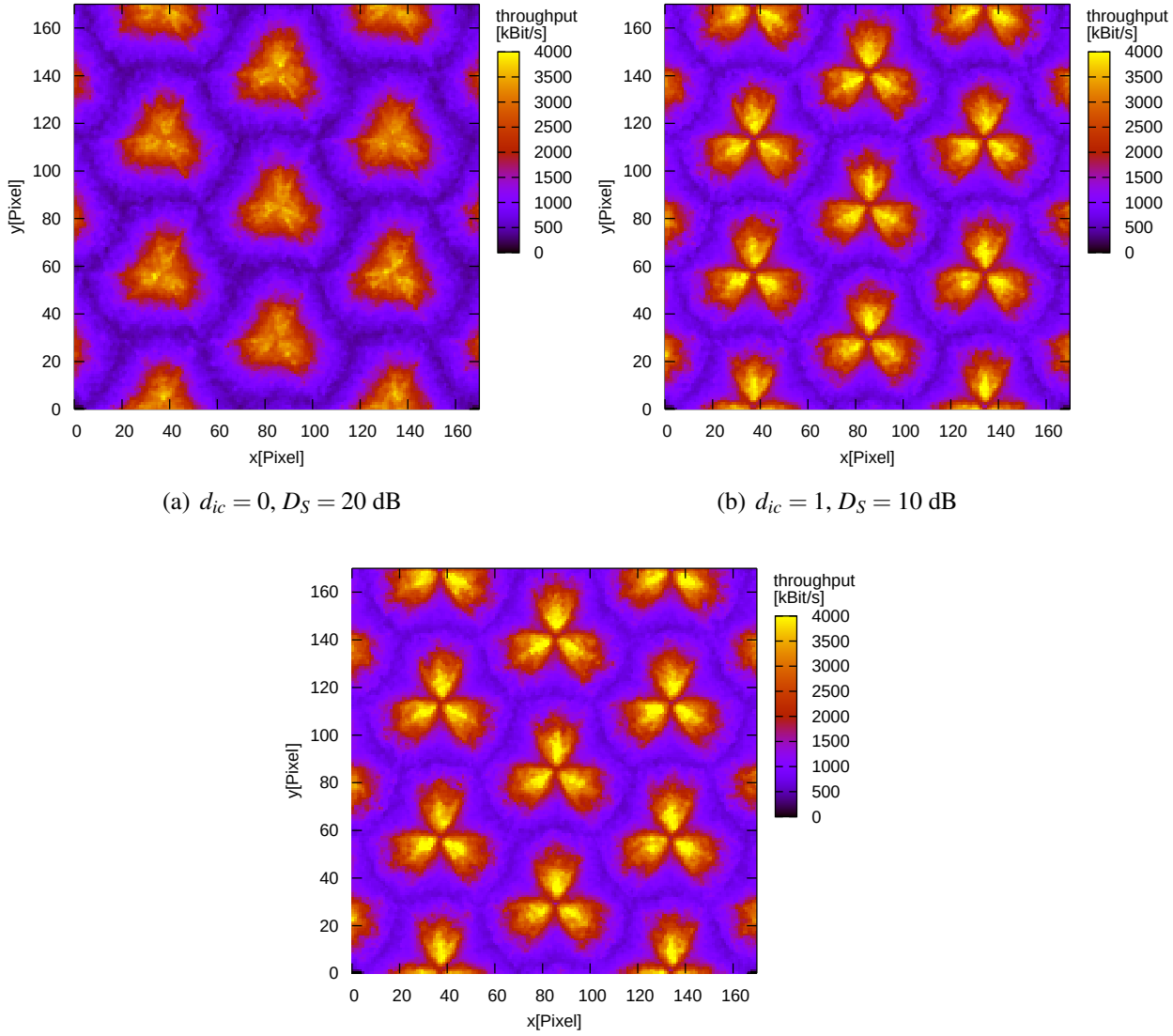

(c) $d_{ic} = 2$, $D_S = 10$ dB

**Figure 7.8:** Area-dependent SINR mean for systems with graph-based global coordination.

hand, the results presented in this Section clearly indicate that considering only one interferer is not sufficient, since a one-tier coordination already considers one full tier of interfering base stations. Yet, there is room for performance improvements when moving to a two-tier coordination.

### 7.2.2 Graph-based global coordination with advanced interference graph

This Section evaluates the possibility of an advanced interference graph creation according to Section 4.6.6. The results from the previous Section show that a one-tier or two-tier coordination with a fairly low desired SIR $D_S$ achieves an excellent cell edge performance while falling short with respect to the aggregate throughput. On the other hand, a zero-tier coordination



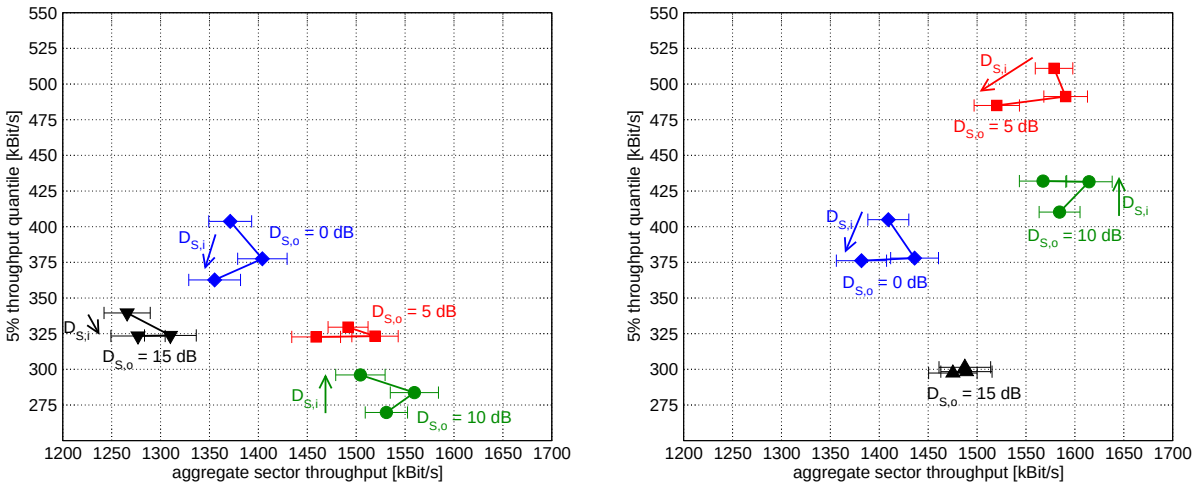

(a) $d_{ic} = 0$, $D_S = 20$ dB

(b) $d_{ic} = 1$, $D_S = 10$ dB

(c) $d_{ic} = 2$, $D_S = 10$ dB

**Figure 7.9:** Area-dependent throughput mean for systems with graph-based global coordination.

provides an increased aggregate performance while falling short with respect to the throughput quantile. When recalling the observations for the vertex degree in Fig. 4.13, it can be seen that both configurations feature an interference graph with a relatively small vertex degree compared to the optimum case of a two-tier coordination with $D_S = 10$ dB. This suggests that a separate generation of zero-tier and one/two-tier interference graphs and a subsequent merging of both graphs will lead to a lower vertex degree as in the two-tier case while providing an optimized SIR within the area. As discussed before, a lower vertex degree will lead to a lower chromatic number of the interference graph and hence potentially increase the system performance. Merging of the two graphs is done simply by including an edge in the merged graph whenever one of the two original graphs contains an edge, as discussed in Section 4.6.6.




(a) $d_{ic,o} = 1$ (b) $d_{ic,o} = 2$

**Figure 7.10:** Advanced generation of interference graphs for different parameters of $D_{S,i}$ and $D_{S,o}$, $d_{ic,i} = 0$, $D_{S,i} = \{15, 20, 25\}$ dB.

The performance of a system with a global coordination based on the combination of interference graphs is plotted in Fig. 7.10 for a combination of a zero-tier with a one-tier interference graph (Fig. 7.10(a)) and a two-tier interference graph (Fig. 7.10(b)). In the following, the zero-tier graph with coordination diameter $d_{ic,i} = 0$ will be denoted as inner graph, which is generated for a desired minimum SIR $D_{S,i}$, and the one/two-tier graph with coordination diameter $d_{ic,o}$ will be denoted as outer graph, generated for a desired minimum SIR $D_{S,o}$.

Figure 7.10 shows a strong dependence of the system performance on $D_{S,o}$. As $D_{S,o}$ is increased from 0 dB in the one-tier case, the aggregate throughput performance increases, while the cell edge performance decreases. When $D_{S,o}$ is increased further to $D_{S,o} = 15$ dB, the cell edge performance starts to increase again, however at a significant aggregate throughput penalty. The optimum performance is therefore obtained for values of $D_{S,o}$ between 0 dB and 5 dB. For the case of $d_{ic,o} = 2$, the optimum configuration is even more pronounced at values of $D_{S,o}$ between 5 dB and 10 dB. The desired minimum SIR $D_{S,i}$ of the inner graph needs to be chosen in the already known range between 15 dB and 25 dB, where it can be seen from Fig. 7.10 that its influence is much smaller than the influence of $D_{S,o}$.

As a conclusion, $d_{ic,o}$ needs to be set to 2 in order to achieve a good performance. In this case, the cell edge throughput can be increased by about 20% compared to regular global two-tier coordination while maintaining the same high aggregate throughput.

## 7.3 Graph-based Local Coordination with FFR

This Section deals with the combination of FFR with local graph-based coordination ($d_{ic} = 0$) as it was also studied in [Nec07c, Nec07d]. Overlapping frequency bands as shown in Fig. 4.3(b) are used. Two different methods for the assignment of mobile terminals to the reuse 1 or reuse

3 bands are investigated. First, the distance-based FFR performs this assignment based on the distance $d_{MT}$ of the mobile terminal to its serving base station, which will be evaluated in Section 7.3.1. Second, an assignment based on SINR-measurements will be evaluated in Section 7.3.2.

### 7.3.1 Distance-based FFR

For the distance-based assignment, a distance ratio $d_{13}$ is introduced. If $2 \cdot {}^{d_{MT}}/_{d_{BS}} < d_{13}$, the mobile terminal is served in the reuse 1 band, otherwise it is served in the reuse 3 band. The parameter $d_{13}$ thus allows modifying the size of the inner reuse 1 area. Choosing $d_{13} = 0$ will result in a pure frequency reuse 3 setup, and large values of $d_{13}$ will yield a frequency reuse 1 system.

Figure 7.11 shows the utilization of resources ρ and the median of the sector SIR depending on the distance ratio $d_{13}$. If $d_{13}$ is increased, the cell area where a reuse of 3 is enforced becomes smaller and the utilization of resources increases. At the same time, the median of the SIR decreases. Naturally, this will lead to a tradeoff. Figure 7.12 therefore shows the total sector throughput depending on $d_{13}$ for different values of $D_S$. A desired minimum SIR $D_S$ of 20 dB delivers the best results. This is in line with the results of pure interference graph based coordination without FFR in Fig. 7.6 for a coordination diameter of zero tiers. With respect to the distance ratio $d_{13}$, a value of about 0.6 delivers the best results, which nicely fits the results of [Ste03a].

Figure 7.13 plots the 5% throughput quantile over the aggregate sector throughput for $D_S =$ 20 dB. With respect to both the quantile and the total throughput, the performance of the global graph based coordination scheme with inter-cell coordination cannot be met. The performance is rather comparable to the previously investigated zero-tier coordination scheme, where the additional FFR now allows to trade off the throughput quantile and the aggregate sector throughput. From the chart we can see that the aggregate throughput can be pushed to an almost as high throughput as in the globally coordinated system while sacrificing $50-70\%$ of the cell border performance.

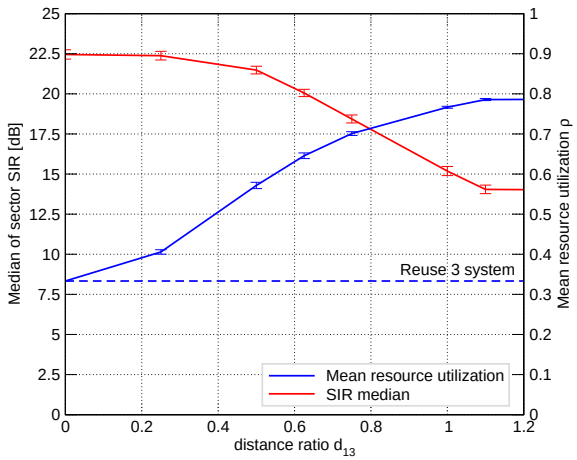



**Figure 7.11:** Distance-based FFR with 0-tier coordination: Median of SIR and mean utilization of resources, $D_S = 20$ dB.

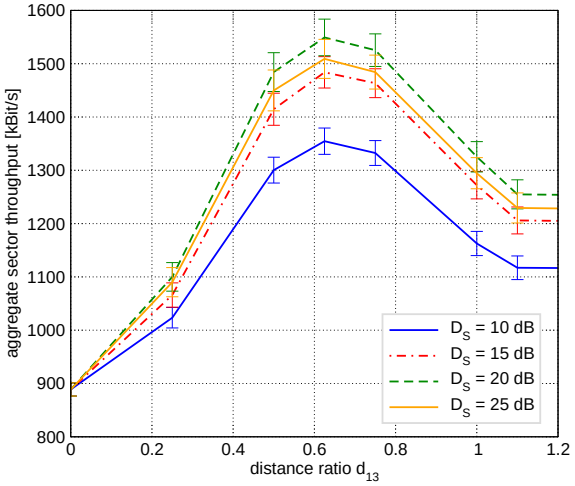



**Figure 7.12:** Distance-based FFR with 0-tier coordination: Average sector throughput over $d_{13}$ for different values of $D_S$.

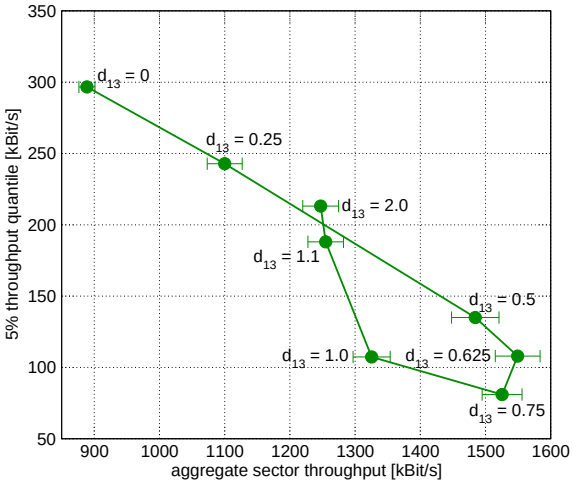



**Figure 7.13:** Distance-based FFR: 5% throughput quantile depending on the total sector throughput, $D_S = 20$ dB.

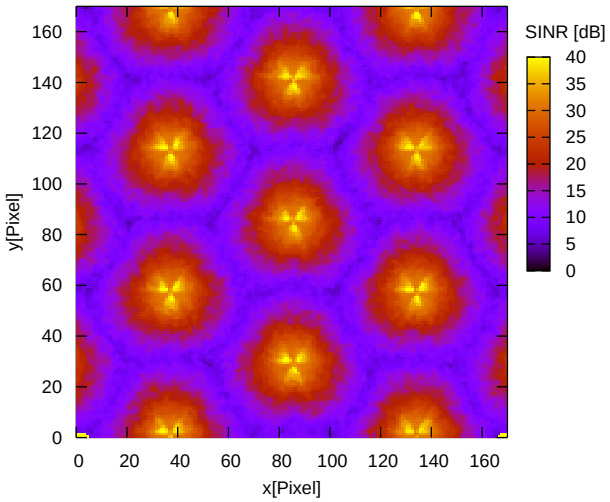



**Figure 7.14:** Distance-based FFR: area-dependent mean of SINR, $d_{13} = 0.625$, $D_S = 20$ dB.

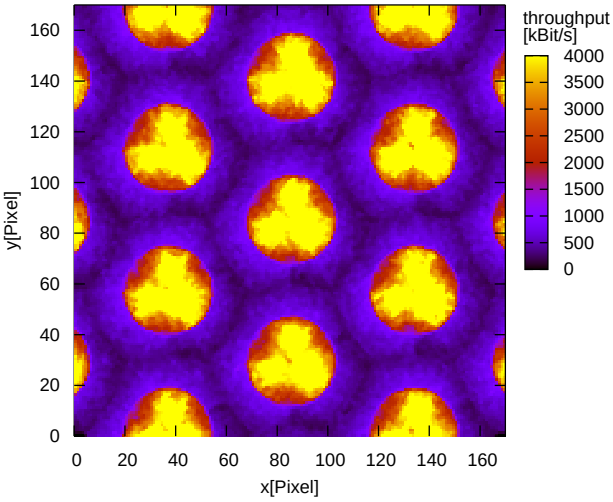



**Figure 7.15:** Distance-based FFR: area-dependent throughput mean, $d_{13} = 0.625$, $D_S = 20$ dB.

The mean SINR within the observation area is shown in Fig. 7.14 for the optimal parameter choice $d_{13} = 0.625$ and $D_S = 20$ dB. Compared to the same system without FFR in Fig. 7.8(a), the SINR is improved in the border areas of the cells as intended. Due to the reduced resource utilization in the reuse 3 areas, this does not allow any conclusions for the achievable throughput. The area throughput is therefore plotted in Fig. 7.15. Compared to the area-dependent throughput of the system without FFR in Fig. 7.9(a), the throughput strongly increases in the reuse 1 areas close to the base stations, while a significant decrease can be observed in the border areas between base stations. Owing to the local coordination, there is no overly strong throughput decrease in the border areas between cell sectors of the same base station. These results confirm the observations for the 5% throughput quantile from before.

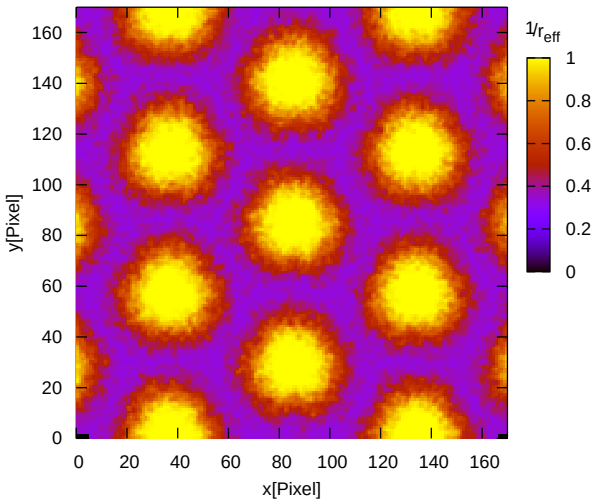



**Figure 7.16:** SINR-based FFR (coordinated): area-dependent effective reuse, illustrated with $1/r_{eff}$. $th_{low} = 15$ dB, $th_{up} = 25$ dB, $D_S = 20$ dB.

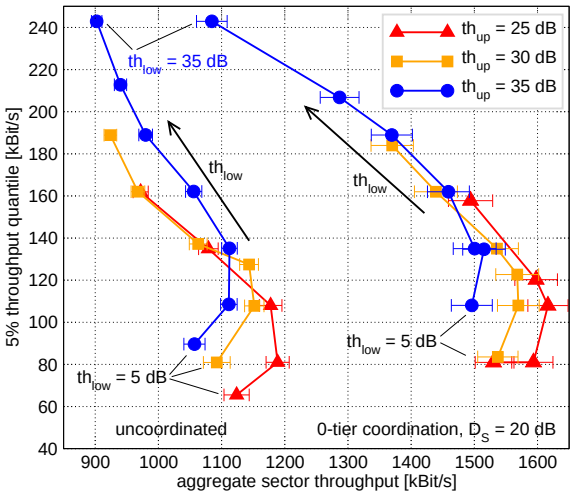



**Figure 7.17:** SINR-based FFR: 5% throughput quantile depending on aggregate sector throughput, $D_S = 20$ dB.

Fig. 7.15 reveals a sharp edge at the given distance ratio, where the throughput drops by a factor of $4-5$. This is rather undesirable, and can be avoided by the SINR-based FFR evaluated in the following Section.

### 7.3.2 SINR-based FFR

According to the classification in Section 4.5.2.3, the assignment of mobile terminals to the reuse partitions can be done based on SINR measurements of recently received data frames. In the following, the SINR estimation according to Section 6.7, which is also used for the burst profile management, will be used. To take into account the high variability of the instantaneous SINR, the decision regarding the reuse 1 or reuse 3 area is based on a hysteresis. This is done by introducing an upper SINR threshold $th_{up}$ and a lower SINR threshold $th_{low}$. If the SINR estimate rises above $th_{up}$, the mobile terminal will be put into the reuse 1 partition. If the SINR estimate falls below $th_{low}$, it will be put back into the reuse 3 partition.

As a first result, Fig. 7.16 plots the average reuse factor which a mobile terminal experiences within the observation area. As expected, the cell borders are covered with a reuse of 3, while large portions of the cell area are covered with an effective reuse of $1-2$. A sharp edge as with the distance-based FFR is avoided. Figure 7.17 plots the 5% throughput quantile over the total sector throughput for different SINR thresholds. Plotted are the results for the described system with zero-tier coordination, and additionally for the same system with deactivated coordination. All points of one curve represent different values of $th_{low}$ and are spaced 5 dB apart with the first point representing $th_{low} = 5$ dB and the last point $th_{low} = th_{up}$. Based on the previous results for a system with zero-tier coordination, $D_S$ is set to 20 dB. From Fig. 7.17 it can be seen that the uncoordinated system cannot match the performance of the coordinated system with respect to the aggregate throughput. In both cases, $th_{up}$ and particularly $th_{low}$ allow to

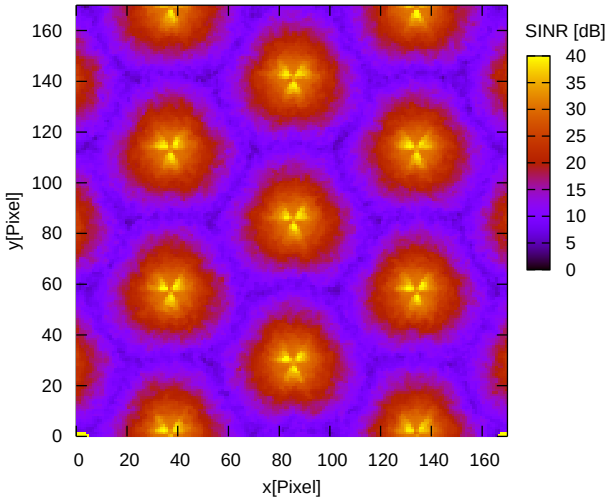



**Figure 7.18:** SINR-based FFR (coordinated): area-dependent mean of SINR, $th_{low} = 15$ dB, $th_{up} = 25$ dB, $D_S = 20$ dB.

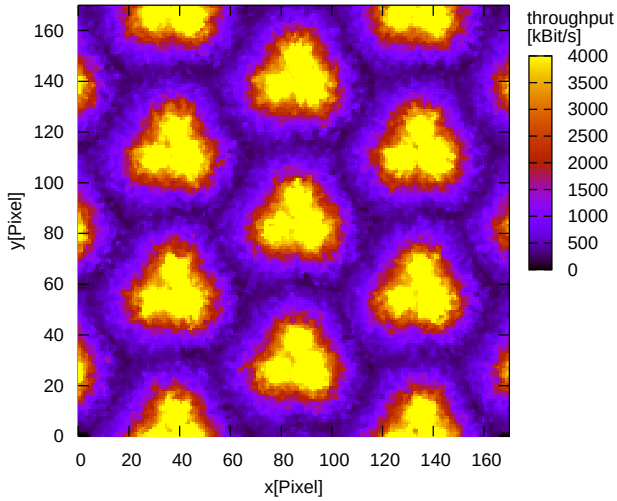



**Figure 7.19:** SINR-based FFR (coordinated): area-dependent throughput mean, $th_{low} = 15$ dB, $th_{up} = 25$ dB, $D_S = 20$ dB.

trade off the aggregate throughput and the cell edge throughput. The SINR-based FFR slightly outperforms the distance-based FFR with respect to both the aggregate throughput and the cell-edge throughput.

The mean SINR within the observation area in Fig. 7.18 shows only marginal differences to the SINR distribution obtained for distance-based FFR in Fig. 7.14. However, a significant difference can be observed for the throughput distribution plotted in Fig. 7.19. In particular, the SINR-based assignment shows a soft degradation of the throughput performance when moving from the cell center to the edge, avoiding a sharp edge as with the distance-based FFR.

Summarizing the results, the performance of a system with inter-cell coordination can be matched regarding the aggregate sector throughput. With respect to the cell border performance, the performance of the locally coordinated system is significantly worse, as inter-cell coordination allows a much better control of the interference caused by neighboring base stations.

## 7.4 Distributed Graph-based Interference Coordination

This Section evaluates the performance of a system with distributed graph-based interference coordination according to Chapter 5. This algorithm has a large number of configurable parameters. In particular, the involved genetic algorithm offers various degrees of freedom. Section 7.4.1 studies these degrees of freedom and evaluates the performance based on indirect performance metrics. After the identification of good parameters, Section 7.4.2 presents throughput results depending on the configuration of the inner and outer interference graph. Subsequently, the impact of the signaling delay is studied in Section 7.4.3. Finally, Section 7.4.4 studies the update period of the inner interference graph, and Section 7.4.5 elaborates on the impact of the terminal mobility.

### 7.4.1 Configuration of the genetic algorithm

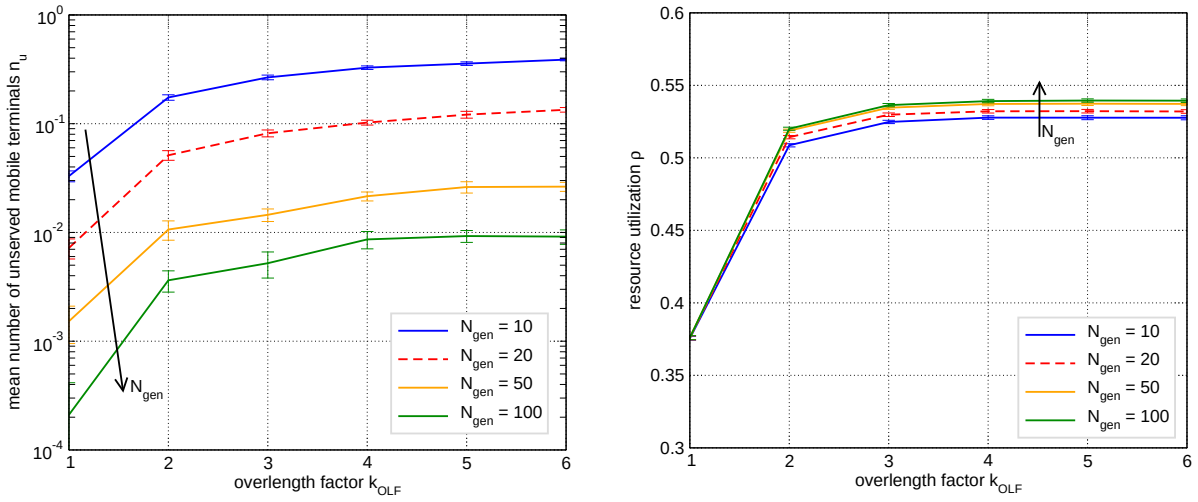
This Section deals with the configuration of the genetic algorithm. Section 7.4.1.1, Section 7.4.1.2, and Section 7.4.1.3 study different genetic representations, namely Random Genomes, Weighted Genomes, and Exact Genomes, respectively. The performance of various parameters of the genetic algorithm is evaluated based on indirect performance metrics. In particular, the resource utilization $\rho$ and the number of unserved mobile terminals $n_u$ are investigated by means of Monte-Carlo simulations (see Section 6.9.1). This allows a computationally much more efficient traversal of the possible parameter range than a full event-driven simulation aiming at throughput results. Note that a higher resource utilization leads to a larger aggregate throughput, while the number of unserved terminals affects the fairness.

All results of this Section were obtained for a configuration of the inner and outer interference graphs of $D_{S,i} = 20$ dB, $d_{ic,i} = 0$, $D_{S,o} = 5$ dB, $d_{ic,o} = 2$. These values were derived from the results obtained for graph-based global coordination with advanced graph generation in Fig. 7.10. The configuration of these parameters will be refined in Section 7.4.2 after the genetic algorithm has been parametrized in this Section. The scaling factor $\kappa$ was chosen larger than $|\boldsymbol{C}|$ in order to assure that the GA focuses on serving all mobile terminals before maximizing the resource utilization. The tournament size of the tournament selection was set to $N_{ts} = 4$.

#### *7.4.1.1 Evaluation of Random Genomes*

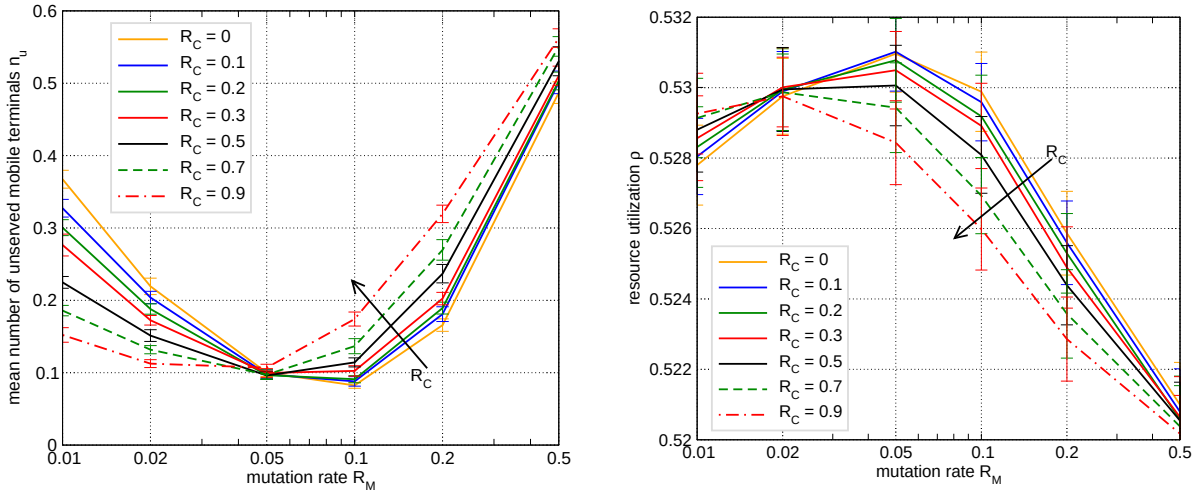
The simplest genetic representation is a Random Genome. It is used with a standard swap mutator, and it can be used with the crossover operators Standard PMX and Random PMX as described in Section 5.3.5. One important parameter in the genome representation is the overlength factor $k_{OLF}$. Figure 7.20 shows the influence of the overlength factor on the algorithm performance. Figure 7.20(a) plots the mean number of unserved mobile terminals $n_u$ in a virtual frame depending on the overlength factor $k_{OLF}$ after different number of generations $N_{gen}$. Alike, the resource utilization $\rho$ is plotted in Fig. 7.20(b). First, it can be observed that a larger overlength factor leads to more unserved mobile terminals in a virtual frame, since it becomes more difficult for the genetic algorithm to rearrange the genes in the genome to form a valid solution to the inner optimization problem. It particularly becomes difficult to meet the constraint of eq. (5.10) which requires every mobile terminal to be served at least once every virtual frame. However, this problem can be counteracted by increasing the number of generations, since long genomes principally allow for the same solutions to the inner optimization problem as short genomes. On the other hand, longer genomes are essential to achieve a good resource utilization, as can be seen from Fig. 7.20(b). Obviously, $k_{OLF}$ needs to be chosen at least equal to 3 in order to achieve a good resource utilization. Smaller values of $k_{OLF}$ can never achieve an as good resource utilization, since every mobile terminal can only be served at most $k_{OLF}$ times in every virtual frame. Consequently, small values of $k_{OLF}$ cannot be compensated by more generations. In the following, all simulations will be done with $k_{OLF} = 3$ if not otherwise noted.

The mutation and crossover rate used above was based on a good guess. A systematic analysis of the parameter range is given in Fig. 7.21(a) for the number of unserved mobile terminals and in Fig. 7.21(b) for the resource utilization. Since it is likely that the genetic algorithm




(a) Mean number of unserved mobile terminals $n_u$ (b) Resource utilization ρ

**Figure 7.20:** Influence of overlength factor $k_{OLF}$ for Random Genomes with Standard PMX. $|\boldsymbol{P}| = 40$, $R_M = 0.1$, $R_C = 0.05$.

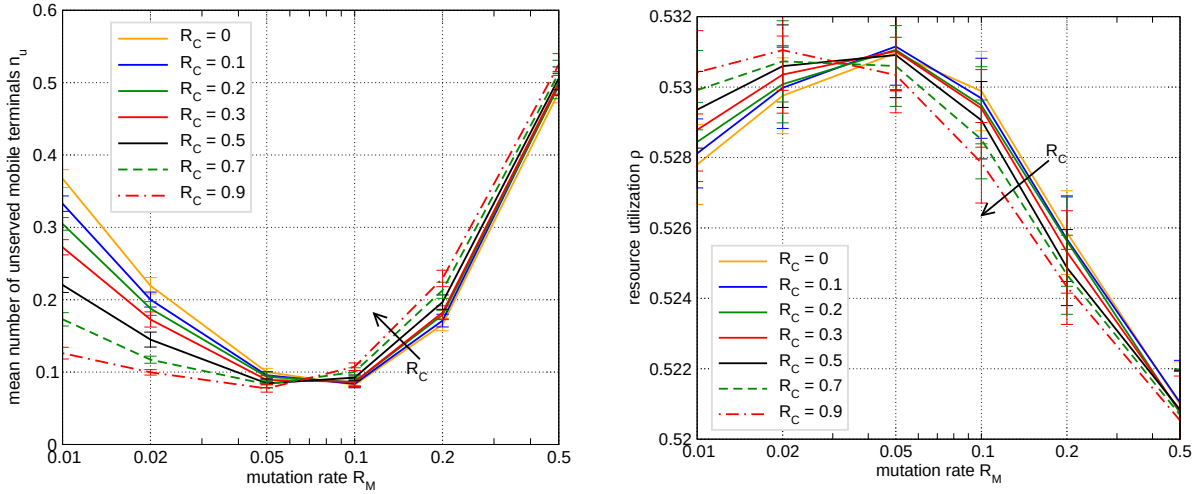



(a) Mean number of unserved mobile terminals $n_u$ (b) Resource utilization ρ

**Figure 7.21:** Choice of mutation and crossover rate for Random Genomes with Standard PMX. $|\boldsymbol{P}| = 40$, $N_{gen} = 20$.

converges to a near-optimal solution for most parameter choices if $|\boldsymbol{P}|$ and $N_{gen}$ are chosen only large enough, the following parameter studies are done with a relatively small population size of $|\boldsymbol{P}| = 40$ and a small number of generations of $N_{gen} = 20$. This allows configuring the genetic algorithm for the case of a limited available compute power while focusing on a fast algorithm convergence.

First, it can be seen that the mutation rate $R_M$ has a bigger influence on the algorithm than the crossover rate $R_C$. With respect to the mutation rate, the best performance is achieved for




(a) Mean number of unserved mobile terminals $n_u$

(b) Resource utilization ρ

**Figure 7.22:** Choice of mutation and crossover rate for Random Genomes with Fully Random PMX. $|\boldsymbol{P}| = 40$, $N_{gen} = 20$.

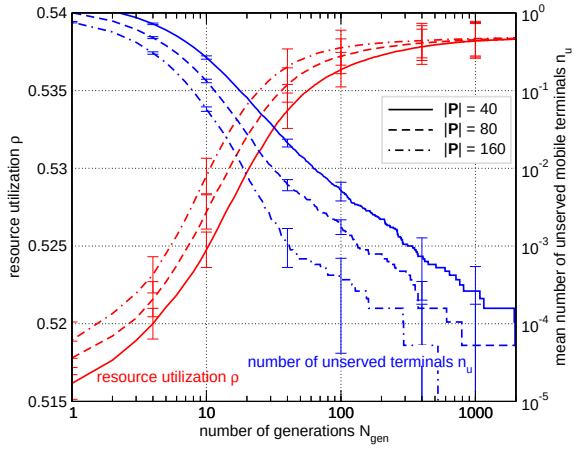



**Figure 7.23:** Convergence behavior of genetic algorithm. $R_M = 0.1$, $R_C = 0.05$.

$R_M \approx 0.05 - 0.1$. In this parameter range, small crossover rates deliver the best performance. In particular, $R_C$ may even be set to 0, which deactivates the crossover operator but still delivers the best performance. Figure 7.22 shows the same results for an identical setup with a Random PMX as described in Section 5.3.5.2. This considerably more complex crossover operator obviously does not achieve any performance improvements and hence is no competition to the much simpler Standard PMX.

A more detailed view on the convergence behavior of the algorithm with Random Genomes and the Standard PMX is given in Fig. 7.23. Plotted is the overall resource utilization ρ and the average number of unserved mobile terminals $n_u$ after a certain number of generations $N_{gen}$. In particular, terminals in unfavorable positions at the cell edge will most likely be unserved for small $N_{gen}$, thus decreasing the cell edge performance. From the graph it can be seen that as few as $N_{gen} = 18$ generations bring the number of unserved terminals below 0.1 at a population size

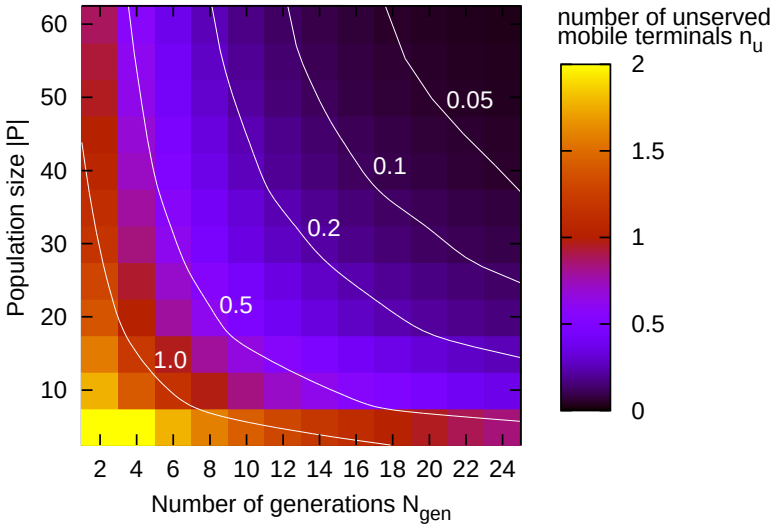
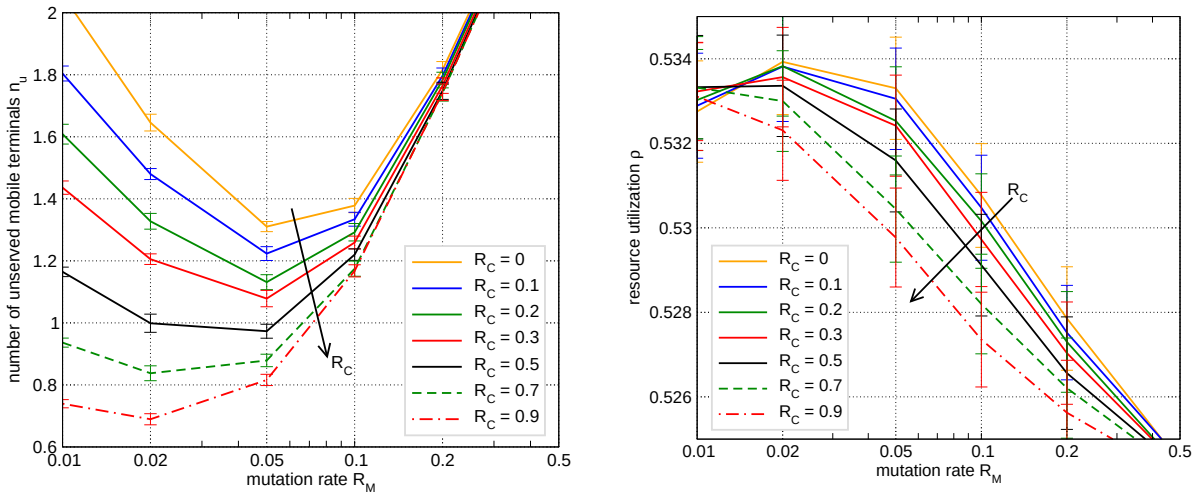



**Figure 7.24:** Mean number of unserved mobile terminals $n_u$ depending on $|\boldsymbol{P}|$ and $N_{gen}$ with Random Genomes and Standard PMX. $R_M = 0.1$, $R_C = 0.05$.

of $|\boldsymbol{P}| = 40$. This means that one terminal remains unserved every 10 MAC frames in average. Hence, compensation for unserved mobile terminals as it was touched on in Section 5.3.7 is not necessary very often, and it will therefore be neglected in the following. For $N_{gen} = 100$, the algorithm already goes into saturation with respect to the resource utilization, while the number of unserved terminals keeps decreasing to negligible values as $N_{gen}$ is increased further. The graph also shows that the resource utilization ρ depends much less on the number of generations than the number of unserved terminals $n_u$. Hence, it can be concluded that this also holds for the aggregate throughput and the cell edge throughput, respectively. A similar observation can be made for the dependence on the population size $|\boldsymbol{P}|$.

The population size and the number of generations are two fundamental parameters in a genetic algorithm. They both have a direct influence on the computational complexity of the genetic algorithm, which is essentially proportional to $N_{gen} \cdot |\boldsymbol{P}|$ if all other parameters remain unchanged. They also directly impact the algorithm performance, where larger values deliver better results. Given a certain maximum computational complexity, both values need to be traded off against each other in order to find the optimal configuration. To do so, Fig. 7.24 plots the average number of unserved mobile terminals $n_u$ depending on $|\boldsymbol{P}|$ and $N_{gen}$ in a two-dimensional chart. It first becomes obvious that both $|\boldsymbol{P}|$ and $N_{gen}$ need to be chosen above a certain threshold, since the number of mobile terminals $n_u$ is relatively high if either of them is chosen too small. Furthermore, it can be seen that both parameters are well traded off when choosing $|\boldsymbol{P}| \approx 2N_{gen}$.

Summarizing these results, it becomes obvious that a relatively small number of generations with a medium population size suffices to achieve a decent performance. This makes the algorithm well implementable and motivates the choice of a relatively small population size of $|\boldsymbol{P}| = 40$ in combination with a small number of generations $N_{gen} = 20$ in the parameter studies.




(a) Mean number of unserved mobile terminals $n_u$

(b) Resource utilization ρ

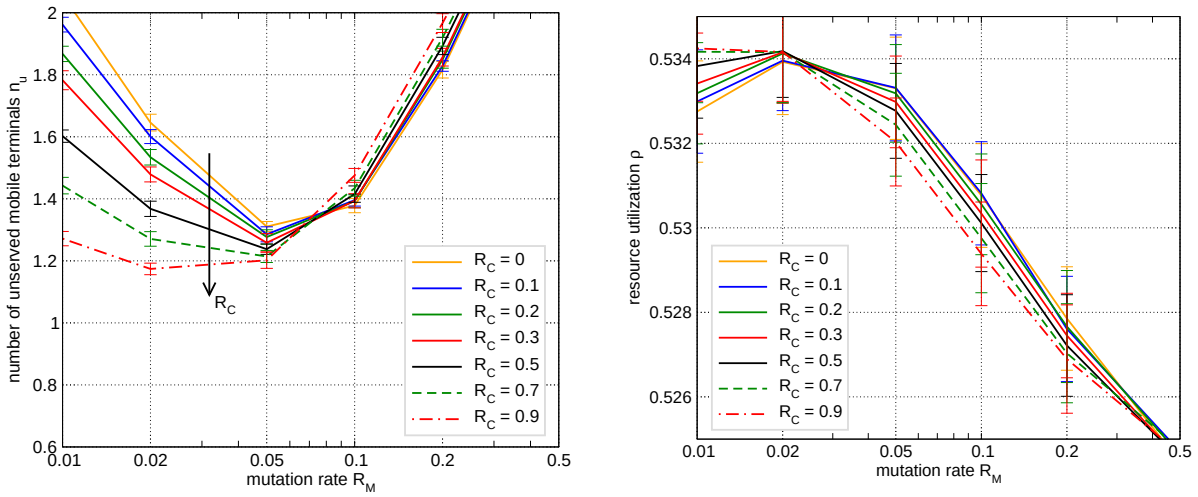
**Figure 7.25:** Parameter choice for Weighted Genomes with Standard PMX. $|\boldsymbol{P}| = 40$, $N_{gen} = 20$.

### *7.4.1.2 Evaluation of Weighted Genomes*

The second basic genetic representation introduced in Section 5.3.1 are Weighted Genomes. Parameter studies for the mutation and crossover rate are shown in Fig. 7.25 and in Fig. 7.26 for the Standard PMX and the Random PMX, respectively. When comparing these results to the results obtained with Random Genomes in Fig. 7.21 and Fig. 7.22, it quickly becomes obvious that a modeling approach based on Weighted Genomes cannot match the model with Random Genomes irrespective of the parameter choice. Further studies reveal that for large enough values of $|\boldsymbol{P}|$ and $N_{gen}$, the performance improves and produces acceptable results with respect to ρ and $n_u$. However, it is obvious that the convergence behavior is far worse than in the case of Random Genomes, and hence, the computational complexity for achieving a comparable performance is much higher.

### *7.4.1.3 Evaluation of Exact Genomes*

The last investigated modeling approach are Exact Genomes. This modeling approach requires specifically adapted mutation and crossover operators, which were described in Sections 5.3.4 and 5.3.5.3, respectively. Besides the mutation and crossover rate, the new mutation operator introduces an additional degree of freedom, namely the swap-in probability $R_{SI}$. Figure 7.27 illustrates the performance with Exact Genomes by plotting the resource utilization ρ over the mean number of unserved terminals $n_u$ for different values of $R_M$, $R_C$, and $R_{SI}$. The graph shows four sets of curves, one for every considered mutation rate $R_M$. Every set contains four curves, one for every considered crossover rate $R_C$ between 0 and 0.2. The curves itself consist of seven data points, one for every considered swap-in probability $R_{SI}$ between 0 and 0.5. Best results are achieved for mutation rates between $R_M = 0.05$ and $R_M = 0.1$, which match the results obtained with Random and Weighted Genomes. As before, small crossover rates achieve better results. In particular, the deactivation of the crossover operator ($R_C = 0$) results in the best




(a) Mean number of unserved mobile terminals $n_u$

(b) Resource utilization ρ

**Figure 7.26:** Parameter choice for Weighted Genomes with Fully Random PMX. $|\boldsymbol{P}| = 40$, $N_{gen} = 20$.

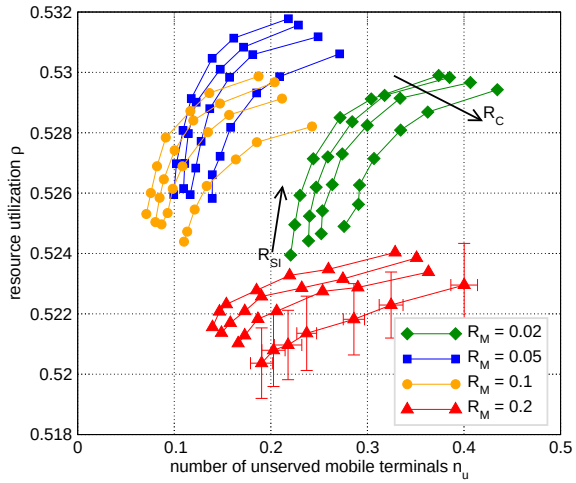
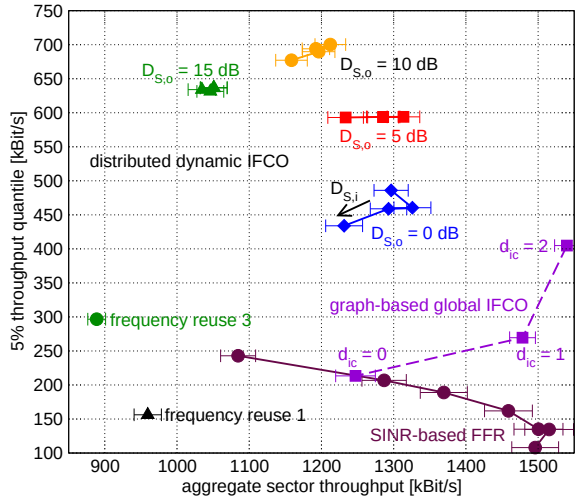



**Figure 7.27:** Resource utilization ρ vs. mean number of unserved mobile terminals $n_u$ for Exact Genomes. $R_C = \{0, 0.05, 0.1, 0.2\}$, $R_{SI} = \{0, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5\}$.

performance, which indicates that the proposed counting crossover operator is not perfect. The swap-in probability $R_{SI}$ allows to trade off ρ and $n_u$.

Compared to the performance with Random Genomes and a Standard PMX in Section 7.4.1.1, no significant performance improvement becomes obvious. On the other hand, Exact Genomes come along with a much higher computational cost due to the involved complicated mutation and crossover operators. Hence, they do not seem like a serious alternative to the amazingly simple Random Genomes, which match the performance of Exact Genomes and outperform the previously studied Weighted Genomes.




| $D_{S,o}$ | $D_{S,i}$ | aggregate throughput | 5% throughput quantile |
|---|---|---|---|
| 5 dB | 15 dB | 1285 kBit/s | 594 kBit/s |
| 5 dB | 20 dB | 1314 kBit/s | 594 kBit/s |
| 5 dB | 25 dB | 1286 kBit/s | 594 kBit/s |
| 5 dB | 30 dB | 1233 kBit/s | 593 kBit/s |
| 10 dB | 15 dB | 1192 kBit/s | 694 kBit/s |
| 10 dB | 20 dB | 1212 kBit/s | 700 kBit/s |
| 10 dB | 25 dB | 1196 kBit/s | 690 kBit/s |
| 10 dB | 30 dB | 1159 kBit/s | 677 kBit/s |

**Figure 7.28:** 5% throughput quantile vs. aggregate throughput for distributed graph-based coordination depending on $D_{S,0}$ and $D_{S,i}$. $t_{C,delay} = 0.5$ s and $t_{C,up} = 0.5$ s. Note that the data points of graph-based global IFCO for different $d_{ic}$ are connected only for illustrative purposes.

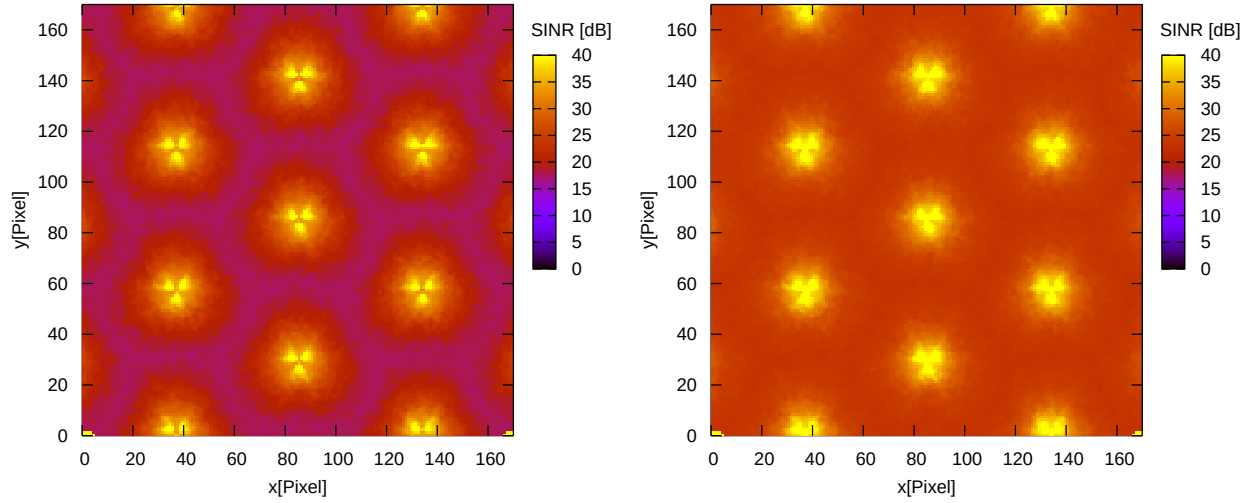
#### *7.4.1.4 Conclusion*

As a consequence from the previous studies, Random Genomes in combination with a Standard PMX will be used for all further simulations. The mutation rate is set to $R_M = 0.1$, and the crossover rate is set to $R_C = 0$. This essentially disables the crossover operator, which has the side effect of significantly reducing the algorithm complexity, in particular with respect to possible hardware implementations as discussed in Section 5.3.8. Furthermore, a small population size of $|\boldsymbol{P}| = 40$ with a small number of generations of $N_{gen} = 20$ will be used, since this is sufficient to achieve a good performance.

### 7.4.2 Throughput results

After the previous studies with roughly estimated configurations of $D_{S,o}$ and $D_{S,i}$, this Section presents a systematic analysis of the throughput performance depending on the configuration of the inner and outer interference graphs. In all cases, the coordination diameter is set to $d_{ic,o} = 2$ due to the results obtained in Section 7.2.2. Figure 7.28 plots the 5% throughput quantile over the aggregate sector throughput for different values of $D_{S,o}$ and $D_{S,i}$. The update period $t_{C,up}$ was set to 0.5 s, and the signaling delay $t_{C,delay}$ to 0.5 s. It turns out that the initial guesses for $D_{S,o}$ and $D_{S,i}$ taken from Fig. 7.10 were very good. Near optimal configurations of these parameters can be taken from Fig. 7.28 depending on whether the cell edge performance or the aggregate performance should be maximized. Within the considered configurations in Fig. 7.28, the cell edge performance is maximized for the choice of $D_{S,o} = 10$ dB. The aggregate throughput performance is maximized for $D_{S,o} = 5$ dB, resulting in a 12% penalty for the cell edge throughput. In both cases, $D_{S,i} = 20$ dB delivers the best results.

The chart also contains the results of the reference systems. For the SINR-based FFR and the graph-based global IFCO, the best results according to the studies in the previous Sections are plotted. Compared to the uncoordinated reference systems, a significant performance gain is




(a) $D_{S,o} = 5$ dB (b) $D_{S,o} = 10$ dB

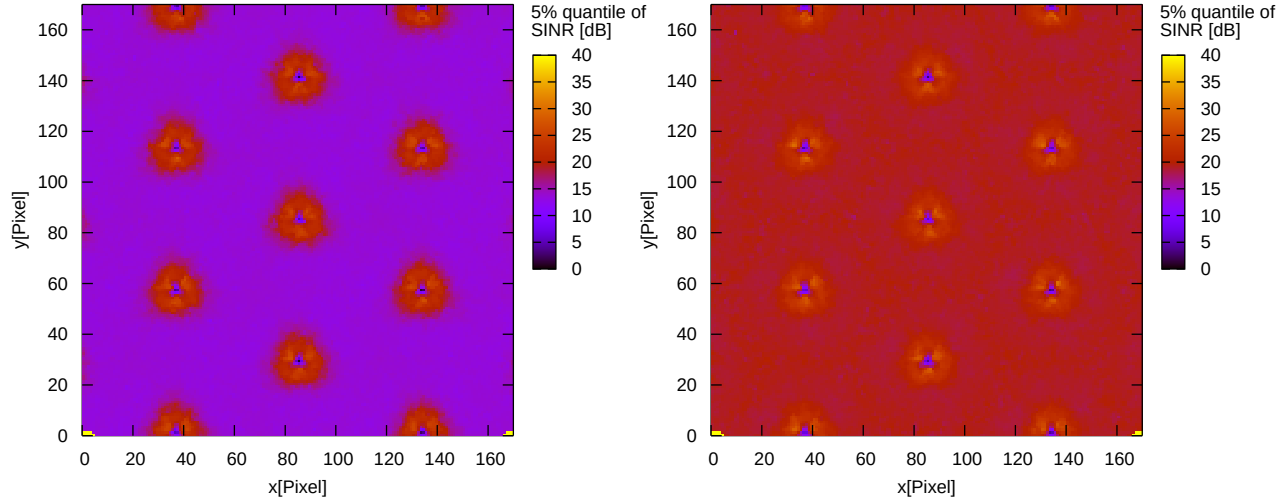
**Figure 7.29:** Area-dependent SINR mean for graph-based distributed coordination. $D_{S,i} = 20$ dB.

achieved with respect to both the aggregate throughput and the throughput quantile. Specifically, compared to the reuse 3 system, the graph-based distributed coordination scheme increases the 5% throughput quantile by more than 130% together with an increase of more than 30% in the aggregate throughput performance if $D_{S,o} = 10$ dB. For $D_{S,o} = 5$ dB, the increase in the 5% throughput quantile is about 100% with a gain of almost 50% in the aggregate throughput. Compared to the graph-based global IFCO, a significant gain in the 5% throughput quantile can be observed, while the aggregate throughput cannot be matched.

Figure 7.29 plots the distribution of the mean SINR within the observation area for $D_{S,o} = 5$ dB and $D_{S,o} = 10$ dB. For $D_{S,o} = 10$ dB, Fig. 7.29(a) shows a better SINR in the cell border areas compared to the case of $D_{S,o} = 5$ dB in Fig. 7.29(b), which goes along with the higher measured throughput quantile for $D_{S,o} = 10$ dB. With respect to the aggregate throughput, no direct conclusion can be drawn from the comparison of Fig. 7.29(a) and Fig. 7.29(b), since the SINR with $D_{S,o} = 10$ dB is better at any location than with $D_{S,o} = 5$ dB.

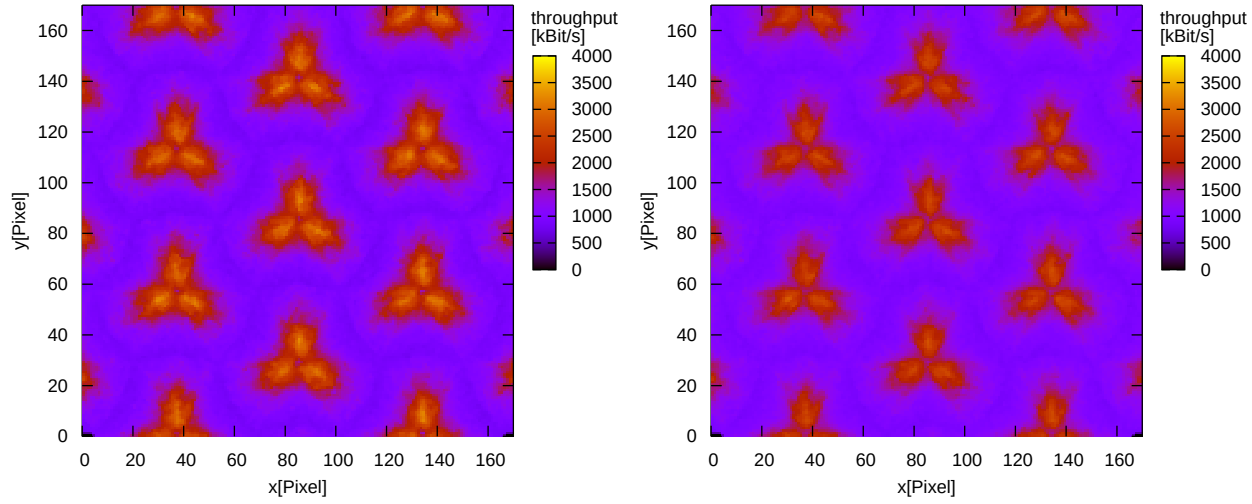
Figure 7.30 plots the distribution of the 5% quantile of the SINR within the observation area for the same configurations. The most important observation is the uniform SINR quantile in large portions of the coverage area. Only close to the base stations does the quantile increase. These uniform conditions prove that the inner optimization actually obeys to the information provided by the central coordinator, which itself is based on a global interference graph and which is able to avoid any inter-cell interference up to a certain desired minimum SIR $D_{S,o}$. Consequently, the SINR quantile in-between the base stations is about 5 dB larger in 7.30(b) compared to 7.30(a).

Finally, Fig. 7.31 plots the throughput distribution within the observation area. Comparing Fig. 7.31(a) with Fig. 7.31(b) clearly reveals the higher aggregate throughput for $D_{S,o} = 5$ dB, since the red high-throughput areas close to the base stations are larger for $D_{S,o} = 5$ dB. The difference in the cell edge throughput can also be observed in the figures, though the difference of 17% in the throughput quantile is not as clearly visible in the colored charts of Fig. 7.31.




(a) $D_{S,o} = 5$ dB (b) $D_{S,o} = 10$ dB

**Figure 7.30:** Area-dependent 5% quantile of SINR for graph-based distributed coordination. $D_{S,i} = 20$ dB.

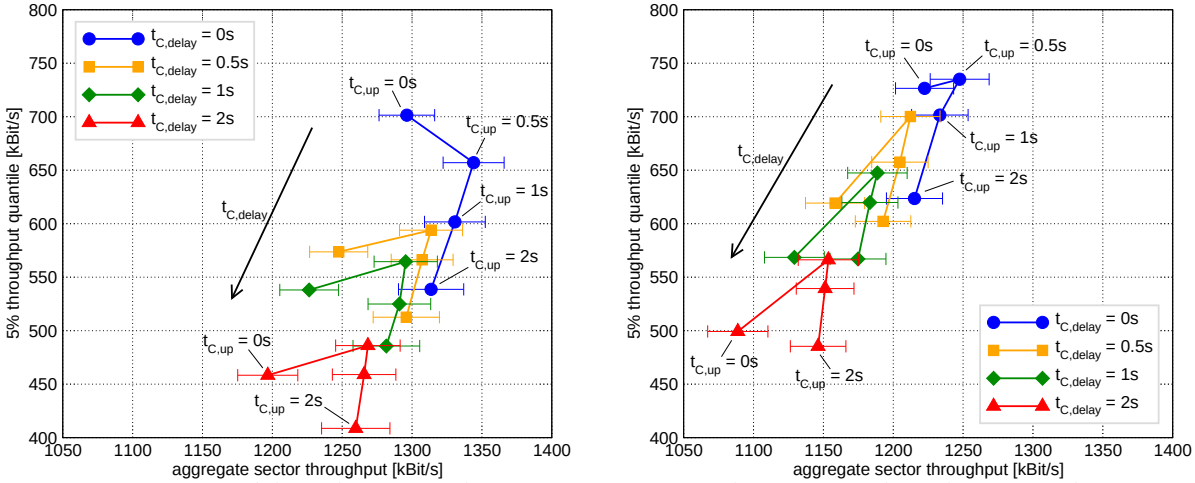



(a) $D_{S,o} = 5$ dB (b) $D_{S,o} = 10$ dB

**Figure 7.31:** Area-dependent throughput mean for graph-based distributed coordination. $D_{S,i} = 20$ dB.

### 7.4.3 Impact of signaling delay and update period with central coordinator

It was discussed in Section 5.7.2 that the signaling delay $t_{C,delay}$ and the update period $t_{C,up}$ have a significant impact on the system performance. This impact is illustrated in Fig. 7.32 for $D_{S,o} = 5$ dB and $D_{S,o} = 10$ dB by plotting the 5% throughput quantile over the aggregate sector throughput for different values of $t_{C,delay}$ and $t_{C,up}$.




(a) $D_{S,0} = 5$ dB and $D_{S,i} = 20$ dB (b) $D_{S,0} = 10$ dB and $D_{S,i} = 20$ dB

**Figure 7.32:** 5% throughput quantile vs. aggregate throughput for distributed graph-based coordination depending on signaling delay $t_{C,delay}$ and update period $t_{C,up}$.

A larger signaling delay $t_{C,delay}$ leads to a performance decrease both with respect to the 5% throughput quantile and the aggregate sector throughput. This is intuitive, since the information from the central coordinator becomes outdated as $t_{C,delay}$ increases. The performance behavior cannot be explained so easily when varying the update period $t_{C,up}$. It can be seen that the system performance increases when increasing $t_{C,up}$ from 0 s to 0.5 s. Only when increasing $t_{C,up}$ further does the performance deteriorate, as one would expect. The slight increase when moving from 0 s to 0.5 s can be traced back to side effects of the burst profile management. The biggest problem that the burst profile management encounters are SINR fluctuations. These are particularly high in a system with beamforming antennas, since the interference that a mobile terminal receives from a particular cell sector strongly depends on the mobile terminal served in that sector on the same resources. For $t_{C,up} = 0$ s, the resource occupation changes in every virtual frame. For $t_{C,up} = 0.5$ s, the resource occupation changes only every 500 ms. This implies a stable resource occupation for 500 ms and consequently, there are no SINR fluctuations if a mobile terminal is served exactly once in a virtual frame. For mobile terminals that are served multiple times in a virtual frame, SINR variations persist within the virtual frame. Fewer SINR fluctuations allow the burst profile management to adjust the MCS more precisely, leading to an increased throughput performance. This circumstance can also be observed in Fig. 7.33, which plots the loss probability of data bursts depending on the update period $t_{C,up}$ for different configurations. As a reference, the loss probability for an uncoordinated reuse 3 system with beamforming antennas is plotted, too. The loss probability in Fig. 7.33 is defined as the overall ratio of unsuccessfully received data bursts to the overall number of transmitted data bursts for a particular mobile terminal, including all unsuccessful retransmissions of the HARQ. The chart confirms the explanation from above, since the loss probability decreases when increasing $t_{C,up}$ from 0 s to 0.5 s. Consequently, the system performance increases. When $t_{C,up}$ is increased further, the impact of outdated information from the central coordinator starts to dominate, leading to a decrease in performance.

Note that an increased signaling delay or update period has a bigger effect on the cell edge performance compared to the aggregate throughput performance, since the information from the central coordinator is mainly responsible for the cell edge throughput.

### 7.4.4 Impact of update period of inner interference graph

In the previous Section, only the signaling delay and update period towards the central coordinator were considered. The inner interference graph was assumed to be updated at the beginning of every virtual frame. In reality, the inner interference graph may be subject to longer update periods as well. The system performance depending on the update period of the inner interference graph $t_{C,innerup}$ is plotted in Fig. 7.34. In particular, Fig. 7.34(a) and Fig. 7.34(b) plot the aggregate sector throughput and the 5% throughput quantile, respectively, for different system configurations. In all cases, $t_{C,innerup} \leq t_{C,up}$ since the inner interference graph can at least be updated as frequently as the outer interference graph. Both figures show virtually no performance deterioration as $t_{C,innerup}$ is increased from 0 s to $t_{C,up}$. The reason for this is that the performance is already dominated by the update period of the outer interference graph.

### 7.4.5 Impact of terminal mobility

All simulations so far were carried out with a terminal mobility of 30 km/h. Besides such a mobile scenario, a nomadic usage is an important use case in particular for IEEE 802.16e networks. In this case, the mobile terminals remain stationary for a certain period of time before they are carried to a new location by their user. It is obvious that in such a relatively static scenario the signaling aspects of Section 7.4.3 are not relevant. Consequently, the performance gains achievable by the system will not deteriorate even for high signaling delays or high update periods. Moreover, the measurement of relevant parameters such as signal levels (see Section 5.6) becomes much easier and can be done with a much lower frequency compared to a fully mobile scenario.

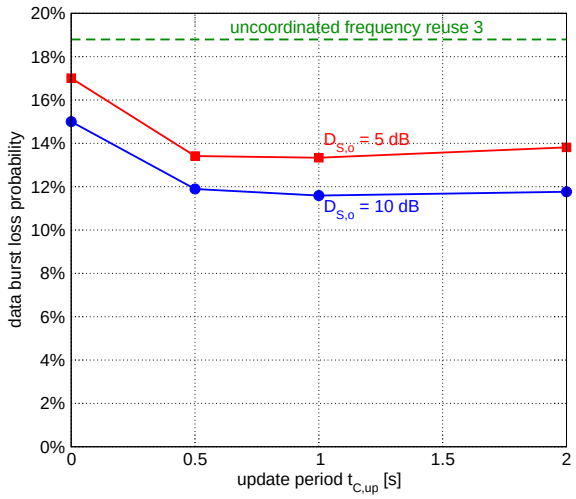
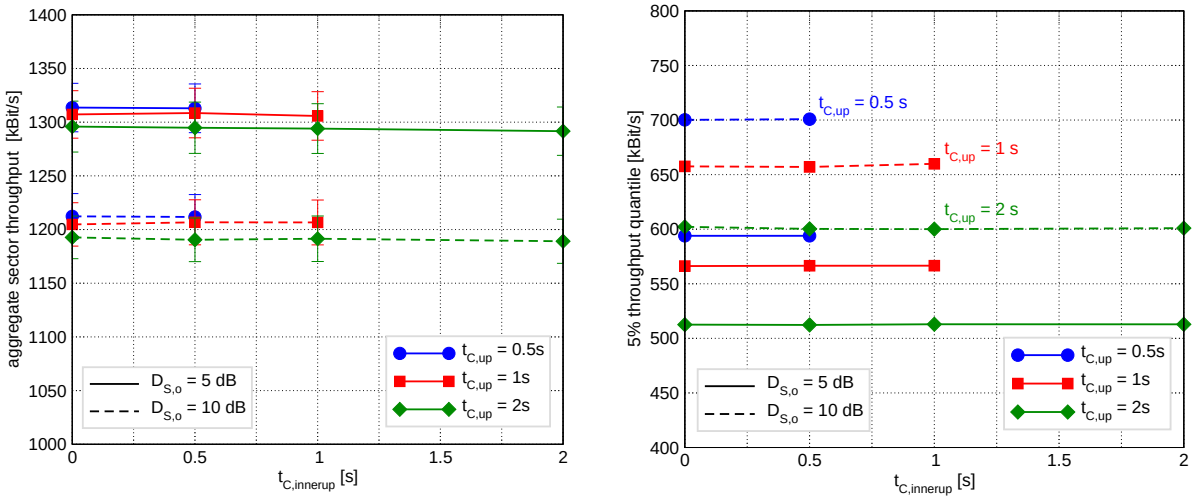



**Figure 7.33:** Loss probability of data bursts depending on $t_{C,up}$. $t_{C,delay} = 0$, $D_{S,i} = 20$ dB.




(a) influence on aggregate sector throughput

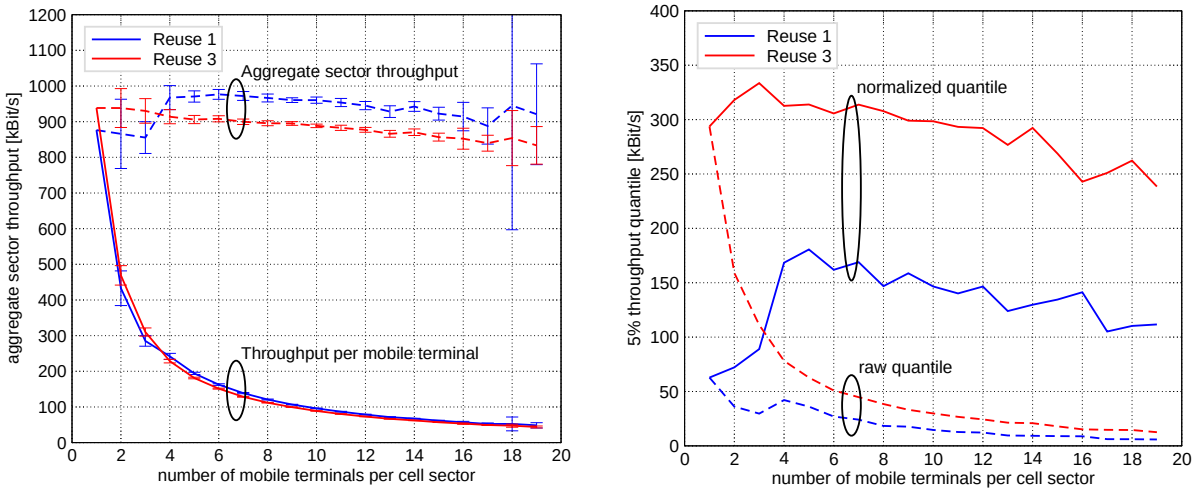
(b) influence of 5% throughput quantile

**Figure 7.34:** Influence of inner update period $t_{C,innerup}$ on 5% throughput quantile and aggregate throughput. $D_{S,i} = 20$ dB, $t_{C,delay} = 0.5$ s.

## 7.5 Poisson User Distribution

All previous simulations were based on a scenario with a uniform terminal distribution, i.e., the number of mobile terminals per cell sector was constant and uniform with $N = 9$. In this Section, a non-uniform terminal distribution according to a Poisson distribution as it was described in Section 6.2 will be considered. The mean number of mobile terminals per cell sector remains unchanged at $N = 9$. It was said in Section 6.8 that in a reuse 1 scenario, the permutation zone is divided into four resource partitions. In a reuse 3 scenario, six resource partitions are used, where two reuse partitions are assigned to every cell sector in order to achieve the frequency reuse factor of 3. Each reuse partition can be used by one mobile. In the case of a non-uniform terminal distribution, cell sectors with less than four (reuse 1) or two (reuse 3) terminals must be allowed to assign more than one resource partition to a mobile terminal in order to achieve a full resource utilization in that sector. For example, a cell sector with only one mobile terminal is allowed to utilize all available resource partitions for that mobile terminal.

The aggregate sector throughput with a non-uniform terminal distribution is illustrated in Fig. 7.35(a) for an uncoordinated frequency reuse 1 and an uncoordinated frequency reuse 3 system. The throughput results are broken down into throughput values obtained in cell sectors with a particular number of mobile terminals on the x-axis of Fig. 7.35(a). Furthermore, the aggregate sector throughput is plotted together with the throughput per mobile terminal, which is obtained by dividing the aggregate throughput by the number of mobile terminals in the respective sector. The chart shows that the aggregate sector throughput is independent of the number of mobile terminals in a cell sector. This is logical, since all available resources can be utilized in all cell sectors. Consequently, the throughput per mobile terminal decreases as the number of mobile terminals increases.



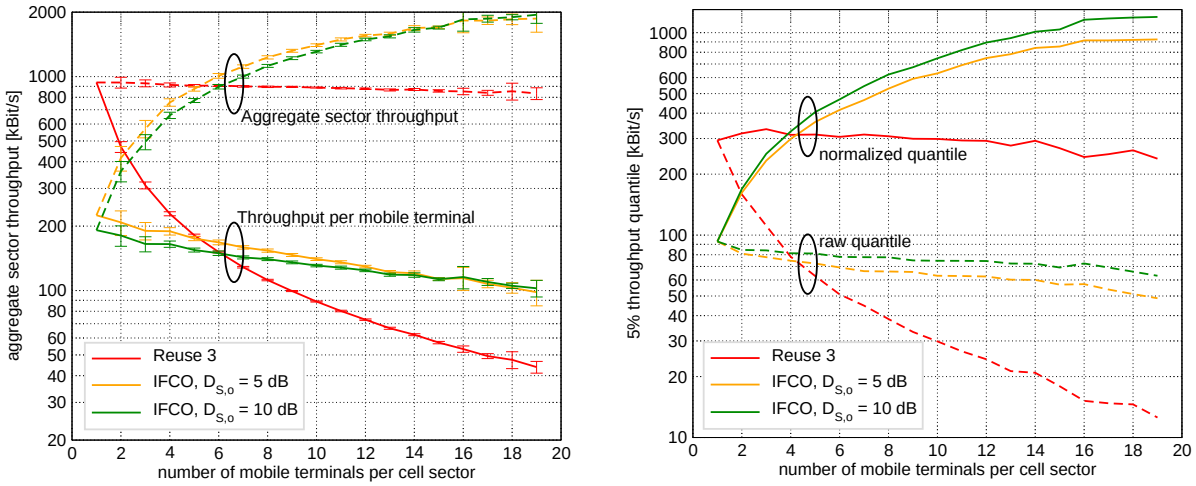

(a) Aggregate sector throughput

(b) 5% throughput quantile

**Figure 7.35:** System performance with Poisson user distribution in cell sectors with $N$ mobile terminals for uncoordinated reuse 1 and reuse 3 reference systems.

Figure 7.35(b) plots the 5% throughput quantile. The raw quantile is the actual 5% throughput quantile as reported by the mobile terminals. The normalized quantile is the same value multiplied by the number of mobile terminals in the cell sector. As the number of mobile terminals increases, the raw quantile decreases significantly, while the normalized quantile decreases only slightly. For the frequency reuse 1 system, the quantile shows a significant drop for $N \leq 3$. The reason for this behavior can be found in the resource allocation strategy. For $N \geq 4$ mobile terminals per cell sector, the permutation zone is subdivided into $N_F = 4$ resource partitions. Every resource partition can thereby be allocated to the transmission for one mobile terminal. For $N = 2$ and $N = 3$ mobile terminals, the permutation zone is subdivided into $N_F = 2$ resource partitions, and for $N = 1$, it is subdivided into $N_F = 1$ partitions in order to achieve a resource utilization of $\rho = 1$. Larger resource partitions imply longer data bursts which consist of more FEC blocks. Consequently, at the same SINR and MCS, a longer data burst is lost with a higher probability. This effect is particularly visible in low SINR conditions since the higher loss probability at some point cannot be compensated by a stronger MCS. Consequently, it has a bigger effect on the 5% throughput quantile than on the aggregate throughput, where the effect is also visible but not overly pronounced.

The performance of the graph-based distributed interference coordination in combination with a non-uniform terminal distribution is shown in Fig. 7.36. The charts show a fundamentally different behavior for the interference coordinated system compared to the uncoordinated frequency reuse 3 system. For small numbers of mobile terminals per cell sector, the performance of the interference coordinated system is lower than the performance of the uncoordinated reuse 3 system, both with respect to the aggregate throughput and the 5% throughput quantile. As the number of mobile terminals in a cell sector increases, the interference coordinated system outperforms the uncoordinated system. It is interesting to note that the performance per mobile terminal far less depends on the number of mobile terminals in the cell sector as it does in the uncoordinated system. This illustrates an important effect mentioned before, namely the ability




(a) Aggregate sector throughput

(b) 5% throughput quantile

**Figure 7.36:** System performance with Poisson user distribution in cell sectors with $N$ mobile terminals for system with distributed graph-based IFCO. $T_{C,delay} = 0.5$ ms, $T_{C,up} = 0.5$ s, $T_{C,innerup} = 0$ s, $D_{S,i} = 20$ dB.

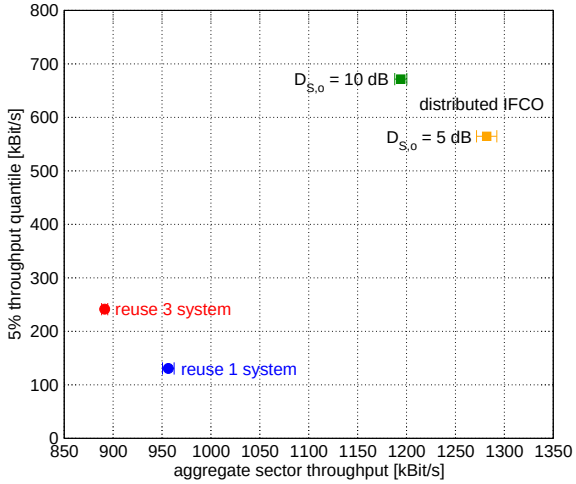



**Figure 7.37:** 5% throughput quantile vs. aggregate sector throughput for non-uniform distribution of mobile terminals. $t_{C,delay} = 0.5$ s, $t_{C,up} = 0.5$ s.

to perform load balancing with respect to the assigned resources between differently loaded sectors.

After this detailed analysis, Fig. 7.37 concludes this Section by comparing the average throughput quantile and aggregate throughput values for the just considered system. The graph shows that the performance gains of the graph-based distributed interference coordination can be maintained even with non-uniform terminal distributions in the same order of magnitude as with a uniform terminal distribution.

## 7.6 On-Off Traffic Sources

In this final Section, the performance of the distributed IFCO algorithm given a traffic load according to the on-off traffic model from Section 6.5 will be evaluated. Empty input buffers of mobile terminals are treated according to the third option discussed in Section 5.4.3, which is very easy to realize and requires only minor modifications to the so far used algorithms. For the uplink direction, ARQ status reporting was disabled in order to avoid small amounts of data to be transmitted in the downlink direction during the off-periods of the traffic model. In order to measure the 5% throughput quantile, measurement samples were only considered if they were larger than 0.

The mean duration of the on-off period was set to $E[T_P] = 20$ s. The signaling delay with the central coordinator was set to $T_{C,delay} = 0.5$ s, and the update period to $T_{C,up} = 0.5$ s. The update period of the inner interference graph was set to $T_{C,innerup} = 0$ s, and the desired minimum SIR of the inner graph was set to $D_{S,i} = 20$ dB.

If the buffer for mobile terminal $m_i$ runs empty, it takes until the next coordination cycle for the respective color set $\boldsymbol{C}_i$ to be cleared. During that time, the virtual frame duration is longer than necessary, and the inner optimization should be able to fill up gaps in the resource occupation of a virtual frame caused by terminal $m_i$ with other terminals. Therefore, the overlength factor $k_{OLF}$ needs to be chosen larger than before. This is illustrated in Fig. 7.38, which plots the aggregate sector throughput depending on the overlength factor for different values of $D_{S,o}$ and $A_F$. Based on these results and to be on the safe side for filling up all gaps, a value of $k_{OLF} = 6$ is chosen for all following simulations with the on-off traffic scenario.

The aggregate sector throughput depending on the activity factor $A_F$ is plotted in Fig. 7.39 for the interference coordinated reuse 1 systems with $D_{S,o} = 5$ dB and $D_{S,o} = 10$ dB, and for the uncoordinated reference systems with frequency reuse 1 and 3. For $A_F = 1$, the scenario is equivalent to a full buffer scenario and hence delivers about the same results as seen in Section 7.4.2. For $A_F = 0$, there is no data to transmit and hence no throughput is observed. From Fig. 7.39 it can be seen that the performance of the coordinated systems falls behind the per-

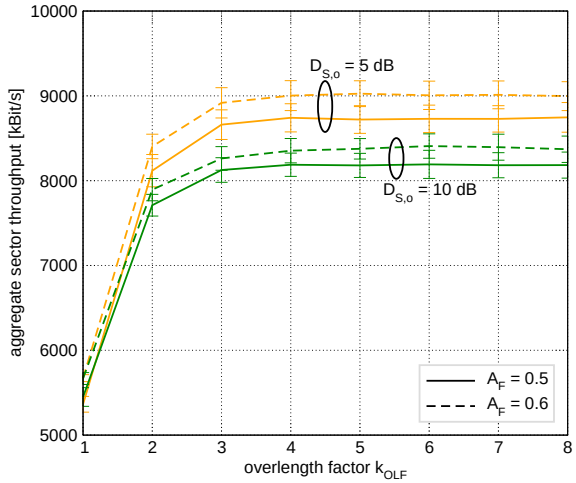



**Figure 7.38:** Aggregate sector throughput depending on overlength factor $k_{OLF}$.

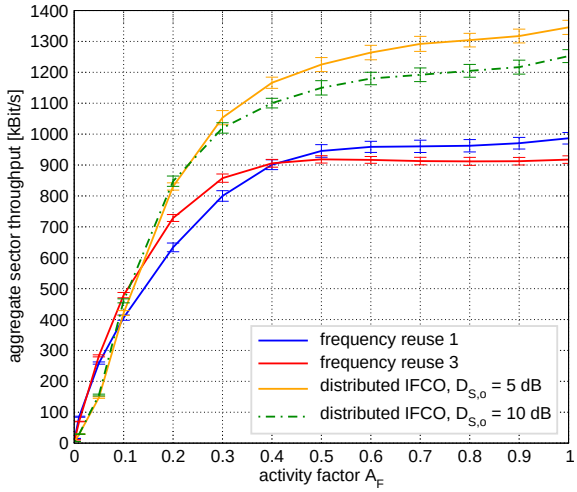



**Figure 7.39:** Aggregate sector throughput depending on activity factor $A_F$.

formance of the uncoordinated reference systems only for $A_F < 0.1$. For $A_F \geq 0.3$, significant performance gains can be observed which are in the same order of magnitude as in the full buffer scenario. Nevertheless, while the uncoordinated reference systems deliver almost the same aggregate throughput for any value of $A_F$ greater than 0.5, the coordinated systems show a small performance decrease as $A_F$ is decreased from $A_F = 1.0$. There are two reasons for this behavior. First, there are packets that remain in the ARQ retransmission buffer once the input buffer queue has been drained. These retransmissions consume only a small amount of air interface resources, but they require resources to be allocated by the central coordinator. Second, as $A_F$ gets smaller, there are more mobile terminals that are randomly assigned to resources thus undermining the coordination. This is illustrated by the average number of randomly assigned mobile terminals plotted in Fig. 7.40. Both effects reduce the aggregate throughput as $A_F$ decreases.

Figure 7.41 furthermore plots the resource utilization ρ depending on $A_F$ for the considered scenarios. For the coordinated systems, the maximum resource utilization is observed for values of $A_F$ between 0.5 and 0.7, while it gets smaller as $A_F$ approaches 1. This correlates with the number of randomly assigned terminals in Fig. 7.40, since randomly assigned terminals increase the resource utilization at the expense of the SIR. For the uncoordinated systems, the resource utilization remains constant at the respective maximum level starting at a particular $A_F$. For the reuse 3 system, the maximum ρ can be observed from about $A_F = 0.5$ onwards, while for the reuse 1 system the maximum ρ is achieved at about $A_F = 0.7$. This is comprehensible, since in a reuse 1 system, more resource partitions have to be filled in a MAC frame (namely 4 resource partitions), than in a reuse 3 system (namely only 2). Therefore, the full resource utilization can be achieved in a frequency reuse 3 system with fewer simultaneously active data flows. Consequently, as $A_F$ is increased, the throughput reaches its maximum level earlier for the reuse 3 system as for the reuse 1 system, as can be seen in Fig. 7.39.

The 5% throughput quantile depending on $A_F$ is plotted in Fig. 7.42(a). The graph shows that large performance gains as in the full buffer scenario are only possible with activity factors $A_F$ close to 1. Nevertheless, for smaller activity factors, the 5% throughput quantile always remains competitive to the uncoordinated frequency reuse 1 system, while being outperformed by the

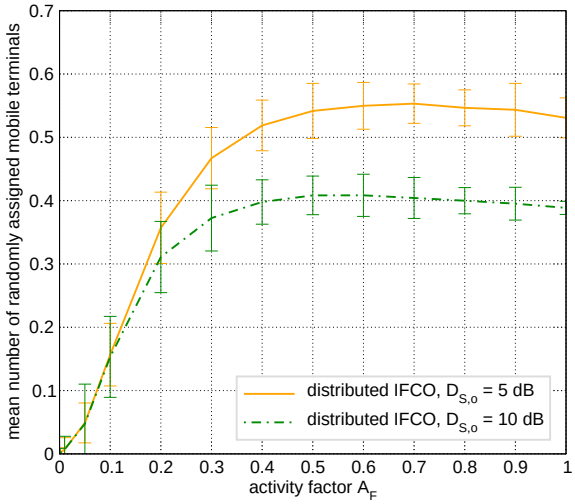



**Figure 7.40:** Mean number of randomly assigned mobile terminals per virtual frame depending on $A_F$.

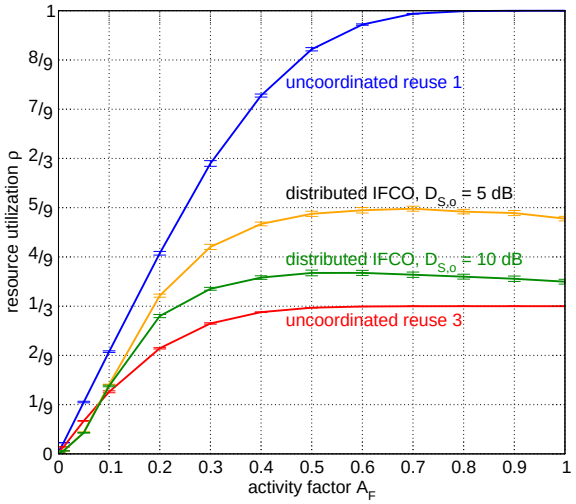
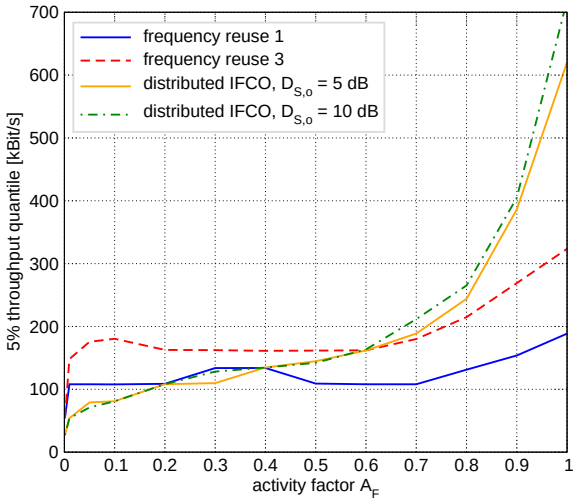
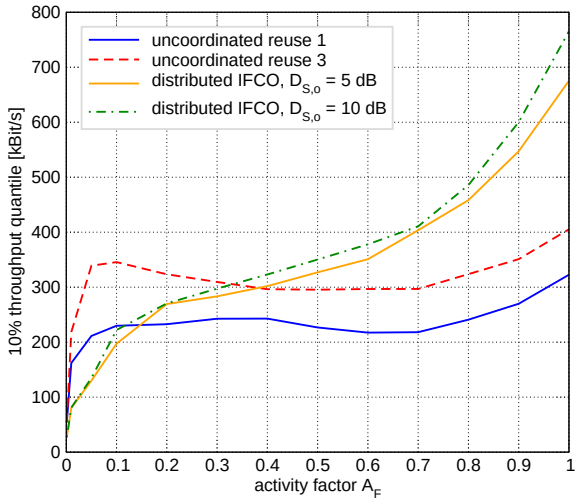



**Figure 7.41:** Mean resource utilization ρ depending on $A_F$.

frequency reuse 3 system for $A_F \leq 0.5$. As mentioned before, only measurement samples that are greater than 0 are considered for measuring the throughput quantile in the case of on-off traffic sources. It was stated in Section 6.9.2 that the short-term period was set to $T_{STP} = 4$ s. If the off-period begins during one of these measurement periods, the resulting measurement sample will be smaller than the actual throughput quantile due to the measurement granularity of 4 s. The same thing applies if the off-period persists during a short-term period, but single packets are delivered at the receiver side that stem from late ARQ retransmissions. This strongly influences the 5% quantile. It is therefore useful to look at the 10% throughput quantile plotted in Fig. 7.42(b), which is not as much affected by these effects. Consequently, the 10% throughput




(a) 5% throughput quantile




(b) 10% throughput quantile

**Figure 7.42:** Throughput quantiles depending on activity factor $A_F$.

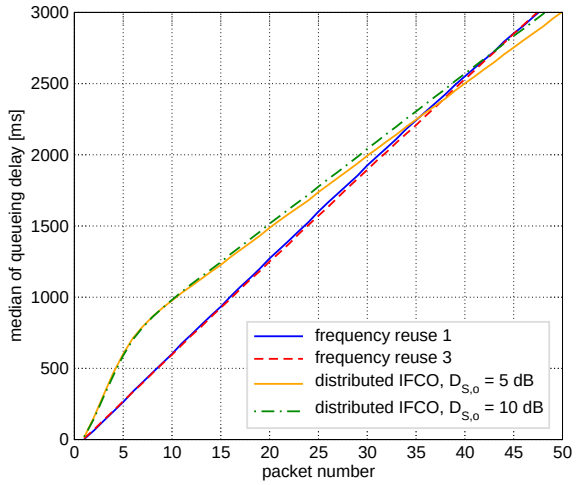


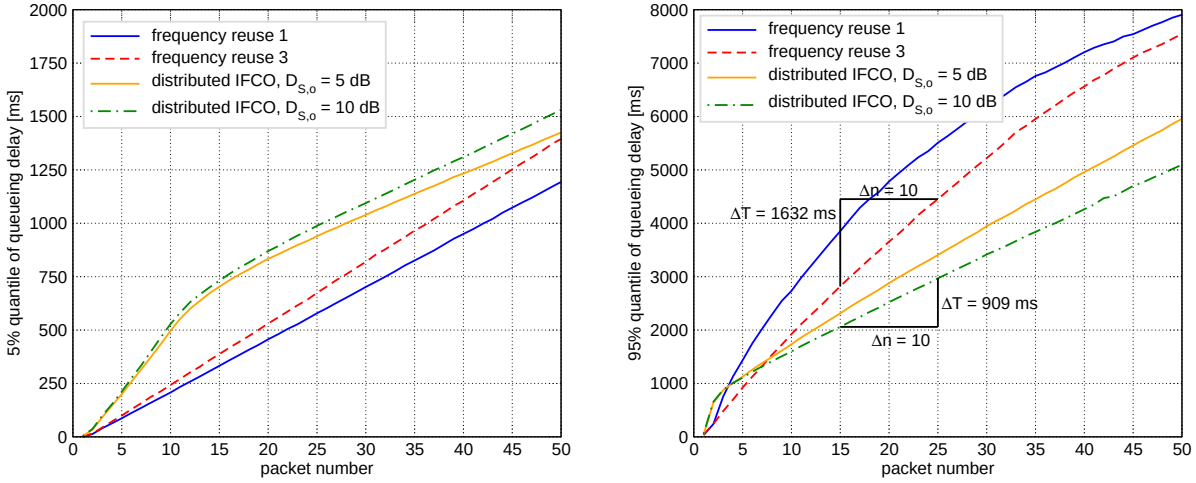

**Figure 7.43:** Median of IP packet queuing delay for $n$-th packet in on-phase. $t_{C,delay} = 0.5$ s, $t_{C,up} = 0.5$ s, $T_{C,innerup} = 0$ ms, $D_{S,i} = 20$ dB.

quantile shows a performance advantage of the coordinated systems already for values of $A_F$ larger than 0.5. Further results regarding the cell edge performance will be derived later from the queuing delay in Fig. 7.44(b).

The above results only consider the steady state of the simulation. However, the transient behavior at the beginning of an on-period is of great interest, since in the beginning a mobile terminal is not included in the coordination process but instead receives randomly chosen resources. In order to evaluate this transient behavior, a new metric is introduced. The new metric measures the time that the $n$-th packet of an on-period spends in the input buffer before it is fetched by lower layers, i.e., the queuing delay of the $n$-th IP data packet of an on-period. The median of the queuing delay is plotted in Fig. 7.43 for the considered scenarios depending on the packet number. It is beneficial to consider the median instead of the mean since the mean is dominated by long queuing delays of cell edge terminals. Cell edge terminals will be considered separately by the 95% queuing delay quantile below.

The inverse of the slope of the curves corresponds to the achieved data rate. Hence, a steeper curve in Fig. 7.43 corresponds to a lower data rate, and a flatter curve corresponds to a higher data rate. First, the median does not reveal any difference between the frequency reuse 1 system and the frequency reuse 3 system, which indicates that the average user experiences the same performance in both systems. The networks with distributed IFCO show a worse throughput performance within the first 750 ms. Afterwards, the slopes of the respective curves decrease, and the throughput is significantly higher than in the uncoordinated systems. Due to the signaling delay of $t_{C,delay} = 500$ ms, all packets transmitted during the first 500 ms are transmitted on randomly assigned resources, where only one resource partition is assigned per virtual frame. Due to the update period of $t_{C,up} = 500$ ms, an additional average of 250 ms elapses until a mobile terminal is participating in the coordination process. The break-even point of the coordinated and the uncoordinated systems is reached between 2.5 and 3 s into the on-period.

The performance of mobile terminals in particularly bad or good reception conditions (cell edge or close to the base station) can be deduced from the 95% and the 5% quantile of the




(a) 5% quantile (good reception conditions)

(b) 95% quantile (bad reception conditions)

**Figure 7.44:** Quantiles of IP packet queuing delay for $n$-th packet in on-phase. $T_{C,delay} = 0.5$ ms, $T_{C,up} = 0.5$ ms, $T_{C,innerup} = 0$ ms, $D_{S,i} = 20$ dB.

queuing delay, respectively. Both are depicted in Fig. 7.44. First of all, both the 5% quantile in Fig. 7.44(a) and the 95% quantile in Fig. 7.44(b) reveal the expected difference between the uncoordinated networks. The uncoordinated reuse 1 system achieves the higher throughput in good reception conditions (eventually leading to a higher aggregate throughput), and the uncoordinated reuse 3 system achieves the better cell edge throughput. The distributed IFCO schemes show the same typical behavior as before, with a bend in the curve at approximately 750 ms. Especially for cell edge terminals, the coordinated systems quickly outperform the uncoordinated reference systems, since the break-even point is already reached between 1 and 1.5 s into the on-period. Note that the slope of the curves decreases as they approach a delay of 10 s, which corresponds to the on-period duration, since no packet may be delayed longer.

The 95% quantile in Fig. 7.44(b) additionally allows a conclusion with respect to the cell edge performance without the errors the direct measurement of the throughput quantiles suffer from. More precisely, the graphs in Fig. 7.44(b) allow the calculation of the achieved throughput in the steady-state with the help of the depicted slope triangles. The throughput $D_{edge,reuse3}$ of the uncoordinated reuse 3 system and $D_{edge,IFCO}$ of the coordinated network with $D_{S,o} = 10$ dB then yield to:

$$D_{edge,reuse3} = \frac{\Delta n \cdot 1500 \text{ Bytes}}{\Delta T} = \frac{10 \cdot 1500 \text{ Bytes}}{1632 \text{ ms}} = 73.5 \text{ kBit/s} \tag{7.1}$$

and

$$D_{edge,IFCO} = \frac{\Delta n \cdot 1500 \text{ Bytes}}{\Delta T} = \frac{10 \cdot 1500 \text{ Bytes}}{909 \text{ ms}} = 132 \text{ kBit/s} \ . \tag{7.2}$$

The achieved steady-state performance gain with respect to the cell edge performance is thus about 80% and therefore more than half the gain in the full-buffer scenario. If transient effects are to be considered in addition, the throughput $\bar{D}$ can be calculated by simply counting the

number of packets since the beginning of the on-period up to a certain time. For example, until $\Delta T = 4$ s:

$$\bar{D}_{edge,reuse3} = \frac{\Delta n \cdot 1500 \text{ Bytes}}{\Delta T} = \frac{22.14 \cdot 1500 \text{ Bytes}}{4000 \text{ ms}} = 66.42 \text{ kBit/s} \tag{7.3}$$

and

$$\bar{D}_{edge,IFCO} = \frac{\Delta n \cdot 1500 \text{ Bytes}}{\Delta T} = \frac{37 \cdot 1500 \text{ Bytes}}{4000 \text{ ms}} = 111 \text{ kBit/s} \ . \tag{7.4}$$

The cell edge performance gain including transient effects within the first 4 seconds is thus still significant and about 67%. In the first 5 seconds, it is almost 71%.

As a conclusion, it can be noted that the very simple approach to deal with empty buffers that was used here already achieves acceptable results. It is therefore a proof-of-concept to show that the proposed distributed IFCO scheme cannot only be used in a full-buffer scenario but also in a scenario with random on-off traffic sources. Further enhancements of the algorithm are possible to improve the performance in such situations, which were partially discussed in Section 5.4.3. These performance improvements remain for further study, where the above results give a very promising perspective on the potential performance gain.

# 8 Conclusion and Outlook

The increasing demand for bandwidth along with the scarcity of the available frequency spectrum requires advanced algorithms to increase the spectral efficiency in emerging cellular networks. A promising approach is the cooperation of network nodes in order to minimize interference in the network. In this monograph, a distributed algorithm for interference coordination (IFCO) was proposed. It aims at increasing the spectral efficiency of the cellular network while at the same time providing a high degree of fairness towards the mobile terminals in unfavorable reception conditions. The algorithm is based on a central coordinator, which periodically collects system state information from the base stations and communicates control information to them. The communication interval can be in the order of seconds, making the distributed architecture well implementable.

Orthogonal Frequency Division Multiple Access (OFDMA) was chosen as the technological basis for the proposed interference coordination scheme. OFDMA is used within several emerging cellular standards, such as IEEE 802.16e and 3GPP Long Term Evolution (LTE). The underlying OFDM transmission system makes it robust in multipath fading environments and allows for a high spectral efficiency. Furthermore, beam steering antennas were used, and the proposed interference coordination algorithm was designed to exploit their capabilities. An introduction to the technological basics, such as fading channels, antenna systems, and OFDMA networks, was given in Chapter 2.

After a thorough investigation of the state-of-the-art in interference coordination in the first half of Chapter 4, the notion of an interference graph was introduced in Section 4.6. An interference graph contains mobile terminals as vertices and critical interference relations among them as edges. These interference relations can be only among cell sectors of one base station (local graph), or they can span multiple base station sites (global graph). The interference graph allows a formal application of graph theoretical methods to the problem of interference coordination. In particular, the problem can be traced back to the well-known vertex coloring problem. This requires the sub-division of the OFDMA MAC frame into resource partitions, where each resource partition is associated with one color. The resource assignment to the transmissions for the individual mobile terminals can then be done by using standard graph-coloring algorithms. This procedure obeys the constraints of the interference graph and thus effectively solves the interference coordination problem. Besides these standard algorithms, a heuristic was presented in Section 4.6.5 that includes external constraints such as QoS requirements into the resource assignment process. Based on these resource assignment strategies, the tradeoff between the overall resource utilization in the network and the Signal-to-Interference-Ratio (SIR) conditions was discussed.

The proposed distributed and dynamic IFCO scheme was described in Chapter 5. The architecture foresees a central coordinator connected to all base stations in a certain coordination area. The central coordinator is responsible for the coordination of interference among base stations. A global outer optimization based on a global interference graph is performed in the central coordinator. The global interference graph is created based on state information collected from the base stations. A variation of the standard vertex coloring problem is solved in the central coordinator, and the results are communicated to the base stations. Every base station solves a local inner optimization problem which takes into account the coloring information provided by the central coordinator. It additionally takes into account local state information and further constraints, such as QoS information. The inner optimization problem was formulated as a Binary Integer Linear Program (BILP). Such problems are generally difficult to solve, and a genetic algorithm was adapted to solve the problem. First, this required the development of a suitable genetic representation. Three variations were eventually considered for further evaluation, and suitable mutation and crossover operators were tailored to these variations.

The proposed algorithm and the reference systems were evaluated with an event-driven simulation system based on the IKR SimLib [SimLib]. Full terminal mobility with a velocity of 30 km/h and both uniform and non-uniform distributions of mobile terminals to the cell sectors were assumed. Additionally, Monte-Carlo simulations were performed in order to explore the large parameter space of the genetic algorithm. The simulation system was based on a detailed model of a cellular IEEE 802.16e network with a comprehensive representation of the MAC layer mechanisms. The physical layer abstraction was based on an average value interface. A detailed description of this model was given in Chapter 6. Furthermore, a fundamental treatment of the underlying modeling concepts in wireless networks was given in Chapter 3.

The metrics used during the performance evaluation were discussed in Section 3.5. They were adopted directly from the evaluation methodology of the 3GPP and the Next Generation Mobile Networks (NGMN) Alliance. In particular, the overall spectral efficiency was considered by means of the aggregate sector throughput. Besides, in order to study fairness, the 5% throughput quantile of the mobile terminal's throughput was taken as an indicator for the cell edge performance. Both throughput metrics were measured at the IP level, thus capturing all effects of retransmissions and padding overhead. In addition to the throughput metrics, a number of indirect metrics was used, such as the Signal-to-Interference-and-Noise-Ratio (SINR) or the resource utilization.

The performance evaluation in Chapter 7 first dealt with the performance of the uncoordinated reference systems with frequency reuse 1 and 3 both with sector and beamforming antennas in a full-buffer scenario. In general, the reuse 1 systems achieve a higher aggregate sector throughput than their corresponding reuse 3 system while at the same time showing a lower cell edge performance. Further reference systems with beamforming antennas were based on a global graph-based coordination in a frequency reuse 1 scenario with the afore mentioned resource assignment heuristic. Such a system showed a significant gain of more than 70% in the aggregate throughput compared to the uncoordinated reuse 3 system. At the same time, the cell edge throughput could be increased by more than 30%. An optimized generation of the interference graph achieved an additional increase in the cell edge throughput of 20%. Even with a local coordination, the performance gain in aggregate throughput was about 40%, while at the same time the cell edge throughput decreased by about 25%. Further performance improvements

could be achieved by combining Fractional Frequency Reuse (FFR) with a local coordination, increasing the aggregate sector throughput by about 80% compared to the uncoordinated reuse 3 system while sacrificing about 65% of the cell edge throughput. However, the application of FFR allows to trade off both performance metrics against each other.

In order to parametrize the genetic algorithm for solving the inner optimization problem, the resource utilization and the number of unserved mobile terminals after an optimization period were evaluated by means of Monte-Carlo simulations. It turned out that the simplest and most efficient genetic representation *Random Genomes* delivered the best results. Moreover, it was shown that the crossover operator does not help to improve the convergence behavior of the genetic algorithm, and hence, the crossover rate can be set to zero thus effectively deactivating the crossover operator. A low mutation rate of around 0.1 was found to be sufficient. Finally, the parameter studies showed that a small population size $|\boldsymbol{P}| = 40$ combined with a small number of generations of only $N_{gen} = 20$ already delivered sufficiently good results. This opens the way to a real-time implementation of the genetic algorithm. In particular, it was concluded that a real-time hardware implementation with a total execution time of only a few ten $\mu$-seconds is possible.

With respect to the throughput, the parametrization of the graph-based distributed algorithm allows to trade off the aggregate throughput with the cell edge throughput. Depending on the parameter choice, the aggregate throughput can be increased by about 30% while increasing the cell edge throughput by more than 130% compared to the uncoordinated frequency reuse 3 system. Alternatively, for a different parameter choice, the aggregate throughput was shown to increase by about 50% with a cell edge throughput gain of about 100%.

Subsequently, the signaling delay to the central coordinator and the update period of the global coordination information was evaluated. This is essential in a distributed system like the one considered here. It was shown that the signaling delay has a higher impact on the system performance than the update period. Nevertheless, both the signaling delay and the update period may be chosen in the order of half a second with the algorithm still delivering almost the same high performance. Even for delays and update periods in the order of seconds, the interference coordinated system still shows significant performance gains over the uncoordinated reference systems. It needs to be noted that all results were obtained with a terminal velocity of 30 km/h. It is obvious that in a scenario with slowly moving or stationary terminals the signaling delay and update period has a far smaller impact on the system performance.

The performance of the distributed interference coordination algorithm was also investigated in combination with a non-uniform distribution of mobile terminals to the cell sectors. This revealed an important property of the distributed interference coordination algorithm, namely the ability to perform load balancing between the differently loaded cells. Altogether, the performance gain observed with a uniform terminal distribution was retained in the non-uniform case.

Finally, an initial performance evaluation was conducted with on-off traffic sources. Minor modifications to the proposed interference coordination algorithm were necessary in order to facilitate the operation in a scenario with empty buffers. A performance gain of about 30% with respect to the aggregate sector throughput compared to the uncoordinated reuse 3 system could be maintained at an activity factor of $A_F = 0.5$ and a mean on-off-period duration of

$E[T_P] = 20$ s. Moreover, a significant cell edge performance gain of about 80% is achievable. Besides these steady state results, an analysis of the transient behavior revealed a slower startup time of the interference coordinated system only within the first few transmitted packets. It needs to be noted that these results were obtained without any sophisticated modifications to the algorithm that could enhance its behavior in a scenario with heterogeneous traffic requirements. The results can therefore be regarded as a proof-of-concept and provide a solid basis for the further development of the algorithm.

As a conclusion, interference coordination proved to be a powerful tool to increase the spectral efficiency of cellular networks. Moreover, a significant increase of the cell edge performance could be achieved. The computational complexity of the proposed algorithm is well manageable. Even the genetic algorithm involved in the inner optimization problem could efficiently be implemented in hardware with an overall execution time in the order of tens of $\mu$-seconds. The tolerance of the algorithm with respect to signaling delay and update period make it well suitable for an implementation in a real network.

Future work could be directed to the inclusion of QoS constraints in the distributed dynamic interference coordination algorithm beyond the already considered utility values. Initial ideas were already given in Section 5.4. Closely related to this issue, the idea of a combined uncoordinated and coordinated operation as discussed in Section 5.4.3 could be further elaborated on. Furthermore, signaling and measurement issues could be treated in more detail. In particular, suitable measurement procedures for the input parameters to the interference coordination algorithms need to be developed. Finally, the extension of the algorithm from a frequency-diverse to a frequency-selective operation would be of great interest, which would also allow investigating the inherent tradeoff between frequency-selective scheduling and interference coordination.

# A 3GPP LTE Network Architecture

In this Chapter, a brief overview of the 3GPP Long Term Evolution (LTE) network architecture is given. Figure A.1 shows a simplified view of the 3GPP LTE network architecture according to [TS 36.300], focusing on the radio access network portion. Compared to the conventional 3GPP UMTS network architecture, the hierarchy level of the Radio Network Controller (RNC) is removed. Instead, the Node B is enriched with functionality and is referred to as eNode B (eNB). Among others, it is responsible for Radio Resource Management (RRM), scheduling, mobility control, and routing of user data towards the Serving Gateway (S-GW).

An IP-based network connects the eNode Bs via the X2 interface. The eNode Bs are connected to the Mobility Management Entity (MME) via the S1-MME interface and to the S-GW via the S1u interface, summarized as S1 interface in Fig. A.1. Among others, the MME is responsible for selection of a new MME in case of handovers with change of MME, mobility between 3GPP access networks, roaming, and authentication. The S-GW is the local mobility anchor for inter-eNode B handover and inter-3GPP mobility, and it is, among others, responsible for packet routing and forwarding, and charging.

Further detail can be found in the respective standard documents [TS 36.300] and [TS 23.401].

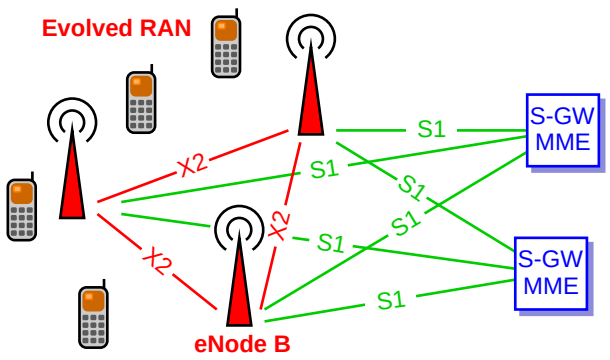



**Figure A.1:** Simplified view of the 3GPP LTE network architecture.

# B Genetic Algorithm Glossary

The following collection briefly explains the most important terms in the area of genetic algorithms.

- **Allele**

  The alleles are the possible values of a genome. For example, for a binary gene, the alleles are 0 and 1.

- **Chromosome**

  See *genome*.

- **Demes**

  Demes are sub-populations that are evolved separately from each other. Exchange of genetic information may occur by *migration*. Natural populations may be separated into demes by rivers or mountains.

- **Fitness**

  Every *individual* or *solution* is associated with a real-valued fitness value which corresponds to how good the solution is. Larger fitness values usually correspond to better solutions.

- **Gene**

  *Genes* are the individual elements a genome consists of. A gene may take on different values. The possible values of a gene are called *alleles*, which may differ for the genes within a genome.

- **Generation**

  *Populations* are evolved over time, which takes place in discrete steps where each step corresponds to one *generation*. One generation represents a stable set of *solutions*, which is then evolved into the next generation by the *genetic operators*.

- **Genetic operators**

  The genetic operators come into action when evolving one generation to the next. They include the selection operator, the mutation operator, and the crossover (or recombination) operator.

- **Genome, Chromosome**

  Genomes or chromosomes are possible solutions to the considered problem. Often, they are coded representations of problem solutions, which first need to be decoded in order to obtain the solution. A genome or chromosome consists of a number of genes.

- **Genotype**

  The genotype is the encoded solution. For example, the genotype of an animal is the actual genome, i.e., the actual genetic information. In contrast, the *phenotype* is the decoded solution.

- **Individual**

  The population in a genetic algorithm consists of several individuals, each of which is represented by a *genome*. The individuals therefore constitute all currently active solutions, and individuals are therefore also referred to as *solutions*.

- **Migration**

  Migration describes the process of *individuals* moving from one *deme* to another, thus exchanging genetic information between them.

- **Phenotype**

  The phenotype is the decoded solution. In contrast to the genotype, the phenotype is the actual appearance of the solution. The phenotype of an animal is its actual physical appearance.

- **Population**

  The collection of *individuals* that is evolved from one generation to the next is referred to as population.

- **Solution**

  See *individual*.

# C Flow Charts of Genetic Operators

This Chapter gives the flow charts of selected genetic operators that were introduced in Chapter 5.3.

Figure C.1 depicts the flow chart of the Standard PMX operator that was presented in Section 5.3.5.1. Note that the operation in Fig. C.1 has to be performed twice, once for each child, thereby swapping *mom* and *dad* in the flow chart.

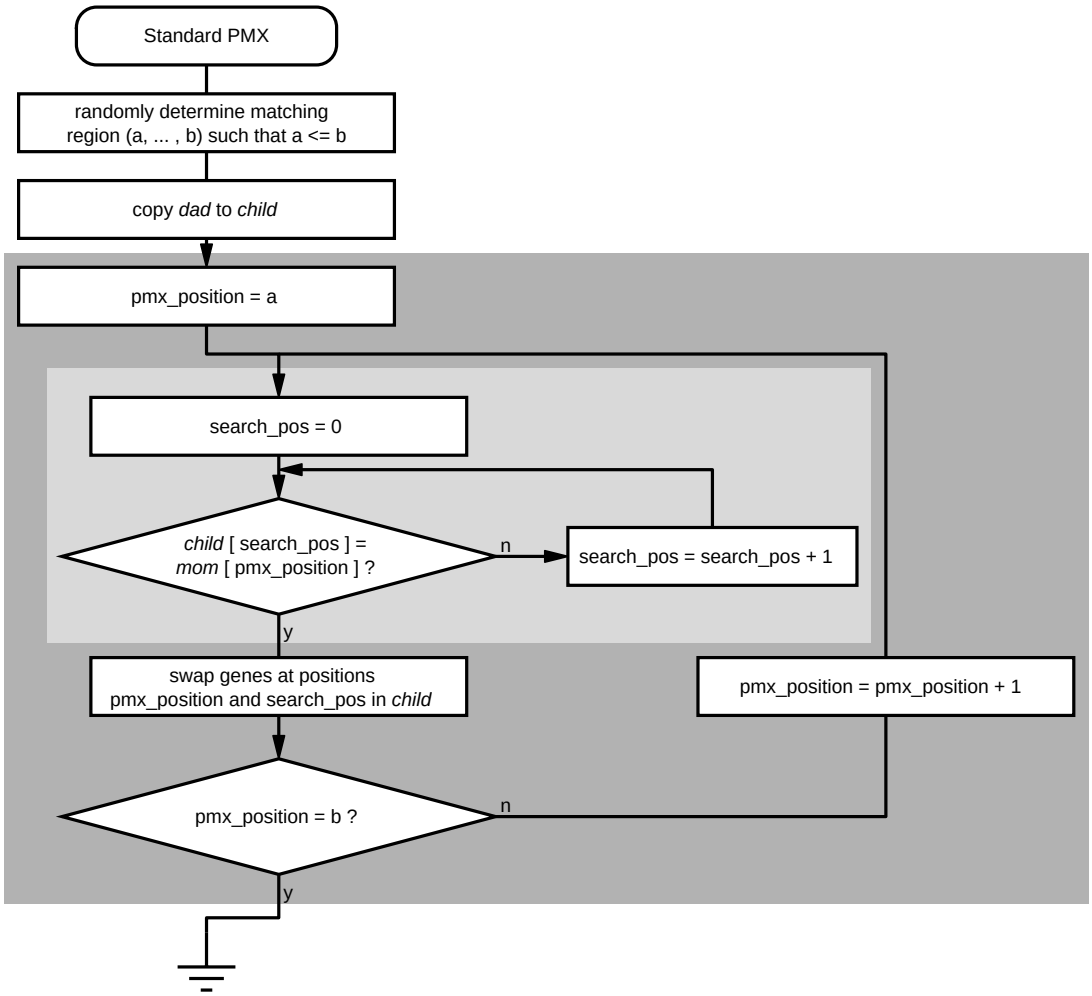



**Figure C.1:** Flow chart for Standard PMX operator.

The flow chart of the more advanced Random PMX from Section 5.3.5.2 is given in Fig. C.2. Again, the operation in Fig. C.2 has to be performed twice, once for each child, thereby swapping *mom* and *dad* in the flow chart.

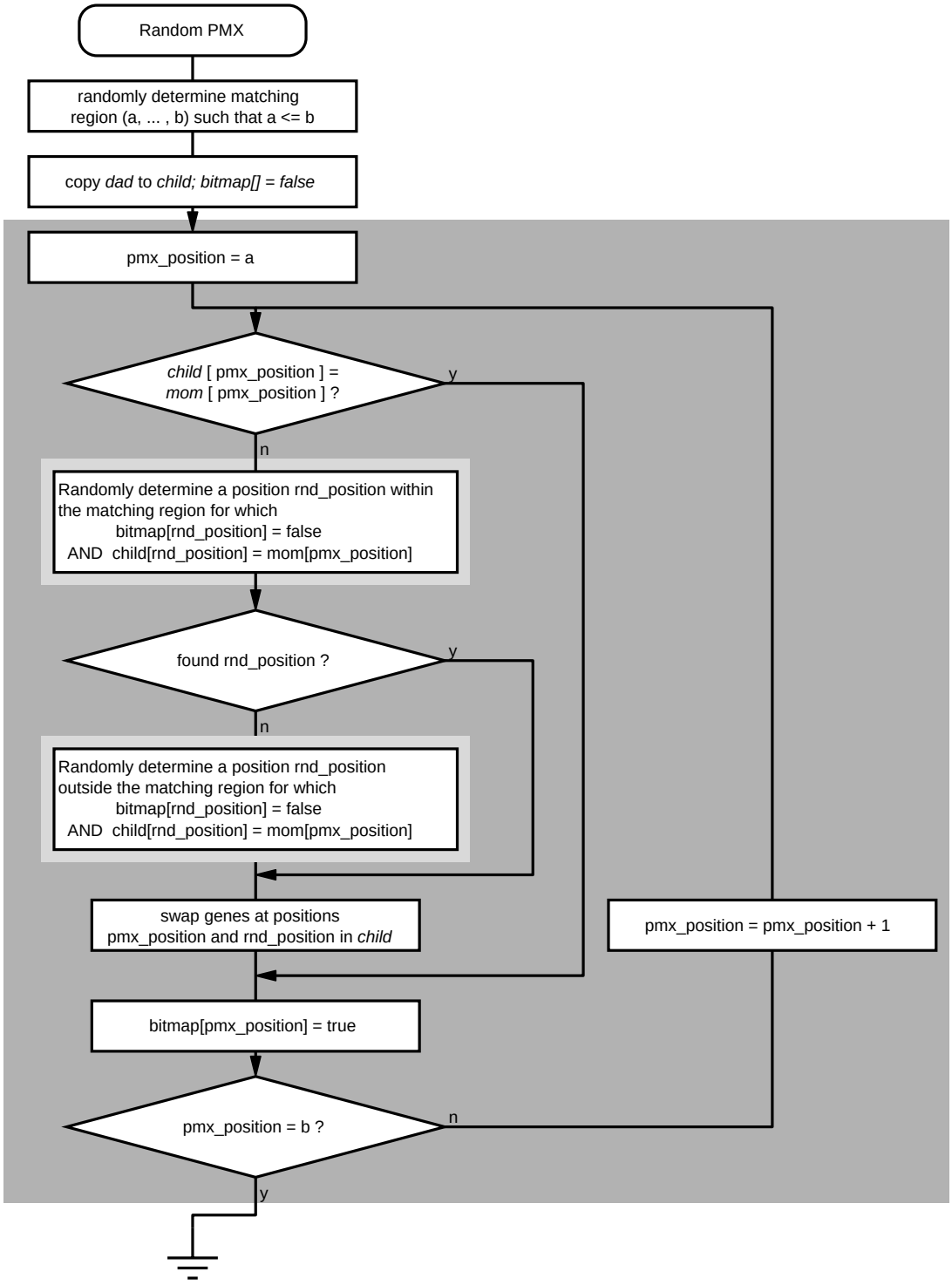



**Figure C.2:** Flow chart for random PMX operator.

# D Impact of Mobility Model

The underlying cellular model of the simulations in this monograph restricts all mobile terminals to their respective cell sector by means of geographic cell sector boundaries. Sector borders are drawn irrespective of the actual shadowing conditions, which would otherwise lead to a mobile terminal being assigned to a neighboring base station. This implies that a mobile terminal may be served by a transceiver which is not the optimal serving transceiver. As a consequence, throughput results are worse in the cell edge areas than they usually would be if the mobile terminal were allowed to choose its serving base station based solely on path loss measurements. In order to ensure that all relative performance results obtained in this monograph are valid even in such a more realistic setup, simulations have been carried out with deactivated shadowing. In such a setup, the just described effect does not occur, which allows to conclude on the impact of the terminal movement restriction on relative performance comparisons.

Fig. D.1 plots the 5% throughput quantile over the aggregate sector throughput for selected configurations of the graph-based globally coordinated system and the uncoordinated reuse 1 and reuse 3 reference systems if shadowing is deactivated. The chart reveals an almost as high performance gain for the graph-based globally interference coordinated systems as in the scenario with shadowing. As a conclusion, the applied mobility model does not result in an unfair advantage of the coordinated systems.

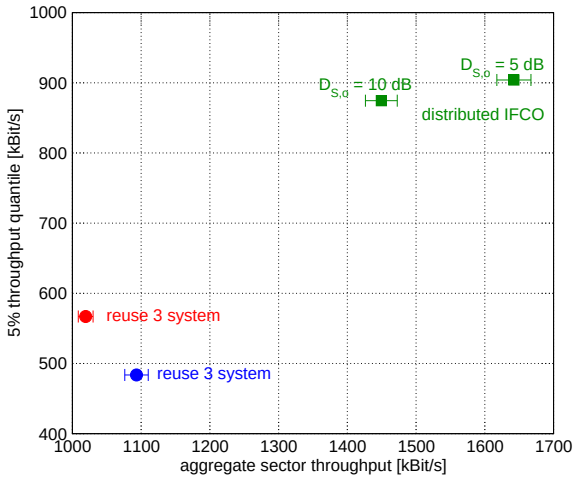



**Figure D.1:** Throughput quantile vs. aggregate sector throughput for scenario with deactivated shadowing. $t_{C,delay} = 0.5$ s, $t_{C,up} = 0.5$ s.

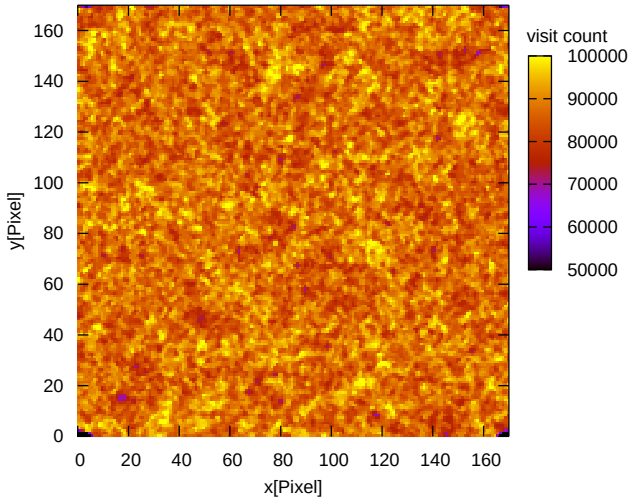



**Figure D.2:** Terminal visit count per pixel, $v = 30$ km/h, $T_R = 40000$ seconds.

The distribution of mobile terminals within the area is illustrated in Fig. D.2. Plotted is the *visit count* per observation pixel, i.e., the number of times a mobile terminal visits a pixel. Thereby, in each MAC frame, the locations of the mobile terminals are captured, and the visit count is increased for the pixel a mobile terminal falls into. The result was obtained for a very long simulated real time of $T_R = 40000$ seconds at a terminal velocity of $v = 30$ km/h. The graph does not reveal any areas with an increased terminal density that correspond to any features of the physical network, such as cell boundaries or base station sites. This allows the conclusion of a uniform terminal distribution throughout the area with the used mobility model.

# Bibliography


[Ahm06] M. H. Ahmed, H. Yanikomeroglu, and S. Mahmoud. Interference management using basestation coordination in broadband wireless access networks. *Wireless Communications & Mobile Computing*, 6(1):95–103, February 2006.

[And93] T. E. Anderson, S. S. Owicki, J. B. Saxe, and C. P. Thacker. High-speed switch scheduling for local-area networks. *ACM Transactions on Computer Systems (TOCS)*, 11(4):319–352, November 1993.

[And07] J. G. Andrews, A. Ghosh, and R. Muhamed. *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall PTR, 2007.

[Bar07] S. J. Barrett. High speed downlink packet access communication in a cellular communication system. *Wipo Patent WO/2007/127543*, November 2007.

[Bau05] D. Baum, J. Hansen, and J. Salo. An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM). In *Proc. 61st IEEE Vehicular Technology Conference (VTC 2005-Spring)*, volume 5, pages 3132–3136, Stockholm, Sweden, May 2005.

[Bea97] B. Beauquier, J.-C. Bermond, L. Gargano, P. Hell, S. Perennes, and U. Vaccaro. Graph problems arising from wavelength-routing in all-optical networks. In *Proc. 2nd Workshop on Optics and Computer Science*, Geneva, Switzerland, April 1997.

[Ber00] H. L. Bertoni. *Radio Propagation for Modern Wireless Systems*. Prentice Hall PTR, 2000.

[Ber02] P. Bergamo, D. Maniezzo, A. Giovanardi, G. Mazzini, and M. Zorzi. An improved markov chain description for fading processes. In *Proc. IEEE International Conference on Communications (ICC 2002)*, volume 3, pages 1347–1351, New York City, NY, USA, April 2002.

[Ber04a] L. T. Berger, T. E. Kolding, P. E. Mogensen, F. Frederiksen, and K. I. Pedersen. Effects of other-sector interference variation on detection, link adaptation, and scheduling in HSDPA. In *Proc. The Nordic Radio Symposium*, Oulu, Finland, August 2004.

[Ber04b] L. T. Berger, T. E. Kolding, P. E. Mogensen, and L. Schumacher. Geometry based other-sector interference modelling for downlink system simulations. In *Proc. 7th International Symposium on Wireless Personal Multimedia Communications (WPMC 2004)*, Abano Terme (Padova), Italy, September 2004.

[Bet01] C. Bettstetter. Mobility modeling in wireless networks: categorization, smooth movement, and border effects. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(3):55–66, July 2001.

[Bet02] C. Bettstetter and C. Wagner. The spatial node distribution of the random waypoint mobility model. In *Proc. German Workshop on Mobile Ad-Hoc Networks (WMAN)*, pages 41–58, Ulm, Germany, March 2002.

[Big74] N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, 1974.

[Bon05] T. Bonald, S. Borst, and A. Proutière. Inter-cell scheduling in wireless data networks. In *Proc. European Wireless (EW 2005)*, pages 566–572, Nicosia, Cyprus, April 2005.

[Bon06] T. Bonald, S. Borst, and A. Proutière. Inter-cell coordination in wireless data networks. *European Transactions on Telecommunications*, 17(3):303–312, May 2006.

[Boo07] C. Boonthum, I. Levinstein, S. Olariu, E. Pigli, E. Shurkova, and A. Zomaya. Mobile computing: Opportunities for optimization research. *Computer Communications*, 30(4):670–684, February 2007.

[Bra07] S. Braun. Optimierung des Uplink-Zugriffes in drahtlosen IEEE 802.16e (WiMAX) Netzen. Diplomarbeit, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, January 2007.

[Bré79] D. Brélaz. New methods to color the vertices of a graph. *Communications of the ACM*, 22(4):251–256, April 1979.

[Bäc00a] T. Bäck, D. B. Fogel, and T. Michalewicz, editors. *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing, January 2000.

[Bäc00b] T. Bäck, D. B. Fogel, and T. Michalewicz, editors. *Evolutionary Computation 2: Advanced Algorithms and Operators*. Institute of Physics Publishing, November 2000.

[Caf98] J. Caffery and G. Stüber. Overview of radiolocation in CDMA cellular systems. *IEEE Communications Magazine*, 36(4):38–45, April 1998.

[Cai03] X. Cai and G. B. Giannakis. A two dimensional channel simulation model for shadow fading processes. *IEEE Transactions on Vehicular Technology*, 52(6):1558–1567, November 2003.

[Cha82] G. J. Chaitin. Register allocation & spilling via graph coloring. In *Proc. 1982 SIGPLAN Symposium on Compiler construction*, pages 98–105, Boston, MA, USA, June 1982.

[Cha92] G. Chan. Effects of sectorization on the spectrum efficiency of cellular radio systems. *IEEE Transactions on Vehicular Technology*, 41(3):217–225, August 1992.

[Cha06] P. Chan, E. Lo, R. Wang, E. Au, V. Lau, R. Cheng, W. H. Mow, R. Murch, and K. Letaief. The evolution path of 4G networks: FDD or TDD? *IEEE Communications Magazine*, 44(12):42–50, December 2006.

[Com84] R. A. Comroe and D. J. Costello, Jr. ARQ schemes for data transmission in mobile radio systems. *IEEE Journal on Selected Areas in Communications*, 2(4):472–481, July 1984.

[COST231] *COST 231 Final Report: Digital Mobile Radio Towards Future Generation Systems*. Office for Official Publications of the European Communities, Luxembourg, 1999.

[Cox71] D. C. Cox and D. O. Reudink. Dynamic channel assignment in high-capacity mobile communications systems. *Bell System Technical Journal*, 50(6):1833–1857, July-August 1971.

[Cox72a] D. C. Cox and D. O. Reudink. A comparison of some channel assignment strategies in large-scale mobile communications systems. *IEEE Transactions on Communications*, 20(2):190–195, April 1972.

[Cox72b] D. C. Cox and D. O. Reudink. Dynamic channel assignment in two-dimensional large-scale mobile radio systems. *Bell System Technical Journal*, 51(7):1611–1629, September 1972.

[Dah07] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming. *3G Evolution: HSPA and LTE for Mobile Broadband*. Academic Press, 2007.

[Dak65] R. J. Dakin. A tree-search algorithm for mixed integer programming problems. *The Computer Journal*, 8(3):250–255, 1965.

[Dar59] C. Darwin. *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. November 1859.

[Das03] S. Das, H. Viswanathan, and G. Rittenhouse. Dynamic load balancing through coordinated scheduling in packet data systems. In *Proc. IEEE Conference on Computer Communications (INFOCOM 2003)*, volume 1, pages 786–796, San Francisco, CA, USA, March 2003.

[Dav91] L. Davis. *Handbook of genetic algorithms*. Thomson Publishing Group, 1991.

[DH95] A. Duel-Hallen, J. Holtzman, and Z. Zvonar. Multiuser detection for CDMA systems. *IEEE Personal Communications*, 2(2):46–58, April 1995.

[Dom05] W. Domschke and A. Drexl. *Einführung in Operations Research*. Springer-Verlag, 6th edition, 2005.

[Ell63] E. O. Elliott. Estimates of error rates for codes on burst-noise channels. *The Bell System Technical Journal*, 42(5):1977–1997, September 1963.

[Eln82] S. Elnoubi, R. Singh, and S. Gupta. A new frequency channel assignment algorithm in high capacity mobile communication systems. *IEEE Transactions on Vehicular Technology*, 31(3):125–131, August 1982.

[Eng73a] J. Engel and M. Peritsky. Statistically-optimum dynamic server assignment in systems with interfering servers. *IEEE Transactions on Communications*, 21(11):1287–1293, November 1973.

[Eng73b] J. Engel and M. Peritsky. Statistically-optimum dynamic server assignment in systems with interfering servers. *IEEE Transactions on Vehicular Technology*, 22(4):203–209, November 1973.

[Erc99] V. Erceg, L. Greenstein, S. Tjandra, S. Parkoff, A. Gupta, B. Kulic, A. Julius, and R. Bianchi. An empirically based path loss model for wireless channels in suburban environments. *IEEE Journal on Selected Areas in Communications*, 17(7):1205–1211, July 1999.

[Erc00] V. Erceg. IEEE 802.16 broadband wireless access working group: Channel models for broadband fixed wireless systems. Technical Report IEEE 802.16.3c-00/47, October 2000.

[ES06] H. El-Sallabi, D. Baum, P. Zetterberg, P. Kyosti, T. Rautiainen, and C. Schneider. Wideband spatial channel model for MIMO systems at 5 GHz in indoor and outdoor environments. In *Proc. 63rd IEEE Vehicular Technology Conference (VTC 2006-Spring)*, volume 6, pages 2916–2921, Melbourne, Australia, May 2006.

[For05] I. Forkel, H. Klenner, and A. Kemper. High speed downlink packet access (HSDPA) – enhanced data rates for UMTS evolution. *Computer Networks*, 49(3):325–340, October 2005.

[Fre02] F. Frederiksen and T. E. Kolding. Performance and modeling of WCDMA/HSDPA transmission/H-ARQ schemes. In *Proc. 56th IEEE Vehicular Technology Conference (VTC 2002-Fall)*, volume 1, pages 472–476, Vancouver, BC, Canada, September 2002.

[GAlib] M. Wall. *GAlib, Version 2.4.7*. Massachusetts Institute of Technology (MIT). Available from: `http://lancet.mit.edu/ga/`.

[Gam91] A. Gamst. Application of graph theoretical methods to GSM radio network planning. In *Proc. IEEE International Symposium on Circuits and Systems (ISCAS 1991)*, volume 2, pages 942–945, Singapore City, Singapore, June 1991.

[Ger96] I. Gerszberg and S. Ramamurthy. Autonomous synchronization of base stations in a digital wireless radiotelephone network. *United States Patent 5,528,597*, June 1996.

[Gil60] E. N. Gilbert. Capacity of a burst-noise channel. *The Bell System Technical Journal*, 39(5):1253–1265, September 1960.

[Giu02] R. Giuliano, F. Mazzenga, and F. Vatalaro. Smart cell sectorization for third generation CDMA systems. *Wireless Communications and Mobile Computing*, 2(3):253–267, May 2002.

[God96] L. Godara. Limitations and capabilities of directions-of-arrival estimation techniques using an array of antennas. In *Proc. IEEE International Symposium on Phased Arrray Systems and Technology*, pages 327–333, Boston, MA, USA, October 1996.

[God97] L. Godara. Application of antenna arrays to mobile communications, part II: Beam-forming and direction-of-arrival considerations. *Proceedings of the IEEE*, 85(8):1195–1245, August 1997.

[God04] L. C. Godara. *Smart Antennas*. CRC Press, 2004.

[Gol85] D. E. Goldberg and R. Lingle, Jr. Alleles, loci, and the traveling salesman problem. In *Proc. 1st International Conference on Genetic Algorithms and their Applications*, pages 154–159, Pittsburgh, PA, USA, July 1985.

[Gol89] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, 1989.

[GSMA08] GSMA Association. 2008 corporate brochure, February 2008.

[Gud91] M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronics Letters*, 27(23):2145–2146, November 1991.

[Hal83] S. Halpern. Reuse partitioning in cellular systems. In *Proc. 33rd IEEE Vehicular Technology Conference (VTC 1983)*, volume 33, pages 322–327, Toronto, ON, Canada, May 1983.

[Hat80] M. Hata. Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, 29(3):317–325, August 1980.

[Hay68] J. Hayes. Adaptive feedback communications. *IEEE Transactions on Communications*, 16(1):29–34, February 1968.

[Her87] A. Hertz and D. de Werra. Using tabu search techniques for graph coloring. *Computing*, 39(4):345–351, December 1987.

[Hig00] T. Higuchi and B. Manderick. Hardware realizations of evolutionary algorithms. In *[Bäc00b]*, chapter 26, pages 253–263. 2000.

[Hoc99] D. Hochfelder. Joel engel, electrical engineer, an oral history conducted in 1999. IEEE history center, Rutgers University, New Brunswick, NJ, USA, 1999.

[Hoe92] P. Hoeher. A statistical discrete-time model for the WSSUS multipath channel. *IEEE Transactions on Vehicular Technology*, 41(4):461–468, November 1992.

[Hoe05] P. Hoeher, S. Badri-Hoeher, W. Xu, and C. Krakowski. Single-antenna cochannel interference cancellation for TDMA cellular radio systems. *IEEE Wireless Communications*, 12(2):30–37, April 2005.

[Hol00] H. Holma. Physical layer performance (chapter 11). In *Wideband CDMA systems*. Lecture at Helsinki University of Technology, 2000.

[Hot03] A. Hottinen, O. Tirkkonen, and R. Wichman. *Multi-antenna Transceiver Techniques for 3G and Beyond*. John Wiley & Sons, Ltd, 2003.

[Hul01] A. Hulbert, M. Purat, and S. Oestreich. Network synchronisation for UTRA TDD. In *Proc. IEEE Vehicular Technology Conference (VTC 2001-Spring)*, volume 4, pages 2804–2807, May 2001.

[Häm97] S. Hämäläinen, P. Slanina, M. Hartman, A. Lappeteläinen, H. Holma, and O. Salonaho. A novel interface between link and system level simulations. In *Proc. ACTS Mobile Communication Summit*, pages 599–604, Aalborg, Denmark, October 1997.

[IEEE80216] IEEE 802.16-2004. *IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, October 2004.

[IEEE80216e] IEEE 802.16e. *IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*, February 2006.

[IEEE80216g] IEEE 802.16g. *Amendment to IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems — Management Plane Procedures and Services*, September 2007.

[IEEE80216h] IEEE 802.16h. *IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment for Improved Coexistence Mechanisms for License-Exempt Operation*, May 2006.

[Ike84] F. Ikegami, S. Yoshida, T. Takeuchi, and M. Umehira. Propagation factors controlling mean field strength on urban streets. *IEEE Transactions on Antennas and Propagation*, 32(8):822–829, August 1984.

[ITU-R M.1225] Recommendation ITU-R M.1225. *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*. International Telecommunication Union, February 1997.

[ITU-T I.130] Recommendation ITU-T I.130. *Method for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN*. International Telecommunication Union, November 1988.

[Jai03] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu. Impact of interference on multi-hop wireless network performance. In *Proc. 9th Annual International Conference on Mobile Computing and Networking*, pages 66–80, San Diego, CA, USA, 2003.

[Jan04] M. Janani, A. Hedayat, T. Hunter, and A. Nosratinia. Coded cooperation in wireless communications: space-time transmission and iterative decoding. *IEEE Transactions on Signal Processing*, 52(2):362–371, February 2004.

[Jeo08] H. Jeon, M. Riegel, and S. Jeong. Transmission of IP over Ethernet over IEEE 802.16 networks. draft-ietf-16ng-ip-over-ethernet-over-802.16-05.txt, IETF, February 2008.

[Joh91] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon. Optimization by simulated annealing: an experimental evaluation; part II, graph coloring and number partitioning. *Operations Research*, 39(3):378–406, May–June 1991.

[Jos99] D. E. Joslin and D. P. Clements. "Squeaky wheel" optimization. *Journal of Artificial Intelligence Research*, 10:353–373, January–June 1999.

[Jun07] V. Jungnickel, M. Schellmann, A. Forck, H. Gäbler, S. Wahls, A. Ibing, K. Manolakis, T. Haustein, W. Zirwas, J. Eichinger, E. Schulz, C. Juchems, F. Luhn, and R. Zavrtak. Demonstration of virtual MIMO in the uplink. In *Proc. IET Smart Antennas and Cooperative Communications Seminar*, London, UK, October 2007.

[Kam08] Private communication with Stephen Kaminski (Alcatel-Lucent Bell Labs, Germany), July 2008.

[Kan92] T. Kanai. Autonomous reuse partitioning in cellular systems. In *Proc. 42nd IEEE Vehicular Technology Conference*, volume 2, pages 782–785, 1992.

[Kas08] M. Kaschub. Performance evaluation of interference coordination in 802.16e systems. Project report, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, June 2008.

[Kau71] A. Kaufmann. *Einführung in die Graphentheorie*. R. Oldenbourg Verlag, 1971.

[Kia06] S. G. Kiani and D. Gesbert. Maximizing the capacity of large wireless networks: Optimal and distributed solutions. In *Proc. IEEE International Symposium on Information Theory (ISIT 2006)*, pages 2501–2505, Seattle, WA, USA, July 2006.

[Kia07] S. G. Kiani, G. E. Oien, and D. Gesbert. Maximizing multicell capacity using distributed power allocation and scheduling. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC 2007)*, pages 1690–1694, Kowloon, China, March 2007.

[Kie07] S. Kiesel and M. Scharf. Modeling and performance evaluation of transport protocols for firewall control. *Computer Networks*, 51(11):3232–3251, August 2007.

[Kim07] K. Kim, G.-M. Yeo, B.-H. Ryu, and K. Chang. Interference analysis and subchannel allocation schemes in tri-sectored OFDMA systems. In *Proc. 66th Vehicular Technology Conference (VTC 2007-Fall)*, pages 1857–1861, Baltimore, MD, USA, October 2007.

[Kli99] T. Klingenbrunn and P. Mogensen. Modelling cross-correlated shadowing in network simulations. In *Proc. 50th IEEE Vehicular Technology Conference (VTC 1999-Fall)*, volume 3, pages 1407–1411, Amsterdam, The Netherlands, September 1999.

[Koc94] H. Kocher and M. Lang. An object-oriented library for simulation of complex hierarchical systems. In *Proc. Object-Oriented Simulation Conference (OOS 1994)*, pages 145–152, Tempe, AZ, USA, January 1994.

[Kou08] C. Koutsimanis and G. Fodor. A dynamic resource allocation scheme for guaranteed bit rate services in OFDMA networks. In *Proc. IEEE International Conference on Communications (ICC 2008)*, pages 2524–2530, Beijing, China, May 2008.

[Koz91] J. R. Koza. Evolving a computer program to generate random numbers using the genetic programming paradigm. In *Proc. 4th International Conference on Genetic Algorithms*, pages 37–44, San Diego, CA, USA, July 1991.

[Kuc99a] A. Kuchar, M. Taferner, M. Tangemann, C. Hoek, W. Rauscher, M. Strasser, G. Pospischil, and E. Bonek. Real-time smart antenna processing for GSM1800 base station. In *Proc. 49th IEEE Vehicular Technology Conference (VTC 1999-Spring)*, volume 1, pages 664–669, Houston, TX, USA, May 1999.

[Kuc99b] A. Kuchar, M. Taferner, M. Tangemann, C. Hoek, W. Rauscher, M. Strasser, G. Pospischil, and E. Bonek. A robust DOA-based smart antenna processor for GSM base stations. In *Proc. IEEE International Conference on Communications (ICC 1999)*, volume 1, pages 11–16, Vancouver, BC, Canada, June 1999.

[Kuh06] M. Kuhn, S. Berger, I. Hammerstrom, and A. Wittneben. Power line enhanced cooperative wireless communications. *IEEE Journal on Selected Areas in Communications*, 24(7):1401–1410, July 2006.

[Kwo99] Y.-K. Kwok and I. Ahmad. Static scheduling algorithms for allocating directed task graphs to multiprocessors. *ACM Computing Surveys (CSUR)*, 31(4):406–471, December 1999.

[Kyö2006] P. Kyösti et al. WINNER II interim channel model. Project Delivery IST-4-027756 WINNER II D1.1.1 V1.1, November 2006.

[Lan60] A. H. Land and A. G. Doig. An automatic method of solving discrete programming problems. *Econometrica*, 28(3):497–520, July 1960.

[Lan99] F. M. Landstorfer. Wave propagation models for the planning of mobile communication networks. In *Proc. 29th European Microwave Week*, pages 320–325, Munich, Germany, October 1999.

[Law91] A. M. Law and W. D. Kelton. *Simulation Modeling & Analysis*. McGraw-Hill, 2nd edition, 1991.

[Li03] G. Li and H. Liu. Downlink dynamic resource allocation for multi-cell OFDMA system. In *Proc. 58th IEEE Vehicular Technology Conference (VTC 2003-Fall)*, volume 3, pages 1698–1702, Orlando, FL, USA, 2003.

[Li06] Y. G. Li and G. L. Stüber, editors. *Orthogonal frequency division multiplexing for wireless communications*. Springer, 2006.

[Lim05] A. Lim, Y. Zhu, Q. Lou, and B. Rodrigues. Heuristic methods for graph coloring problems. In *Proc. ACM Symposium on Applied Computing (SAC 2005)*, pages 933–939, Santa Fe, NM, USA, March 2005.

[Lin84] S. Lin, J. D. J. Costello, and M. J. Miller. Automatic-repeat-request error-control schemes. *IEEE Communications Magazine*, 22(12):5–17, December 1984.

[Lin96] Y. Ling, J. Elling, and T. Toftegaard. Capacity of intelligent underlay and overlay network. In *Proc. IEE Colloquium on Advanced TDMA Techniques and Applications (Digest No: 1996/234)*, pages 5/1–5/9, London, UK, October 1996.

[Lit92] K. Littger. *Optimierung: Eine Einführung in rechnergestützte Methoden und Anwendungen*. Springer-Verlag, 1992.

[Liu06] S. Liu and J. Virtamo. Inter-cell coordination with inhomogeneous traffic distribution. In *Proc. 2nd Conference on Next Generation Internet Design and Engineering (NGI 06)*, València, Spain, April 2006.

[Mad08] S. Madanapalli, S. D. Park, and S. Chakrabarti. Transmission of IPv4 packets over IEEE 802.16's IP convergence sublayer. draft-ietf-16ng-ipv4-over-802-dot-16-ipcs-02.txt, IETF, February 2008.

[Mal94] E. Malkamäki, F. de Ryck, C. Mourot, and A. Urie. A method for combining radio link simulations and system simulations for a slow frequency hopped cellular system. In *Proc. 44th IEEE Vehicular Technology Conference (VTC 1994)*, volume 2, pages 1145–1149, Stockholm, Sweden, June 1994.

[Mar90] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations*. Wiley-Interscience, November 1990.

[Mar92] S. Marić and E. Alonso. Adaptive borrowing of ordered resources for the pan-european mobile communication (GSM) system. In *Proc. IEEE International Conference on Communications (ICC 1992)*, number 3, pages 1693–1697, Chicago, IL, USA, June 1992.

[Mar94] S. Marić, E. Alonso, and G. Metivier. Adaptive borrowing of ordered resources for the pan-european mobile communication (GSM) system. *IEE Proceedings – Communications*, 141(2):93–97, April 1994.

[Mar03] D. Marx. Graph coloring problems and their applications in scheduling. In *Proc. John von Neumann PhD Students Conference*, Budapest, Hungary, October 2003.

[Mas08] M. Masini. Effiziente Implementierung von genetischen Algorithmen auf einem FPGA. Diplomarbeit, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, April 2008.

[Mat96] F. Mattern. Modellbildung und Simulation. In R. Wilhelm, editor, *"Informatik – Grundlagen, Anwendungen, Perspektiven", C.H. Beck*, pages 56–64, 1996.

[Mat07] The MathWorks. *Matlab Optimization Toolbox 3, User's Guide*, September 2007.

[McD00] C. McDiarmid and B. Reed. Channel assignment and weighted coloring. *Networks*, 36(2):114–117, September 2000.

[MD79] V. H. Mac Donald. Advanced mobile phone service: The cellular concept. *The Bell System Technical Journal*, 58:15–41, January 1979.

[Mil05] T. A. Milligan. *Modern Antenna Design*. Wiley-Interscience, 2nd edition, 2005.

[Mog91] P. Mogensen, P. Eggers, C. Jensen, and J. Andersen. Urban area radio propagation measurements at 955 and 1845 MHz for small and micro cells. In *Proc. IEEE Global Telecommunications Conference (GLOBECOM 1991)*, volume 2, pages 1297–1302, Phoenix, AZ, USA, December 1991.

[Muq02] B. Muquet, Z. Wang, G. Giannakis, M. de Courville, and P. Duhamel. Cyclic prefixing or zero padding for wireless multicarrier transmissions? *IEEE Transactions on Communications*, 50(12):2136–2148, December 2002.

[Mut04] A. Mutter, M. C. Necker, and S. Lück. IP-packet service time distributions in UMTS radio access networks. In *Proc. 10th Open European Summer School and IFIP WG 6.3 Workshop (EUNICE 2004)*, pages 71–78, Tampere, Finland, June 2004.

[Nar02] L. Narayanan. Channel assignment and graph multicoloring. *Handbook of wireless networks and mobile computing*, pages 71–94, 2002.

[Nar07] M. Narandzic, C. Schneider, R. Thomä, T. Jämsä, P. Kyösti, and X. Zhao. Comparison of SCM, SCME, and WINNER channel models. In *Proc. 65th IEEE Vehicular Technology Conference (VTC 2007-Spring)*, pages 413–417, Dublin, Ireland, April 2007.

[Nec04a] M. C. Necker and F. Sanzi. Generalized 8-PSK for totally blind channel estimation in OFDM. In *Proc. IEEE Vehicular Technology Conference (VTC 2004-Spring)*, Milano, Italy, May 2004.

[Nec04b] M. C. Necker and G. L. Stüber. Totally blind channel estimation for OFDM on fast varying mobile radio channels. *IEEE Transactions on Wireless Communications*, 3(5):1514–1525, September 2004.

[Nec05a] M. C. Necker. A simple model for the IP packet service time in UMTS networks. In *Proc. International Teletraffic Congress (ITC-19)*, pages 1475–1485, Beijing, China, August 2005.

[Nec05b] M. C. Necker and S. Saur. Statistical properties of fading processes in WCDMA systems. In *Proc. 2nd International Symposium on Wireless Communication Systems (ISWCS 2005)*, Siena, Italy, September 2005.

[Nec06a] M. C. Necker. A comparison of scheduling mechanisms for service class differentiation in HSDPA networks. *International Journal of Electronics and Communications (AEÜ)*, 60(2):136–141, February 2006.

[Nec06b] M. C. Necker. Towards frequency reuse 1 cellular FDM/TDM systems. In *Proc. 9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2006)*, pages 338–346, Torremolinos, Malaga, Spain, October 2006.

[Nec07a] M. C. Necker. Coordinated fractional frequency reuse. In *Proc. 9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2007)*, pages 296–305, Chania, Crete Island, Greece, October 2007.

[Nec07b] M. C. Necker. Integrated scheduling and interference coordination in cellular OFDMA networks. In *Proc. IEEE 4th International Conference on Broadband Communications, Networks, and Systems (Broadnets 2007)*, Raleigh, NC, USA, September 2007.

[Nec07c] M. C. Necker. Local interference coordination in cellular 802.16e networks. In *Proc. 66th Vehicular Technology Conference (VTC 2007-Fall)*, Baltimore, MA, USA, October 2007.

[Nec07d] M. C. Necker. Local interference coordination in cellular OFDMA networks. In *Proc. 12th VDE/ITG Mobilfunktagung*, Osnabrück, Germany, May 2007.

[Nec07e] M. C. Necker. Modelling of multi-cellular 802.16e systems. Project report, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, March 2007.

[Nec07f] M. C. Necker. Performance evaluation of uncoordinated and interference coordinated 802.16e systems. Project report, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, May 2007.

[Nec08a] M. C. Necker. Interference coordination in cellular OFDMA networks. *IEEE Network*, 22(6), November/December 2008.

[Nec08b] M. C. Necker. A graph-based scheme for distributed interference coordination in cellular OFDMA networks. In *Proc. 67th IEEE Vehicular Technology Conference (VTC 2008-Spring)*, pages 713–718, Singapore City, Singapore, May 2008.

[Nec08c] M. C. Necker, M. Köhn, A. Reifert, J. Scharf, and J. Sommer. Optimized frame packing for OFDMA systems. In *Proc. 67th IEEE Vehicular Technology Conference (VTC 2008-Spring)*, pages 1483–1488, Singapore City, Singapore, May 2008.

[Net89] R. Nettleton and G. Schloemer. A high capacity assignment method for cellular mobile telephone systems. In *Proc. 39th IEEE Vehicular Technology Conference (VTC 1989)*, volume 1, pages 359–367, San Francisco, CA, USA, May 1989.

[Net90] R. Nettleton. Traffic statistics in a self-organizing cellular telephone system. In *Proc. 40th IEEE Vehicular Technology Conference (VTC 1990)*, pages 305–310, Orlando, FL, USA, May 1990.

[NGMN2007] NGMN performance evaluation methodology. Version 1.2, Next Generation Mobile Networks Alliance, June 2007.

[Nie97] T. Nielsen, J. Wigard, and P. Mogensen. On the capacity of a GSM frequency hopping network with intelligent underlay-overlay. In *Proc. 47th IEEE Vehicular Technology Conference (VTC 1997)*, volume 3, pages 1867–1871, Phoenix, AZ, USA, May 1997.

[Nis97] V. Nissen. *Einführung in Evolutionäre Algorithmen*. Vieweg Verlag, 1997.

[Nis00] D. Nissani and I. Shperling. Cellular CDMA (IS-95) location, A-FLT (assisted forward link triangulation) proof-of-concept interim results. In *Proc. 21st IEEE Convention of the Electrical and Electronic Engineers in Israel*, pages 179–182, Tel Aviv, Israel, April 2000.

[Nor94] P. Nordin. A compiling genetic programming system that directly manipulates the machine code. In *Advances in genetic programming*, pages 311–331. MIT Press Cambridge, MA, USA, 1994.

[Nos04] A. Nosratinia, T. Hunter, and A. Hedayat. Cooperative communication in wireless networks. *IEEE Communications Magazine*, 42(10):74–80, October 2004.

[Nua07] L. Nuaymi. *WiMAX Technology for Broadband Wireless Access*. John Wiley & Sons, Ltd, 2007.

[Oku68] T. Okumura, E. Ohmori, and K. Fukuda. Field strength and its variability in VHF and UHF land mobile services. *Review of the Electrical Communication Laboratory*, 16(9-10):825–873, September-October 1968.

[Ots96] S. Otsuka. Network synchronization for TDMA cellular communication using signals from mobile stations in neighboring cells. *United States Patent 5,519,710*, May 1996.

[Pap07] A. Papasakellariou. Signaling requirements to support interference coordination in OFDMA based systems. *United States Patent Application 11/627,095*, January 2007.

[Par06] S. Parkvall, E. Englund, M. Lundevall, and J. Torsner. Evolving 3G mobile systems: broadband and broadcast services in WCDMA. *IEEE Communications Magazine*, 44(2):30–36, February 2006.

[Pau04] A. Paulraj, D. Gore, R. Nabar, and H. Bolcskei. An overview of MIMO communications – a key to gigabit wireless. *Proceedings of the IEEE*, 92(2):198–218, February 2004.

[Pei01] J. Peisa and M. Meyer. Analytical model for TCP file transfers over UMTS. In *Proc. International Conference on Third Generation Wireless and Beyond*, San Francisco, CA, USA, June 2001.

[Pin03] L. Ping, L. Liu, and W. Leung. A simple approach to near-optimal multiuser detection: interleave-division multiple-access. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC 2003)*, volume 1, pages 391–396, New Orleans, LA, USA, March 2003.

[Pin07] L. Ping, Q. Guo, and J. Tong. The OFDM-IDMA approach to wireless communication systems. *IEEE Wireless Communications*, 14(3):18–24, June 2007.

[Pla06] S. Plass, X. G. Doukopoulos, and R. Legouable. Investigations on link-level inter-cell interference in OFDMA systems. In *Proc. Symposium on Communications and Vehicular Technology*, pages 49–52, Liege, Belgium, November 2006.

[Poh99] H. Pohlheim. *Evolutionäre Algorithmen*. Springer, Berlin, 1st edition, November 1999.

[Pro07] J. G. Proakis and M. Salehi. *Digital Communications*. McGraw Hill Higher Education, 5th edition, 2007.

[R1-031303] 3GPP TSG RAN WG1#35 R1-031303. System-level evaluation of OFDM — further considerations. Technical document, Ericsson, Lisbon, Portugal, November 2003.

[R1-050407] 3GPP TSG RAN WG1#41 R1-050407. Interference coordination in new OFDM DL air interface. Technical document, Alcatel, Athens, Greece, May 2005.

[R1-050507] 3GPP TSG RAN WG1#42 R1-050507. Soft frequency reuse scheme for UTRAN LTE. Technical document, Huawei, Athens, Greece, May 2005.

[R1-050593] 3GPP TSG RAN WG1 Ad Hoc on LTE R1-050593. Interference coordination for evolved UTRA uplink access. Technical document, Alcatel, Sophia Antipolis, France, June 2005.

[R1-050608] 3GPP TSG RAN WG1 Ad Hoc on LTE R1-050608. Inter-cell interference mitigation based on IDMA. Technical document, RITT, Sophia Antipolis, France, June 2005.

[R1-050738] 3GPP TSG RAN WG1#42 R1-050738. Interference mitigation – considerations and results on frequency reuse. Technical document, Siemens, London, UK, September 2005.

[R1-050763] 3GPP TSG RAN WG1#42 R1-050763. Muting – further discussion and results. Technical document, Ericsson, London, UK, September 2005.

[R1-050764] 3GPP TSG RAN WG1#42 R1-050764. Inter-cell interference handling for E-UTRA. Technical document, Ericsson, London, UK, September 2005.

[R1-050829] 3GPP TSG RAN WG1#42 R1-050829. Coordinated symbol repetition to mitigate downlink inter cell interference. Technical document, Panasonic, London, UK, September 2005.

[R1-050833] 3GPP TSG RAN WG1#42 R1-050833. Interference mitigation in evolved UTRA/UTRAN. Technical document, LG Electronics, London, UK, September 2005.

[R1-050841] 3GPP TSG RAN WG1#42 R1-050841. Further analysis of soft frequency reuse scheme. Technical document, Huawei, London, UK, September 2005.

[R1-050896] 3GPP TSG RAN WG1#42 R1-050896. Description and simulations of interference management technique for OFDMA based E-UTRA downlink evaluation. Technical document, Qualcomm Europe, London, UK, September 2005.

[R1-051059] 3GPP TSG RAN WG1#42bis R1-051059. Inter-cell interference mititgation for EUTRA. Technical document, Texas Instruments, San Diego, CA, USA, October 2005.

[R1-051085] 3GPP TSG RAN WG1#42bis R1-051085. Resource allocation for interference mitigation with symbol repetition in E-UTRA downlink. Technical document, ETRI, San Diego, CA, USA, October 2005.

[R1-060135] 3GPP TSG RAN WG1 Ad Hoc on LTE R1-060135. Interference mititgation by partial frequency reuse. Technical document, Siemens, Helsinki, Finland, January 2006.

[R1-060209] 3GPP TSG RAN WG1 Ad Hoc on LTE R1-060209. System simulation results for downlink interference coordination. Technical document, Alcatel, Helsinki, Finland, January 2006.

[R1-060291] 3GPP TSG RAN WG1#44 R1-060291. OFDMA downlink inter-cell interference mitigation. Technical document, Nokia, Denver, CO, USA, February 2006.

[R1-060368] 3GPP TSG RAN WG1#44 R1-060368. Performance of inter-cell interference mitigation with semi-static frequency planning for EUTRA downlink. Technical document, Texas Instruments, Denver, CO, USA, February 2006.

[R1-060369] 3GPP TSG RAN WG1#44 R1-060369. Signaling requirements to support semi-static frequency planning for inter-cell interference mitigation in EUTRA downlink. Technical document, Texas Instruments, Denver, CO, USA, February 2006.

[R1-060401] 3GPP TSG RAN WG1#44 R1-060401. Interference mitigation via power control and FDM resource allocation and UE alignment for E-UTRA uplink and TP. Technical document, Motorola, Denver, CO, USA, February 2006.

[R1-060586] 3GPP TSG RAN WG1#44 R1-060586. Downlink and uplink inter-cell interference co-ordination/avoidance – impact on the specifications. Technical document, Ericsson, NTT DoCoMo, Denver, CO, USA, February 2006.

[R1-060864] 3GPP TSG RAN WG1#44bis R1-060864. Overview of resource management techniques for interference mitigation in EUTRA. Technical document, Texas Instruments, Athens, Greece, March 2006.

[R1-060905] 3GPP TSG RAN WG1#44bis R1-060905. Adaptive fractional frequency reuse. Technical document, Nortel, Athens, Greece, March 2006.

[R1-061032] 3GPP TSG RAN WG1#44bis R1-061032. Further system-level simulation results for downlink interference coordination. Technical document, Alcatel, Athens, Greece, March 2006.

[R1-061391] 3GPP TSG RAN WG1#45 R1-061391. Downlink interference coordination. Technical document, Siemens, Shanghai, China, May 2006.

[R1-061455] 3GPP TSG RAN WG1#45 R1-061455. System-level simulations for evaluation using downlink interference coordination. Technical document, Alcatel, Shanghai, China, May 2006.

[R1-061716] 3GPP TSG RAN WG1 LTE Ad Hoc R1-061716. Uplink power control adaptation for EUTRA. Technical document, Motorola, Cannes, France, June 2006.

[R1-061852] 3GPP TSG RAN WG1 LTE Ad Hoc R1-061852. Adaptive fractional frequency reuse. Technical document, Nortel, Cannes, France, June 2006.

[R1-062150] 3GPP TSG RAN WG1 Ad Hoc R1-062150. Adaptive fractional frequency reuse. Technical document, Nortel, Tallinn, Estonia, August 2006.

[R1-063379] 3GPP TSG RAN WG1#47 R1-063379. Power control email discussion summary. Technical document, Nokia, Riga, Latvia, November 2006.

[R1-070040] 3GPP TSG RAN WG1#47bis R1-070040. DL power allocation for dynamic interference avoidance in E-UTRA. Technical document, Motorola, Sorrento, Italy, January 2007.

[R1-070043] 3GPP TSG RAN WG1#47bis R1-070043. Performance comparison of uplink power control methods. Technical document, Motorola, Sorrento, Italy, January 2007.

[R1-070099] 3GPP TSG RAN WG1#47bis R1-070099. Frequency domain channel-dependent scheduling considering interference to neighbouring cell for E-UTRA uplink. Technical document, NTT DoCoMo, Sorrento, Italy, January 2007.

[R1-070284] 3GPP TSG RAN WG1#47bis R1-070284. Fractional time re-use interference co-ordination in E-UTRA downlink. Technical document, Texas Instruments, Sorrento, Italy, January 2007.

[R1-070442] 3GPP TSG RAN WG1#47bis R1-070442. Inter-cell interference control. Technical document, Qualcomm Europe, Sorrento, Italy, January 2007.

[R1-070611] 3GPP TSG RAN WG1#47bis R1-070611. Power setting for interference coordination including pilot and control channels. Technical document, Alcatel-Lucent, Sorrento, Italy, January 2007.

[R1-070667] 3GPP TSG RAN WG1#48 R1-070667. Uplink inter-cell interference management – sensitivity to delayed information via backhaul. Technical document, Qualcomm Europe, St. Louis, MO, USA, February 2007.

[R1-072456] 3GPP TSG RAN WG1#49 R1-072456. On inter-cell interference coordination schemes without/with traffic load indication. Technical document, Ericsson, Kobe, Japan, May 2007.

[R1-072840] 3GPP TSG RAN WG1#49bis R1-072840. Fractional time re-use interference co-ordination in E-UTRA DL. Technical document, Texas Instruments, Orlando, FL, USA, June 2007.

[R1-072921] 3GPP TSG RAN WG1#35 R1-072921. Impact of uplink inter-cell interference coordination on uplink power control. Technical document, Alcatel-Lucent, Orlando, FL, USA, June 2007.

[R1-072974] 3GPP TSG RAN WG1#49bis R1-072974. Downlink interference coordination. Technical document, Nokia & Nokia Siemens Networks, Orlando, FL, USA, June 2007.

[R1-073187] 3GPP TSG RAN WG1#49bis R1-073187. Interference coordination framework with results. Technical document, Alcatel-Lucent, Orlando, FL, USA, June 2007.

[R1-073567] 3GPP TSG RAN WG1#50 R1-073567. Improving cell-edge performance by inter-cell interference coordination. Technical document, Samsung, Athens, Greece, August 2007.

[R1-074641] 3GPP TSG RAN WG1#51 R1-074641. Discussion on the DL interference coordination. Technical document, Nortel, Jeju, Korea, November 2007.

[R1-081873] 3GPP TSG RAN WG1#53 R1-081873. Semi-static interference coordination method. Technical document, Alcatel-Lucent, Kansas City, MO, USA, May 2008.

[R1-081877] 3GPP TSG RAN WG1#53 R1-081877. LTE – IMT advanced – candidate technologies. Technical document, Alcatel-Lucent, Kansas City, MO, USA, May 2008.

[R1-081910] 3GPP TSG RAN WG1#53 R1-081910. Use case of OI/HII indicators for uplink ICIC. Technical document, Mitsubishi Electric, Kansas City, MO, USA, May 2008.

[R2-052906] 3GPP TSG RAN WG2&3#49 R2-052906. Proposal for the C-plane architecture. Technical document, Siemens, Seoul, Korea, November 2005.

[R4-060117] 3GPP TSG RAN WG4#38 R4-060117. Analysis for simulation scenario definition to interference mitigation studies. Technical document, Nokia, Denver, CO, USA, February 2006.

[Rah08a] Private communication with author of [Rah08b], May 2008.

[Rah08b] M. Rahman and H. Yanikomeroglu. Interference avoidance through dynamic downlink OFDMA subchannel allocation using intercell coordination. In *Proc. 67th IEEE Vehicular Technology Conference (VTC 2008-Spring)*, pages 1630–1635, Singapore City, Singapore, May 2008.

[Ram89] R. Ramaswami and K. Parhi. Distributed scheduling of broadcasts in a radio network. In *Proc. 8th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 1989)*, volume 2, pages 497–504, Ottawa, ON, Canada, April 1989.

[Rap02] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, 2nd edition, 2002.

[RFC4968] S. Madanapalli (Ed.). Analysis of IPv6 link models for IEEE 802.16 based networks. RFC 4968, IETF, August 2007.

[RFC5121] B. Patil, F. Xia, B. Sarikaya, J. Choi, and S. Madanapalli. Transmission of IPv6 via the IPv6 convergence sublayer over IEEE 802.16 networks. RFC 5121, IETF, February 2008.

[RFS-SEC] RFS Radio Frequency Systems. *Technical Data Sheet: SEC35-120ANVH*.

[Rin47] D. H. Ring. Mobile telephony – wide area coverage – case 20564. Technical memoranda, Bell Telephone Laboratories, December 1947.

[Rud00] G. Rudolph. Evolution strategies. In *[Bäc00a]*, chapter 9, pages 81–88. 2000.

[Sal88] T. Salvalaggio. On the application of reuse partitioning. In *Proc. 38th IEEE Vehicular Technology Conference (VTC 1988)*, pages 182–185, Philadelphia, PA, USA, June 1988.

[Sal95] H. Salgado, M. Sirbu, and J. Peha. Spectrum sharing through dynamic channel assignment for open access to personal communications services. In *Proc. IEEE International Conference on Communications (ICC 1995)*, volume 1, pages 417–422, Seattle, WA, USA, June 1995.

[Sar00] J. Sarma and K. De Jong. Generation gap methods. In *[Bäc00a]*, chapter 28, pages 205–211. 2000.

[Sar01] C. Saraydar and A. Yener. Adaptive cell sectorization for CDMA systems. *IEEE Journal on Selected Areas in Communications*, 19(6):1041–1051, June 2001.

[Sav07] P. Savazzi and L. Favalli. Dynamic cell sectorization using clustering algorithms. In *Proc. 65th IEEE Vehicular Technology Conference (VTC 2007-Spring)*, pages 604–608, Dublin, Ireland, April 2007.

[Sch97] E. R. Scheinerman and D. H. Ullman. *Fractional Graph Theory: A Rational Approach to the Theory of Graphs*. Wiley-Interscience, 1st edition, August 1997.

[Sha02] A. Sharma. WCDMA and CDMA2000 compared. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC 2002)*, volume 1, pages 23–24 vol.1, 2002.

[Sim07] A. Simonsson. Frequency reuse and intercell interference co-ordination in E-UTRA. In *Proc. 65th IEEE Vehicular Technology Conference (VTC 2007-Spring)*, pages 3091–3095, Dublin, Ireland, April 2007.

[SimLib] Institute of Communication Networks and Computer Engineering, Universität Stuttgart. *IKR Simulation Library, Version 2.6*. Available from: http://www.ikr.uni-stuttgart.de/IKRSimLib/.

[Siv89] K. Sivarajan, R. McEliece, and J. Ketchum. Channel assignment in cellular radio. In *Proc. 39th IEEE Vehicular Technology Conference (VTC 1989)*, volume 2, pages 846–850, San Francisco, CA, USA, May 1989.

[Sou73] J. F. Soufflet. The warehouse problem: A review. *Regional and Urban Economics*, 3(2):187–216, May 1973.

[Spe05] J. Speidel. Multiple Input Multiple Output (MIMO) - Drahtlose Nachrichtenübertragung hoher Bitrate und Qualität mit Mehrfachantennen. *Telekommunikation Aktuell*, 59(07-08), July-August 2005.

[Sta91] T. Starkweather, S. McDaniel, K. Mathias, D. Whitley, and C. Whitley. A comparison of genetic sequencing operators. In *Proc. 4th International Conference on Genetic Algorithms*, pages 69–76, San Diego, CA, USA, July 1991.

[Sta02] D. Staehle, K. Leibnitz, and K. Heck. Effects of soft handover on the UMTS downlink performance. In *Proc. 56th IEEE Vehicular Technology Conference (VTC 2002-Fall)*, volume 2, pages 960–964, Vancouver, BC, Canada, September 2002.

[Ste03a] M. Sternad, T. Ottosson, A. Ahlen, and A. Svensson. Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks. In *Proc. 58th IEEE Vehicular Technology Conference (VTC 2003-Fall)*, volume 4, pages 2486–2490, Orlando, FL, USA, October 2003.

[Ste03b] M. Sternad. Reuse partitioning and system capacity in the adaptive OFDM downlink of the wireless IP project target system. Technical report, Uppsala University, July 2003.

[Ste05] A. Stephens. 802.11 "decrypted". *ACM SIGCOMM Computer Communication Review*, 35(2):91–93, April 2005.

[Stu06] P. Stuedi and G. Alonso. Computing throughput capacity for realistic wireless multihop networks. In *Proc. 9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2006)*, pages 191–198, Torremolinos, Malaga, Spain, October 2006.

[Stü00] G. L. Stüber. *Principles of Mobile Communication*. Kluwer Academic Publishers, 2nd edition, 2000.

[Sys89] G. Syswerda. Uniform crossover in genetic algorithms. In *Proc. 3rd Inernational Conference on Genetic Algorithms*, pages 2–9, Arlington, VA, USA, June 1989.

[Tak93] T. Takenaka, T. Nakamura, and Y. Tajima. All-channel concentric allocation in cellular systems. In *IEEE International Conference on Communications (ICC 1993)*, volume 2, pages 920–924, Geneva, May 1993.

[Tho88] R. Thomas, H. Gilbert, and G. Mazziotto. Influence of the moving of the mobile stations on the performance of a radio mobile cellular network. In *Proc. 3rd Nordic Seminar on Digital Land Mobile Radio Communications*, pages 1–9, Copenhagen, Denmark, September 1988.

[Tor06] J. Torres Rodicio. Design and evaluation of routing protocols in IEEE 802.16 (WiMAX) mesh networks. Master thesis, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, December 2006.

[TR 101 112] TR 101 112. *Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS*. European Telecommunications Standards Institute, April 1998.

[TR 3.018] 3GPP TR 3.018. *Evolved UTRA and UTRAN; Radio Access Architecture and Interfaces (Release 7)*. 3rd Generation Partnership Project, September 2006.

[Tra02] V. Tralli and M. Zorzi. Markov models for the physical layer block error process in a WCDMA cellular system. In *Proc. Global Telecommunications Conference (GLOBECOM 2002)*, volume 2, pages 1925–1929, Taipei, Taiwan, November 2002.

[Tra05] V. Tralli and M. Zorzi. Markov models for the physical layer block error process in a WCDMA cellular system. *IEEE Transactions on Vehicular Technology*, 54(6):2102–2113, November 2005.

[TS 23.234] 3GPP TS 23.234. *3GPP system to Wireless Local Area Network (WLAN) interworking; System description*. 3rd Generation Partnership Project, June 2006.

[TS 23.401] 3GPP TS 23.401. *General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access*. 3rd Generation Partnership Project, June 2008.

[TS 25.814] 3GPP TS 25.814. *Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7)*. 3rd Generation Partnership Project, June 2006.

[TS 25.996] 3GPP TS 25.996. *Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 6)*. 3rd Generation Partnership Project, September 2003.

[TS 36.213] 3GPP TS 36.213 V8.2.0. *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 8)*. 3rd Generation Partnership Project, March 2008.

[TS 36.300] 3GPP TS 36.300. *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*. 3rd Generation Partnership Project, March 2008.

[Vil05] R. Vilzmann, C. Bettstetter, and C. Hartmann. On the impact of beamforming on interference in wireless mesh networks. In *Proc. IEEE Workshop on Wireless Mesh Networks*, Santa Clara, CA, USA, September 2005.

[Wal88] J. Walfisch and H. Bertoni. A theoretical model of UHF propagation in urban environments. *IEEE Transactions on Antennas and Propagation*, 36(12):1788–1796, December 1988.

[Wan95] H. S. Wang and N. Moayeri. Finite-state markov channel – a useful model for radio communication channels. *IEEE Transactions on Vehicular Technology*, 44(1):163–171, February 1995.

[Wat98] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.

[Web08] A. Weber. Performance studies on LTE-Advanced in the Easy-C project. In *27th meeting of the VDE/ITG Fachgruppe 5.2.4*, June 2008.

[Wei03] U. Weiss. *Leistungssteigerung zellularer Mobilfunksysteme durch Anwendung von Makrodiversität und dynamischem Ressourcenmanagement*. 83. Bericht über verkehrstheoretische Arbeiten, Institute of Communication Networks and Computer Engineering, Universität Stuttgart, 2003.

[Wei05] H.-Y. Wei, S. Ganguly, R. Izmailov, and Z. Haas. Interference-aware IEEE 802.16 WiMax mesh networks. In *Proc. IEEE Vehicular Technology Conference (VTC 2005-Spring)*, volume 5, pages 3102–3106, Stockholm, Sweden, May 2005.

[Whi85] J. Whitehead. Cellular spectrum efficiency via reuse planning. In *Proc. 35th IEEE Vehicular Technology Conference (VTC 1985)*, volume 35, pages 16–20, Boulder, CO, USA, May 1985.

[Wig97] J. Wigard, T. Nielsen, P. Michaelsen, and P. Mogensen. Improved intelligent underlay-overlay combined with frequency hopping in GSM. In *Proc. 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 1997)*, volume 2, pages 376–380, Helsinki, Finland, September 1997.

[Wil96] V. Wille, S. Irons, and A. King. Capacity increase in cellular radio networks using existing base station sites. In *Proc. IEE Colloquium on Propagation Aspects of Future Mobile Systems (Digest No: 1996/220)*, pages 4/1–4/8, London, UK, October 1996.

[WiM06] Mobile WiMAX – Part I: A technical overview and performance evaluation. Technical report, WiMAX Forum, February 2006.

[WiM07a] WiMAX forum mobile system profile. Release 1.0, WiMAX Forum, May 2007.

[WiM07b] WiMAX system evaluation methodology. Version 1.0, WiMAX Forum, January 2007.

[WiM08a] WiMAX forum network architecture, stage 2: Architecture, tenets, reference model and reference points. Release 1, version 1.2, WiMAX Forum, January 2008.

[WiM08b] WiMAX forum network architecture, stage 2: Architecture, tenets, reference model and reference points, 3GPP–WiMAX interworking. Release 1, version 1.2, WiMAX Forum, January 2008.

[WiM08c] WiMAX forum network architecture, stage 3: Detailed protocols and procedures. Release 1, version 1.2, WiMAX Forum, January 2008.

[Wu05] Y. Wu, Q. Zhang, W. Zhu, and S.-Y. Kung. Bounding the power rate function of wireless ad hoc networks. In *Proc. 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005)*, volume 1, pages 584–595, Miami, FL, USA, March 2005.

[Xia06] W. Xiao, R. Ratasuk, A. Ghosh, R. Love, Y. Sun, and R. Nory. Uplink power control, interference coordination and resource allocation for 3GPP E-UTRA. In *Proc. 64th IEEE Vehicular Technology Conference (VTC 2006-Fall)*, pages 1–5, Montréal, QC, Canada, September 2006.