♦ Enhancement of Throughput and Fairness in 4G Wireless Access Systems by Non-Orthogonal Signaling

Joerg Schaepperle and Andreas Rüegg

Besides having significantly enhanced throughput, fourth generation (4G) cellular wireless access systems are expected to be cost efficient and provide sufficient performance even at the cell edge. This paper proposes a new method, called radial space-division multiple access (RDMA) for improving the system's spectral efficiency significantly and providing fairness among users without increasing the number of antennas and without requiring fast backbone communication and synchronization between base stations. The basic idea is to approach the capacity of the multi-user channel under fairness constraints by exploiting not only the angular dimension of space, as in space-division multiple access (SDMA), but also the radial dimension. Non-orthogonal signaling for simultaneous transmission of multiple signals over the same radio resource and joint detection in the receiver are used. Significant throughput gains are achieved with low complexity methods requiring only long term channel statistics and no instantaneous channel state information at the transmitter. Detailed link and first simple system-level simulation results in the context of the evolution of IEEE 802.16/ Worldwide Interoperability for Microwave Access (WiMAX) systems towards 802.16m show that system throughput can be doubled without requiring more advanced analog hardware or increasing bandwidth or transmit power.

Introduction

Although today the Shannon capacity can be nearly achieved for point-to-point links in wireless systems by practical modulation and coding methods, the aggregate throughput in practical cellular systems is typically far away from the theoretical limits.

From information theory, it is known that orthogonal signaling does not achieve maximum capacity in all cases. Instead, several users must be served at the same time on the same frequency and/or code to approach the performance limits. For maximizing the aggregate throughput (sum rate) in downlink, dirty paper coding is a theoretical concept to achieve capacity. However, sum rate maximizing is neither an appropriate objective in real systems with fairness requirements, nor is the algorithm for dirty paper coding practically realizable with a large number of users.

Especially in the case where additional fairness constraints between users with different channel

Bell Labs Technical Journal 13(4), 59–78 (2009) © 2009 Alcatel-Lucent. Published by Wiley Periodicals, Inc. Published online in Wiley InterScience (www.interscience.wiley.com) • DOI: 10.1002/bltj.20336

Panel 1. Abbreviations, Acronyms, and Terms

3GPP—3rd Generation Partnership Project
4G—Fourth generation
AWGN—Additive white Gaussian noise
CCCF—Complementary cumulative density function
CCDF—Complementary cumulative distribution function
CDMA—Code division multiple access
CP—Cyclic prefix
CTC—Convolutional turbo code
dB—Decibel
FDMA—Frequency division multiple access
IEEE—Institute of Electrical and Electronics Engineers
IFFT—Inverse fast Fourier transform
IMT—International Mobile Telecommunications
ITU—International Telecommunication Union
LTE—Long Term Evolution
MCS—Modulation and coding scheme
MMSE—Minimum mean square error
MS—Mobile station
MU—Multi-user
OFDM—Orthogonal frequency division multiplexing
OFDMA—Orthogonal frequency division multiple access
PER—Packet error rate
PL—Path loss
PUSC—Partial usage of sub-channels
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying
RDMA—Radial space-division multiple access
SDMA—Space division multiple access
SIC—Successive interference cancellation
SINR—Signal-to-interference-plus-noise ratio
SNR—Signal-to-noise ratio
SU—Single-user
TDMA—Time division multiple access
WiMAX—Worldwide Interoperability for Microwave Access

qualities have to be taken into account, there are few theoretical results on system capacity and practically realizable algorithms.

This paper describes a way to improve system throughput under fairness constraints in a cellular system by using well-known low complexity component methods like superposition coding and multi-user detection to design a practically realizable system with substantially improved performance. We restrict ourselves to easily understandable basic concepts and algorithms, which in appropriate combination, allow a considerable performance improvement without claiming to achieve optimum performance. The throughput achieved can be considered as a lower bound for the capacity.

The methods proposed can be used to enhance current orthogonal frequency-division multiple access (OFDMA)-based systems like IEEE 802.16/Worldwide Interoperability for Microwave Access (WiMAX), for example, in the scope of the ongoing IEEE 802.16m project, which aims to develop a new air interface to meet International Mobile Telecommunications (IMT)-Advanced requirements or in 3rd Generation Partnership Project (3GPP*) Long Term Evolution (LTE)-Advanced.

System Description

The wireless access systems considered in this paper consist of a number of base stations and a number of user terminals. The base stations are connected via a fixed network. The user terminals communicate with the base stations via wireless links. We focus on the case where one base station acts as a serving base station and the interference from the other base stations and their users can be treated as uncorrelated noise.

All user terminals share the wireless channel for communication with the base stations. Multiple-access methods like time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA) can be used to split the channel into radio resources that can be assigned to a link between a base station and a terminal. These methods use mutual orthogonal or nearly orthogonal signals for different links. Additionally, when multiple antennas are used at the

base stations, space-division multiple access methods can be used to separate users by orthogonal spatial signatures. The radio resources can be assigned to one or multiple links between a base station and user terminals.

For analyzing the complete communication system, we have to study both the downlink transmission from the base station to the user terminals and the uplink transmission from the user terminals to the base station. In information theory, these two cases are described by a broadcast channel with one transmitter and several receivers for the downlink, and a multiple-access channel with multiple transmitters and one receiver in the uplink. In both cases, we typically have several links between different user terminals and the base station on the same radio resource.

Theoretical Fundamentals

In the section immediately following, we present the basic transmission principles of the proposed system and analytical calculations of the link level performance limits in an additive white Gaussian noise (AWGN) channel both in uplink and in downlink. In the next section, we show how the system performance can be calculated from the link level results.

Link Level Performance

In this section, we study the achievable rates on the single radio resources resulting from dividing the common channel by OFDMA and TDMA into mutual orthogonal radio resources. We will see how throughput can be increased by assigning the single resources to more than one user. To reduce complexity, and since a large part of the possible gain is achieved with a small number of simultaneous users, we restrict ourselves to a two-user case.

In the following, capacities and rates are normalized to the bandwidth, i.e., we consider achievable spectral densities. The signal powers are considered in the bandwidth of one subchannel, which is the smallest allocatable bandwidth in the system.

Uplink. In the uplink, two terminals transmit to the base station simultaneously on the same subchannel. The corresponding schematic of this so-called multiple-access channel is shown in **Figure 1**.

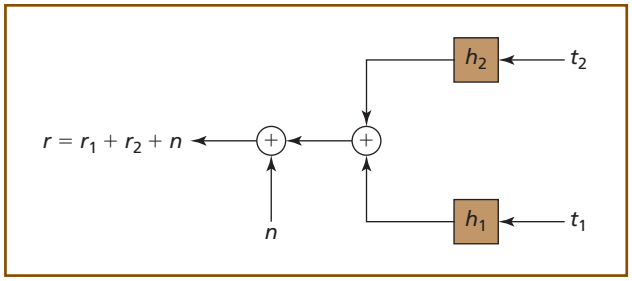


Figure 1.
Two user multiple-access channel in uplink.

The terminals have individual power constraints, i.e., both of them can send simultaneously with a maximum power P.

The uplink receive signal at the base station can be written as

$$r = h_1 t_1 + h_2 t_2 + n,$$

where h_1 and h_2 are the channel gains between terminal 1 and terminal 2, respectively, and the base station, t_2 and t_2 are the corresponding transmit signals of the terminals, and n is noise added in the receiver.

In an AWGN channel, if the mean transmit powers are denoted by P_1 and P_2, the constant channel power gains are $G_1 = |h_1|^2$ and $G_2 = |h_2|^2$, and the signals are assumed to be complex Gaussian distributed, the single user channel capacities normalized to the bandwidth can be written per reference [1] as:

$$C_{SU,1} = \log_2\left(1 + \frac{P_1 G_1}{N}\right)$$

$$C_{SU,2} = \log_2\left(1 + \frac{P_2 G_2}{N}\right),$$

where N is the total power of Gaussian distributed thermal noise in the relevant bandwidth.

If both users transmit simultaneously, the capacity of the two-user channel can be described by the border of the rate region in **Figure 2**. Points A and B are the points where one of the users achieves its single user capacity and the other user transmits at zero rate. All points below and on the colored line are achievable. The dashed lines comprise points with the same weighted sum rate, where the weights are the inverse of single-user rates; the dash-dotted lines

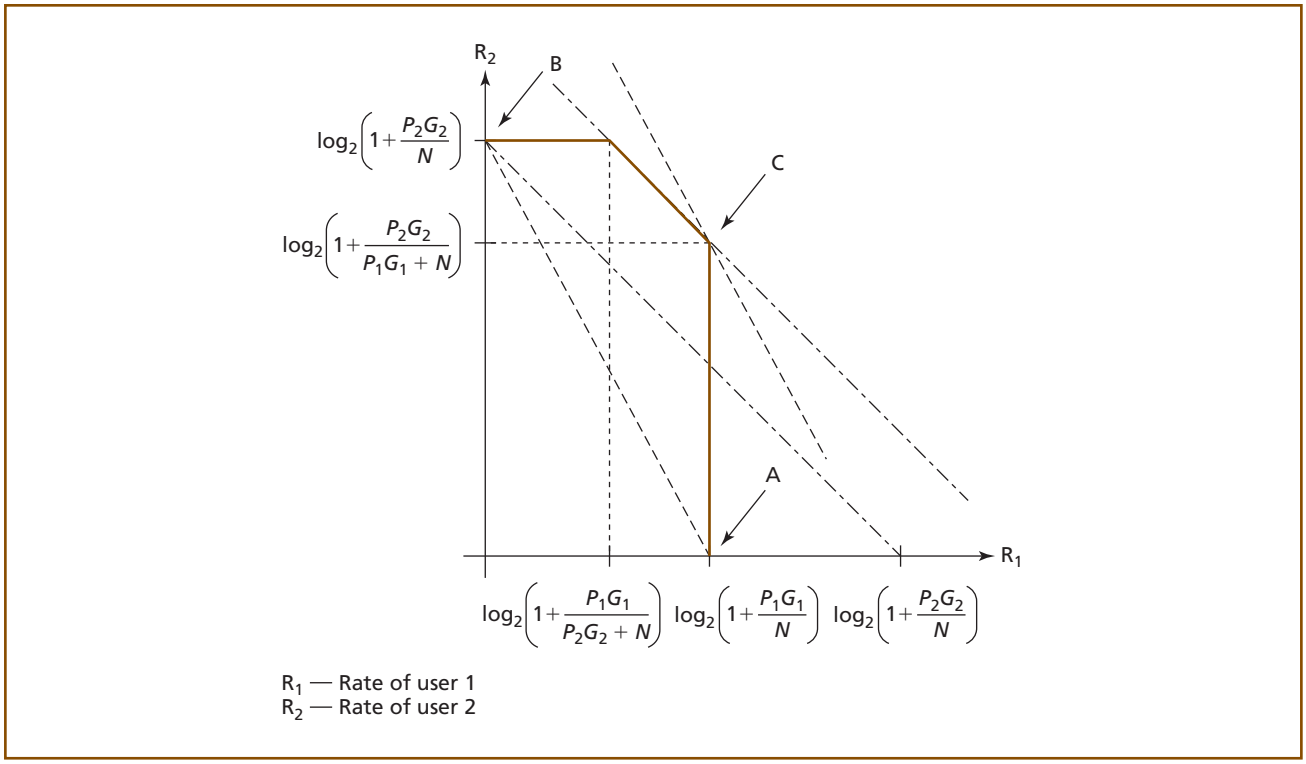


Figure 2.
Rate region for the two-user multiple-access channel in uplink with Gaussian signals.

are those with an equal sum rate. At point C, the weighted sum rate is maximized. This point can be achieved by first decoding signal 2 with signal 1 and the thermal noise as interference, then subtracting it from the sum signal and decoding signal 1 with only thermal noise as interference. Here it is assumed that signal 2 can be ideally compensated for. The simultaneously achievable rates are:

$$C_1 = \log_2\left(1 + \frac{P_1 G_1}{N}\right)$$

$$C_2 = \log_2\left(1 + \frac{P_2 G_2}{P_1 G_1 + N}\right)$$

Typically, signal 1 can come from a far user, and signal 2 from a near user. In this case, and when both are transmitting at the maximum power, signal 2 arrives with higher power at the base station receiver. As can be seen from Figure 2, the sum throughput can be enhanced significantly from that of the single user case, where the subchannel has to be shared in TDMA by the different users, which yields rates on the lower dashed line. The achievable gain is higher for large differences between the channel gains. Fairness is taken into account by maximizing the weighted sum rate instead of the sum rate, which in this case means that the user with the smaller channel gain achieves his single user capacity (point C in the figure).

Downlink. In the downlink, the base station transmits simultaneously over a so-called broadcast channel to two users as shown in **Figure 3**. The two user case transmit signal can be written as

$$t = t_1 + t_2 = a_1 s_1 + \alpha_2 s_2,$$

where s_1, s_2 are normalized transmit signals with mean power 1 and a_1, a_2 are weighting factors to assign fractions of the total transmit power to the different users. In this case, the sum power is limited by the power constraint of the base station. Assuming that the power of the signal sent to user 2 is

$$P_2 = \alpha P,$$

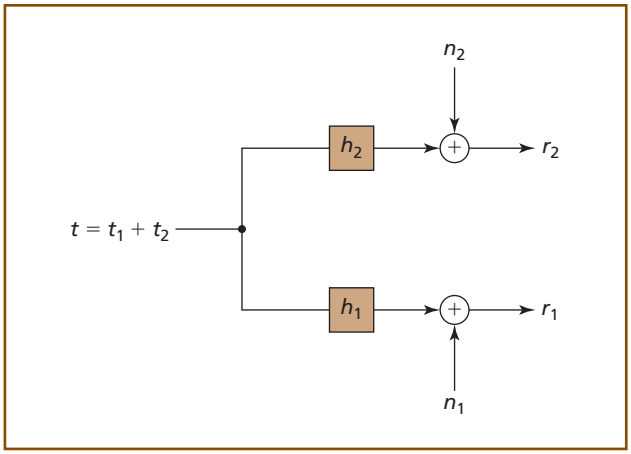


Figure 3.
Two-user broadcast channel in downlink.

the power of the signal sent to user 1 is

$$P_1 = (1 - \alpha)P,$$

where P is the total transmit power and α is the power distribution factor. Since the two signals are uncorrelated, the total power of the sum signal is the sum of the powers of its component signals.

The receive signal at terminal 1 is

$$r_1 = h_1(a_1 s_1 + a_2 s_2) + n_1.$$

And the receive signal at terminal 2 is

$$r_2 = h_2(a_1 s_1 + a_2 s_2) + n_2$$

where n_1 and n_2 represent Gaussian distributed noise signals. In an AWGN channel, the capacities for simultaneous transmission to two users can be given as a function of the power distribution factor:

$$C_1(\alpha) = \log_2\left(1 + \frac{(1-\alpha)PG_1}{\alpha PG_1 + N}\right) = \log_2\left(\frac{PG_1 + N}{\alpha PG_1 + N}\right)$$

$$C_2(\alpha) = \log_2\left(1 + \frac{\alpha PG_2}{N}\right) = \log_2\left(\frac{\alpha PG_2 + N}{N}\right).$$

where N is the noise power. The region of the achievable rates is shown in **Figure 4** (see, for example, reference [2]). Again, points A and B reflect the

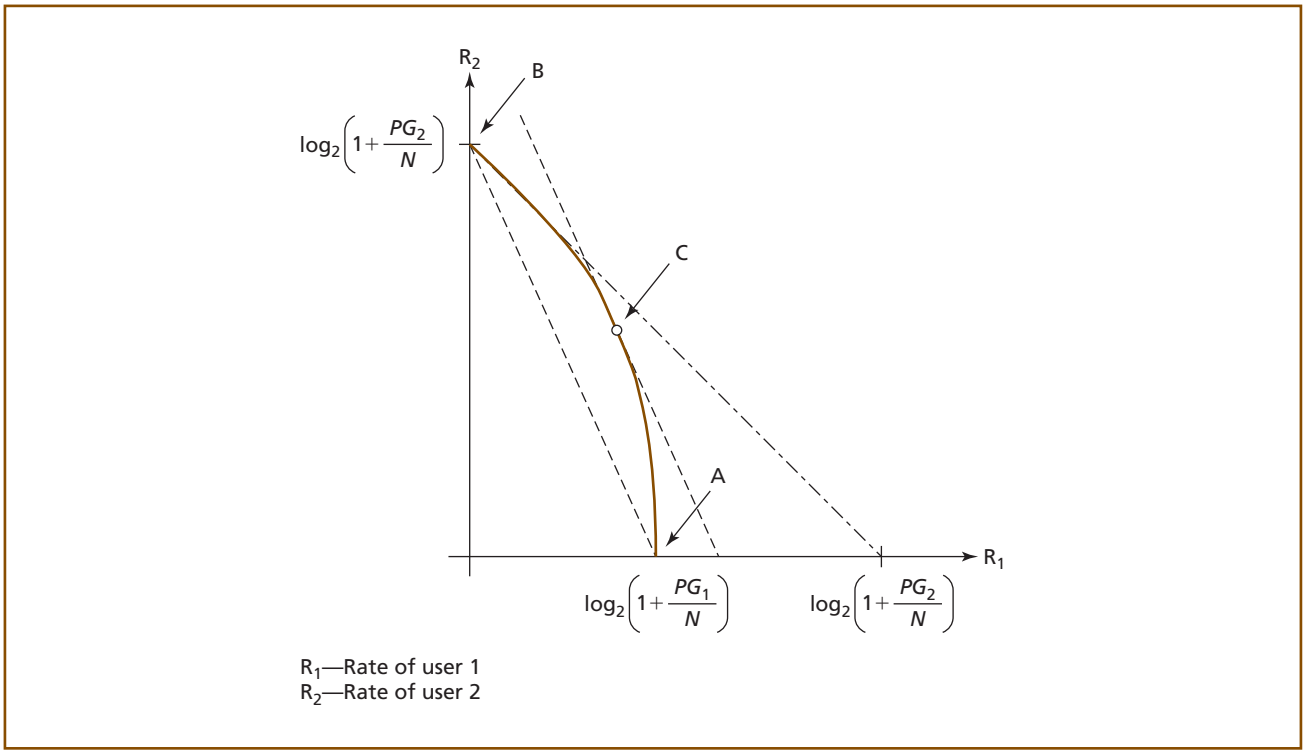


Figure 4.
Rate region for the two-user broadcast channel in downlink with Gaussian signals.

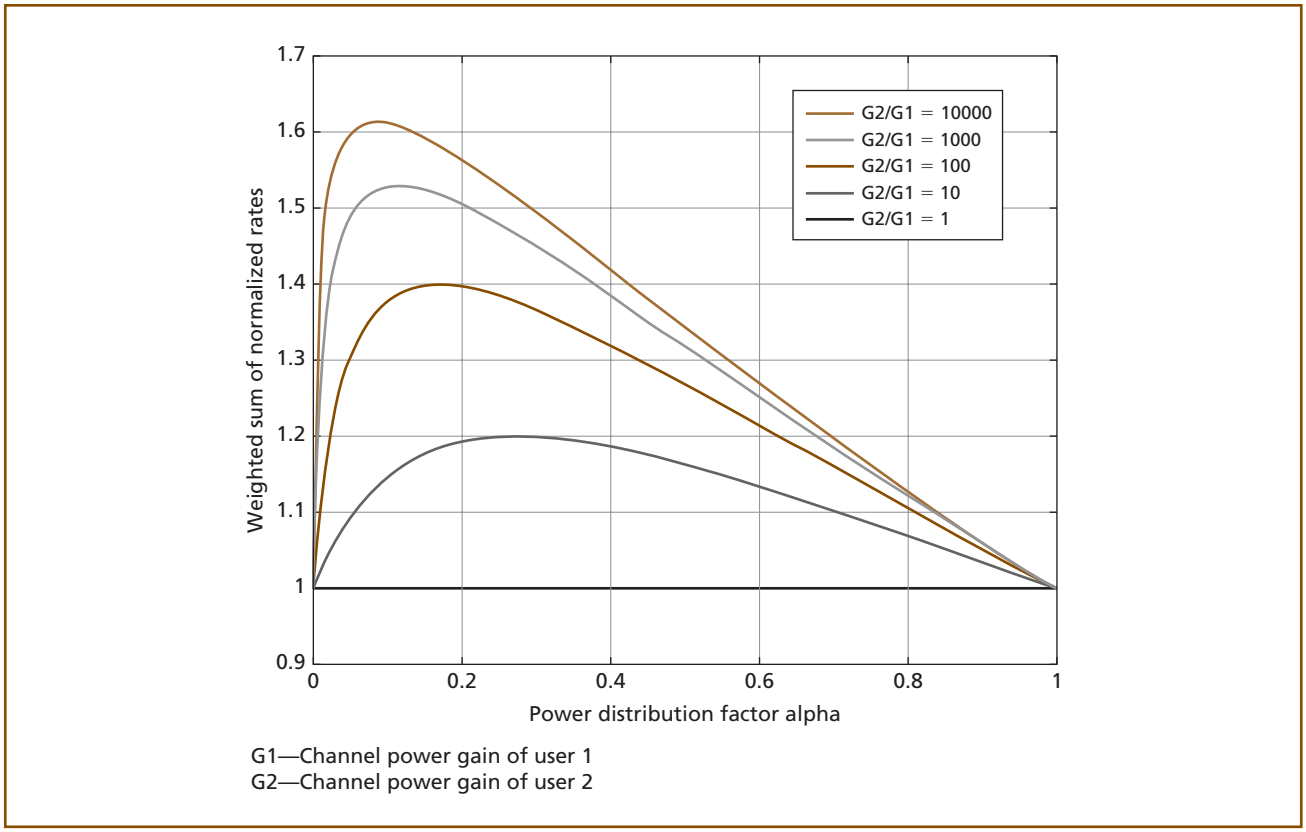


Figure 5.
Weighted sum rate as a function of the power distribution factor α for different ratios (G_2/G_1) of the channel power gains.

single-user cases and point C is the operating point with maximized weighted sum rate.

Figure 5 shows the weighted sum rate as a function of the power distribution factor α

$$C = \frac{1}{C_1(0)} C_1(\alpha) + \frac{1}{C_2(1)} C_2(\alpha)$$

with the inverse single user capacities as weights for different ratios of the channel power gains. Each of the curves has a maximum for a certain power distribution between the two user signals.

The optimum power distribution factor α_{opt}, for which the weighted sum rate is maximized, can be calculated by setting the first derivative of the weighted sum rate with respect to α to zero:

$$\alpha_{\text{opt}} = \frac{NG_1 \ln\left(\frac{PG_2 + N}{N}\right) - NG_2 \ln\left(\frac{PG_1 + N}{N}\right)}{PG_1 G_2 \ln\left(\frac{PG_1 + N}{PG_2 + N}\right)} \quad (1)$$

Figure 6 shows the gain in sum rate when adding a second user to a radio resource that would otherwise be used by a single cell edge user with a signal-to-noise ratio (SNR) of about 0 decibels (dB). If the channel power gains differ by a factor of 1000 (30 dB), the sum rate is more than seven times the cell edge rate. Since in conventional systems a large fraction of resources is allocated to terminals near the cell edge, this gain significantly influences the overall system performance.

System Level Performance

A measure for the system level performance is the aggregate system throughput under fairness constraints. Some fairness constraint is necessary, since optimizing the sum rate of all users results in a system that primarily serves users with good channels.

When serving one user per radio resource, the aggregate system throughput is a weighted sum over the rates $R_{\text{SU},i}$ achieved in all radio resources.

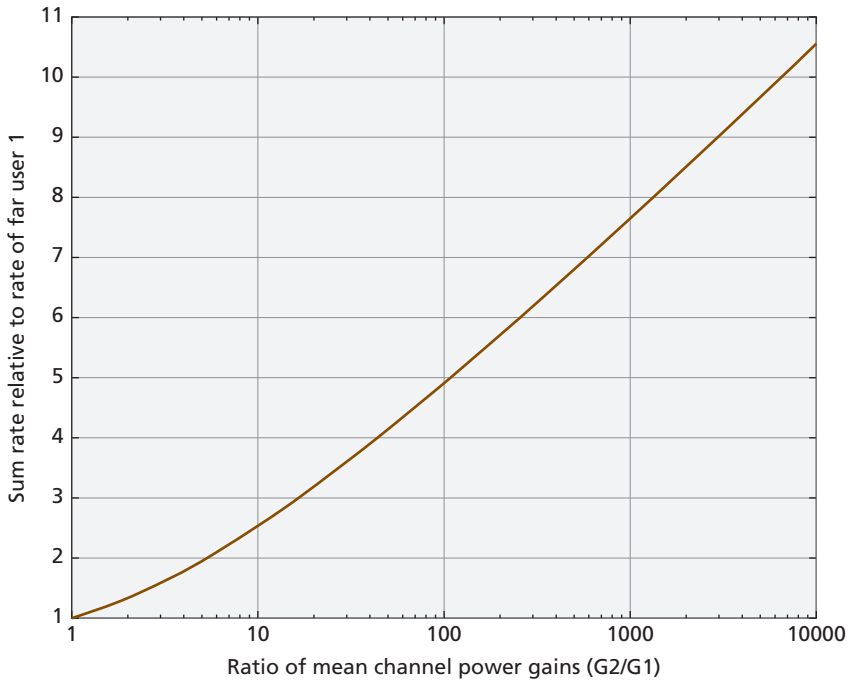


Figure 6.
Sum rate for two users relative to the rate of a cell edge user.

The weights a_i are the fractions of the radio resources used for the corresponding link i:

$$\sum_{i=1}^{L} a_i R_{\mathrm{SU},i} \text{ with } \sum_{i=1}^{L} a_i = 1$$

where L is the number of resources.

In the case where two users are assigned to each radio resource, the throughput is

$$\sum_{i=1}^{L} a_i [R_{i,1} + R_{i,2}]$$

where $R_{i,1}$ and $R_{i,2}$ are the rates achieved for each of the two links on the same resource.

A scheduling algorithm groups the users to pairs and determines the weights a_i for all the pairs of users.

Since the achievable rates depend on the signal-to-interference-plus-noise ratio (SINR) of the corresponding links, the system throughput according to the above equations depends on the distribution of channel gains G that result from the distribution of distances of the terminals from the base station and a path loss model.

Uplink. In the uplink, $R_{i,1}$ can be the same as $R_{\mathrm{SU},i}$ since there is no sum power constraint and the compensation of the second signal can be considered ideal. The throughput of cell edge users can be improved by assigning more radio resources to these users than in the case where only one user is assigned to each radio resource and serving users near to the base station on the same radio resources as cell edge users.

Downlink. In the downlink, $R_{i,1}$ will be slightly smaller than $R_{\mathrm{SU},i}$, since the sum power of the two signals is constrained and there is additional interference from the second user. This is also reflected in the rate diagram in Figure 4, where we can see that in point C the rate of user 1 is smaller than the single-user rate in point A. On the system level this can be compensated by assigning a larger fraction of the radio

resources to the far users, similar as in the uplink. But gains in aggregate system throughput will be smaller in downlink than in uplink.

Performance Evaluation in a Realistic System Setup

To analyze how well the theoretical gains can be realized in practice, we extend the investigations of the proposed non-orthogonal multi-user access scheme towards realistic systems and scenarios. For this purpose, we use a multi-link physical layer simulator that allows detailed uplink and downlink link level simulations of a WiMAX/IEEE 802.16e based system setup. The system parameters of the simulation are listed in **Table I**. The results obtained will extend our knowledge related to the feasibility of reaching the theoretical gains in a realistic scenario. Compared to the theoretic approach, the following assumptions are different:

- Random binary data source and realistic modulation and coding (instead of Gaussian distributed amplitudes).
- Fading channel model (in addition to AWGN).
- Receiver algorithms: equalization, demodulation, and decoding.

Table I. System parameters for simulation.

Parameter	Value
FFT size	1024
Carrier frequency	2.5 GHz
Bandwidth	10 MHz
Burst size	Such that 240 subcarriers are occupied
FEC block sizes	Following the IEEE 802.16e algorithm for CTC coding
Modulations	QPSK, 16 QAM, 64 QAM
Coding rates	1/2, 3/4, 2/3 (only for 64 QAM)
Cyclic prefix length	1/8 of the FFT size. → 128 samples
Power control	Ideal, on frame basis

CTC—Convolutional turbo code
FEC—Forward error correction
FFT—Fast Fourier transform
IEEE—Institute of Electrical and Electronics Engineers
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying

Link level simulation results with realistic assumptions will allow the identification of operational multi-user configurations (modulation and coding schemes of different user pairs) and the scenarios/SNR regions in which they can be applied. These results build the foundation of a basic system level simulation where users are distributed in a realistic way over a single cell and heuristic scheduling allows the evaluation of cell throughput gains compared to orthogonal single-user access. The terms single-user (SU) access and multi-user (MU) access are used in the simulation context regarding the occupation of one single radio resource. This means that, even though orthogonal OFDMA is a MU access method, the term SU access is used if only one user is scheduled on one time-frequency resource of the OFDMA frame. With MU access, we denote the superposition of two users on one OFDMA resource in a non-orthogonal way.

In the section following, the simulator functionality is described in more detail before the different simulation scenarios and the results are presented.

System Description and Parameterization

Superposition with two users, as described in the previous sections, is employed such that their bursts overlap entirely. In the link level simulation, the two superimposed users are scheduled alone in one frame and the data is placed on one single orthogonal frequency division multiplexing (OFDM) symbol. The allocation size is fixed and spans 240 data subcarriers. For channels with high frequency selectivity, users with low mobility, and perfect channel knowledge at the receiver, the performance of this allocation is comparable to a diversity exploiting allocation scheme such as partial usage of sub-channels (PUSC) in WiMAX.

Uplink and downlink simulation scenarios are set up by combining and appropriately parameterizing the basic functional blocks: transmitter, receiver, and channel. Before introducing the architecture of the uplink and downlink simulator, the functionality of these three basic modules is explained.

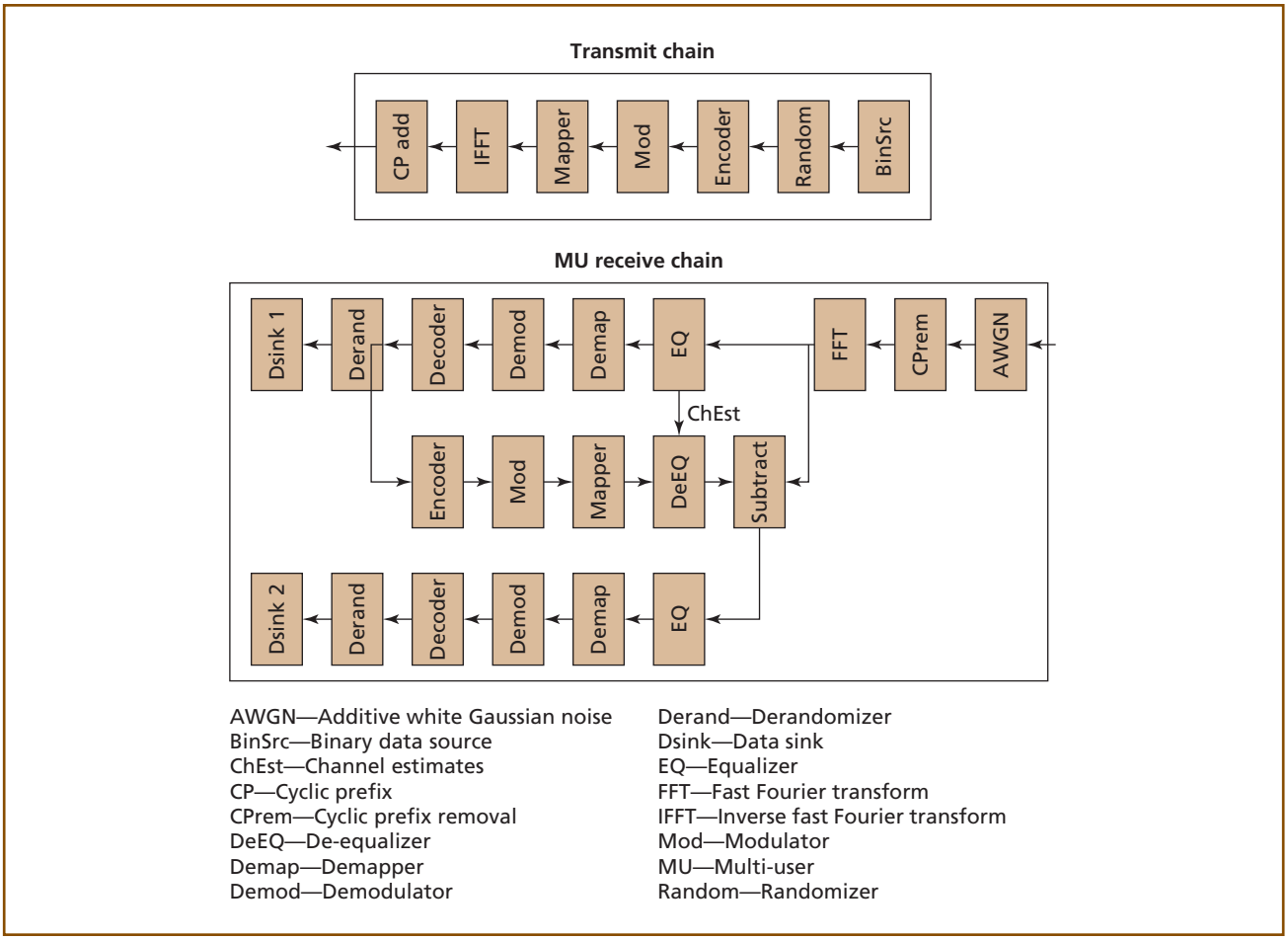


Figure 7.
Functional block diagram of transmitter and MU receiver.

Transmit chain. As shown in **Figure 7**, a random binary source generates the bit stream for transmission. The randomizer, convolutional turbo code (CTC) encoder and modulator are compliant with the IEEE 802.16e standard [3]. The sequence of complex values at the output of the modulator is mapped on a contiguous block of subcarriers out of the 1024 available subcarriers in one OFDM symbol. After the 1024-point inverse fast Fourier transform (IFFT) a cyclic prefix (CP) with 1/8 of the sequence length (128 samples) is added. The transmit power can be varied to simulate a broad set of different superposition scenarios.

Channel. The signal is distorted by a multipath fading model with a variable path loss component. Path loss is equal to the inverse of the channel power gain mentioned in the theory section. The multipath environment is modeled following the International Telecommunication Union (ITU) Pedestrian B delay profile and the Doppler spectrum represents the Jakes model.

For each frame that is transmitted, an independent random channel is generated.

Receive chain. The receiver implements state-of-the-art successive interference cancellation (SIC) as described in [4] but adapted to the present system. The receiver has knowledge of the modulation and coding schemes (MCSs) and power levels of the two combined user signals and decodes the stronger signal first. The incoming composite signal is distorted by additive white Gaussian noise of fixed power.

Detection and decoding of the first user is similar to a common receiver with the signal of user 2 acting

as interference in addition to the noise. The signal is equalized by a minimum mean square error (MMSE) algorithm with perfect channel knowledge and demodulated with a realistic (non-ideal) bit-by-bit soft demapper. Then, iterative turbo decoding with six iterations on low precision quantized values is performed. As shown in Figure 7, at this point, data can be derandomized and the first user's bit stream is available or re-encoded, modulated, and mapped to the frame again. In the de-equalizer, the impact of the channel is applied and the resulting signal is ideally the signal that would have been received in the absence of any other users. After subtracting this reconstruction from the composite signal received, the detection and decoding of the second signal is done in a way similar to that of the first user.

Uplink Multi-Link Simulation

The uplink scenario is modeled as depicted in **Figure 8**, where two terminals transmit independent information at the same time on the same radio resource (subcarriers). The multiple access channel fades independently for the two users and the receiver sees a superposition of the two signals. The power of each user signal in the composite receive signal (directly related to its SNR) is an important metric and is directly linked to the performance of MU communication. Through SNR1 and SNR2, we understand the signal-to-noise ratio of one user's isolated signal in the absence of the second interfering signal. Simulation results can be described by a three-dimensional error rate surface as a function of SNR1 and SNR2 for each of the two users and different MCS configurations, i.e., the MCS of the two users. It can be observed that the user 1 (far end) packet error rate (PER) surface is usually more restrictive since it has as a bigger packet error rate than that of user 2. For different MCS configurations, analysis of the simulation output shows that this is true in more than 90 percent of the SNR points and if it is not, the difference between PER 1 and PER 2 are on average smaller than one percent. This fact is somehow intuitive if we remember that the correct decoding of user 1 depends directly on the decoding of user 2. If the two users have unequal MCSs, the impact is less obvious. The operation regions are therefore mainly defined by the user 1 performance. A set of simulation outcomes for user 1 is presented in **Figure 9** for an ITU Pedestrian B, 3 kilometer per hour (km/h) channel. The two-dimensional PER contour lines at PER $= 10^{-2}$ are plotted as a function of the two users' SNR values for different MCS configurations. All the points enclosed in the upper right area of such a curve have PER $\leq 10^{-2}$.

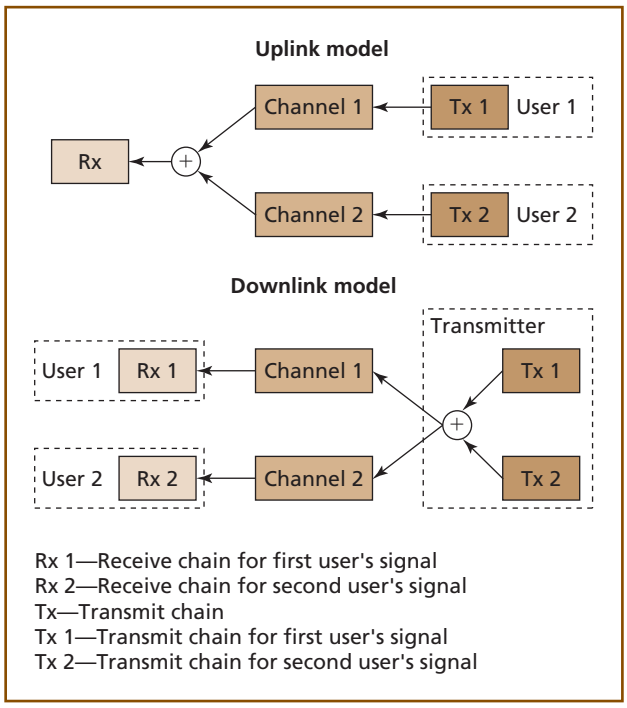


Figure 8.
Uplink and downlink model of multi-link simulation.

Knowledge of SNR1 and SNR2 at the receiver is sufficient to decide about the best MCS configuration. The SNR information, in combination with the operation regions from the previous figure, is crucial for the scheduler in a real system to allocate resources. This means, that the proposed method of superposition only depends on long term channel statistics like path loss which is an important advantage over other physical layer techniques for throughput enhancement. In addition to the channel statistics, a real scheduler would have knowledge of the transmit power of each terminal to decide if it is beneficial to increase/decrease transmit power at certain terminals.

From the previous operation region figure, we can see that several MCS configurations lie within

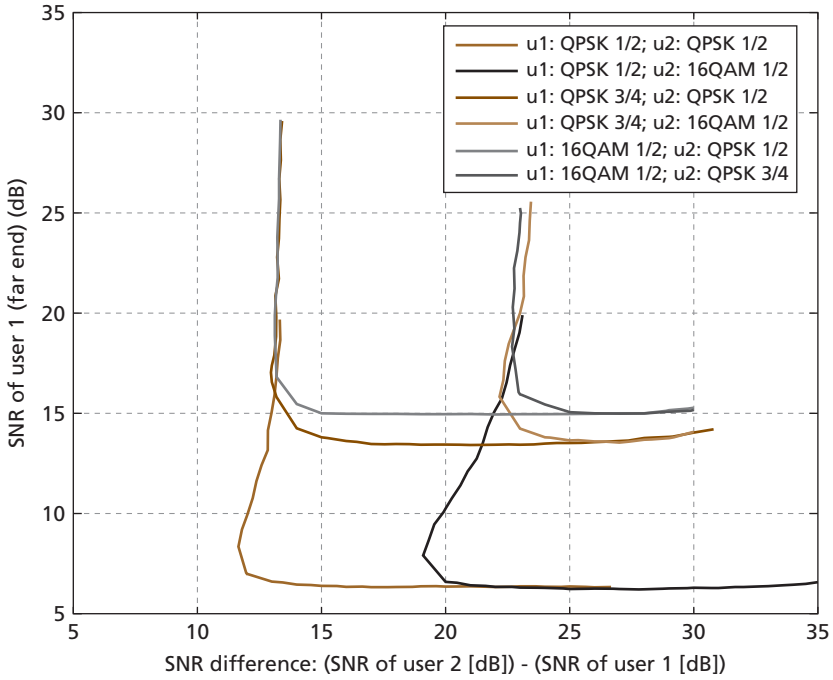


Figure 9.
User 1 simulation outcomes for Pedestrian B, 3 km/h channel.

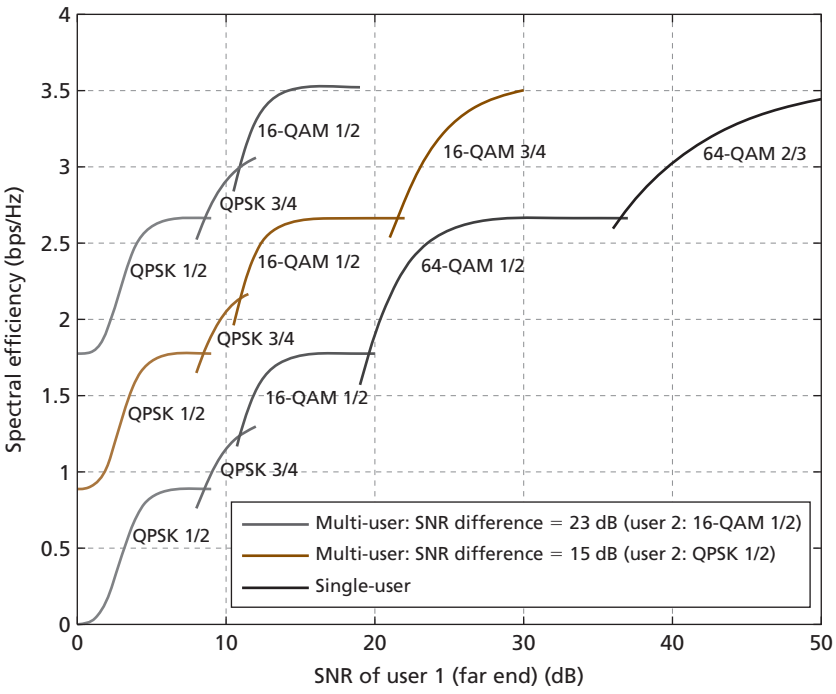
reasonable regions below SNR values of about 40 dB and can be used later on for system level investigations. Several MCS configurations that are not shown here with more robust MCS, i.e. with repetition coding, operate in even smaller SNR regions.

In **Figure 10**, the results are presented in a way to show the gain of MU transmission over SU transmission on a Pedestrian B, 3 km/h channel in specific path loss scenarios indicated by spectral efficiency. The baseline performance for comparison is the well-known SU curve of spectral efficiency for different SNR values (lower curve). The spectral efficiency on the ordinate is calculated as

$$e = (1 - PER) \cdot \frac{R_{CTC} \cdot M}{(1 + R_{CP})} \qquad (2)$$

where R_{CTC} is the code rate of the CTC, M is the modulation order (e.g., 4 for 16 QAM) and R_{CP} is the CP ratio. **Table II** lists the maximum spectral efficiencies (PER = 0) of different MCS configurations. In the MU case, the spectral efficiency is obtained from the sum of the two users' spectral efficiencies. The first column of Table II follows formula (2), whereas in the second column of Table II, the division by $(1 + R_{CP})$ is omitted to indicate spectral efficiency without taking into account CP. For the different spectral efficiency and throughput indications in this document, the values including CP are used.

The multi-user curves in Figure 10 show the gain in spectral efficiency obtained from transmitting another, stronger signal on top of the single user signal, which is shown by the lower curve. The user 2 signal is stronger due to its position in the cell (path loss), or its transmit power. The SNR value on the abscissa still indicates the SNR of user 1 (far end) even though the signal is now superimposed by the signal of user 2.



dB—Decibel
MU—Multi-user
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying
SNR—Signal-to-noise ratio
SU—Single-user

Figure 10.
Effective spectral efficiency of SU and MU access in uplink (Pedestrian B, 3 km/h).

For the two MU setups, shown by the upper curves, the path loss difference between the two users is fixed at values of 15 dB and 23 dB, and the modulation scheme of user 2 is QPSK 1/2 and 16 QAM 1/2 respectively.

It can be seen, especially for scenarios where user 1 has low SNR and is therefore using a robust MCS, that the spectral efficiency can potentially be increased by a factor of three. Further, it is important to note, that real cell edge scenarios show SNR values of zero and below. This demands repetition coding, (on top of QPSK rate 1/2) leading to higher robustness and again lower spectral efficiency than that of QPSK 1/2. It is obvious that in these scenarios, the gain obtained through MU access is even greater.

Downlink Multi-Link Simulation

Based on the transmit and receive chains presented, a downlink model is defined like that depicted in Figure 8. The signals from the two transmit chains residing now in one single transmitter (base station) are summed, and the resulting signal enters the broadcast channel. The composite signal received on each terminal has experienced an independent channel with different fading sequences and path losses.

The basic situation in downlink is different from that of the uplink, especially because power has to be shared between the two users that are served at the same time. The scheduler in the simulation distributes the power according to the optimum power distribution formula, shown in equation (1), and accounts for fairness by optimizing the weighted sum rate as described in the related section.

Further, we assume fixed cell size and fixed total transmit power, chosen such that a single user can be served at low PER (e.g., 10^{-2}) with the most robust modulation and coding scheme in the simulation

Table II. Maximum spectral efficiencies for different single-user and multi-user configurations.

Modulation and coding scheme	Maximum spectral efficiency (bps/Hz)	
	Consider CP	Without CP
Single-user		
QPSK 1/2 Rep 4	0.222	0.25
QPSK 1/2 Rep 2	0.444	0.5
QPSK 1/2	0.889	1
QPSK 3/4	1.333	1.5
16 QAM 1/2	1.778	2
16 QAM 3/4	2.667	3
64 QAM 1/2	2.667	3
64 QAM 2/3	3.556	4
Multi-user		
QPSK 1/2; QPSK 1/2	1.778	2
QPSK 1/2; QPSK 3/4	2.222	2.5
QPSK 1/2; 16 QAM 1/2	2.667	3
QPSK 1/2; 16 QAM 3/4	3.556	4
QPSK 3/4; QPSK 3/4	2.667	3
QPSK 3/4; 16 QAM 1/2	3.111	3.5
QPSK 3/4; 16 QAM 3/4	4	4.5
16 QAM 1/2; 16 QAM 1/2	3.556	4
16 QAM 1/2; 16 QAM 3/4	4.444	5
16 QAM 3/4; 16 QAM 3/4	5.333	6

bps—Bits per second
CP—Cyclic prefix
Hz—Hertz
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying

(QPSK, coding rate 1/2). In reality, in a cell-edge scenario, repetition coding is needed to deal with lower SNR values but here we limit the investigation to the smallest MCS without repetition coding. The SNR values for QPSK 1/2 at PER = 10^{-2} are 6 dB for Pedestrian B, 3 km/h and 2.5 dB for AWGN. The position of user 1 (far end) in the cell is fixed by setting its path loss to 50 dB. Therefore, a maximum transmit power of 56 dB (relative to the noise level at the receivers) in the case of Pedestrian B, 3 km/h channel is necessary. For an AWGN simulation a margin of 1.5 dB is added on top of the necessary 2.5 dB SNR because the AWGN SU PER curve is steeper and slight losses in SNR (due to interference coming from the superimposed user and non-ideal SIC) considerably increase the PER. The only variable parameter is the path loss difference between the two users, and with that, the position of the second (near end) user in the cell. **Table III** summarizes the above mentioned downlink simulation parameters for AWGN and Pedestrian B, 3 km/h simulations.

As in the uplink, the total spectral efficiency of both users together is calculated for different MCS configurations based on the simulation results over the Pedestrian B, 3 km/h channel. In **Figure 11**, for different MCSs of user 2 (user 1 is fixed with QPSK 1/2), the spectral efficiency is plotted against the path loss difference between the two users. The dashed line indicates the maximum reachable SU spectral efficiency of user 1 with QPSK 1/2. The MU spectral efficiency obviously depends on the path loss difference between the two users (abscissa) in a way such that for higher values, where user 2 is closer to the base station, the spectral efficiency becomes higher. We observe that like in uplink, spectral efficiency gains in the downlink are considerable if a user has to be served with a low MCS due to its bad channel (high path loss), and another user can be superimposed with enough path loss difference.

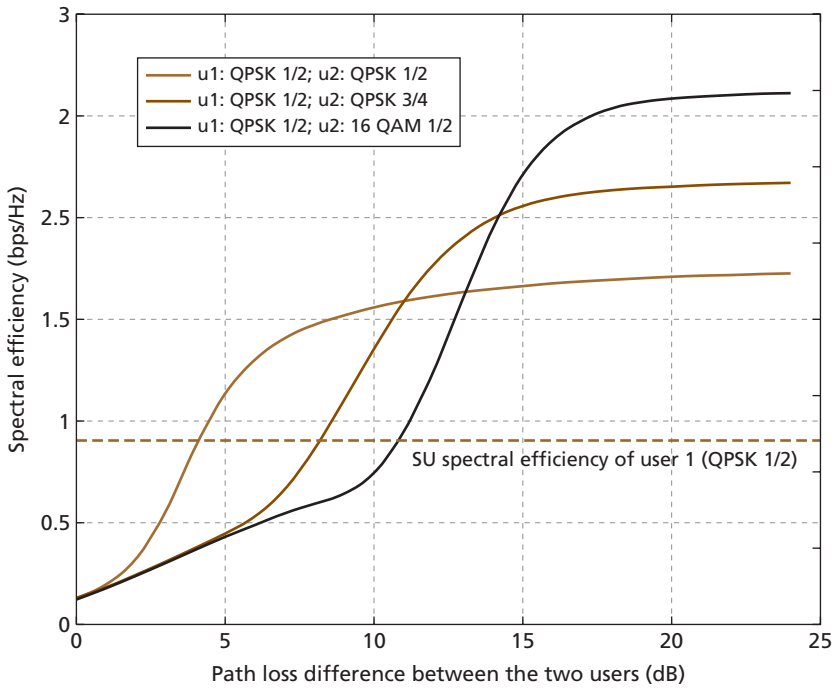
Uplink Cell Throughput Simulation

In the previous link level performance investigations, gains were shown in scenarios where one user is in a low SNR region and has a sufficient difference in path loss from another user. In this chapter, realistic gains over the entire cell shall be identified with real scheduling of user pairs and user positions uniformly distributed over the cell area. The SNR distribution over the cell is obtained from a log-distance path loss model with a path loss exponent of 3.5. The users are paired (scheduled) based on the operation regions (SNR1, SNR2) provided by the previous multi-link simulation results which are taken as abstraction of the physical layer. Additionally, repetition coding with 1 and 3 repetitions is used in combination with QPSK 1/2 to be able to simulate realistic

Table III. Assumptions for the downlink multi-link simulation scenario.

Parameter	Value	
	AWGN	Pedestrian B, 3 km/h
Far end user channel gain	−50 dB	−50 dB
Far end user modulation and coding scheme (necessary SU SNR at PER = 10^{-2})	QPSK 1/2 (2.5 dB)	QPSK 1/2 (6 dB)
Far end user margin with respect to single-user QPSK 1/2 operation point PER = 10^{-2}	1.5 dB	0 dB
Resulting necessary total transmit power	54 dB (relative to noise level)	56 dB (relative to noise level)

AWGN—Additive white Gaussian noise
dB—Decibel
km/h—Kilometers/hour
PER—Packet error rate
QPSK—Quadrature phase shift keying
SNR—Signal-to-noise ratio
SU—Single-user



bps—Bits per second
dB—Decibel
h—Hour
Hz—Hertz
km—Kilometer
MU—Multi-user
QAM—Quadrature amplitude modulation
QPSK—Quadrature phase shift keying
SNR—Signal-to-noise ratio
SU—Single-user

Figure 11.
Effective spectral efficiency of MU access in downlink with user 1 near the cell border (Pedestrian B, 3 km/h).

cell edge situations. For each SNR1/SNR2 combination, the MCS configuration with the highest sum spectral efficiency is chosen by the scheduler.

The modeling of the cell assumes 30 users uniformly spread over one sector of the circular cell area. The maximum transmit power is fixed such that the SNR at the cell edge is −3 dB. No antenna gains, implementation losses, or other effects of this kind are considered.

We assume all the mobile stations (MSs) are transmitting at maximum power as long as the SNR of the strong signal at the receiver stays below reasonable values. The performance of MU access is compared to SU access where every user is scheduled in a way to get the same amount of orthogonal resources (e.g., time) that we are calling one resource unit U. Over an AWGN channel, this corresponds to proportional fair scheduling. The total available orthogonal physical resource for scheduling all of the N users is equal to $N \cdot U$. The MU scheduling mechanisms proposed herein will not subdivide resource unit U, but possibly will not share available resource units equally among all the users. The scheduler bases its decisions on the radial dimension (path loss) only; it is assumed that the path loss to each MS is known.

The MU gains are strongly linked to the scheduling algorithms used. In this section, simple, heuristic scheduling algorithms and their cell throughput gains over SU access are presented for transmission on a Pedestrian B, 3 km/h channel. These scheduling algorithms can mainly be divided into two groups: equally fixed resource allocation (A1), and variable resource allocation (A2). The first group covers allocations where orthogonal resources are equally shared between users, i.e., each user occupying the same number of resources. This is obtained by building $\frac{N}{2}$ pairs and assigning $2U$ resources to each pair. In this way, as opposed to SU access, in MU access, a user is always scheduled on $2U$ resources. An obvious upper bound of gain in cell throughput of this pairwise equal resource allocation scheduling is doubled spectral efficiency and can be reached only if all the users are paired and can transmit superimposed with the same MCS used in SU access. In this case, every user's throughput would be doubled independently of his channel, which provides for very fair scheduling. Scheduling algorithms of type A1 have the advantage that the SU throughput for each user can be easily guaranteed to be the lower bound of throughput. This is done by comparing the MU throughput obtained for each pair to the users' SU throughputs and choosing the better one. The type A1 algorithms proposed here use this approach.

The non-uniform path loss distribution of the users makes it impossible to always find $\frac{N}{2}$ user pairs in good MU operation regions, especially when transmitting over fading channels where larger path loss differences between users are necessary. Therefore, it is beneficial (for the exploitation of the channel resource and therefore the cell throughput) to provide flexible allocation of resource units (A2) where the number of times users are scheduled in pairs can differ for different users. This increased freedom makes it more difficult to control the fairness in terms of a lower bound of throughput as in the A1 algorithms.

In the following, four different scheduling algorithms for non-orthogonal MU access (S1 − S4) are described and the resulting cell throughput gains with respect to orthogonal SU access are listed in **Table IV**. The column titled "MU fraction" indicates the fraction of users that were actually scheduled with non-orthogonal superposition because if the scheduling of a pair does not lead to higher spectral efficiency than

Table IV. Cell throughput gains of non-orthogonal MU access compared to SU access in uplink over a Pedestrian B, 3 km/h channel.

Scheduling algorithm		Gain	MU fraction
S1	Random	29%	38%
S2	Minimum PL difference variation	40%	58%
S3	Maximum PL difference	43%	54%
S4	Variable MU resource allocation	110%	100%

AWGN—Additive white Gaussian noise
km/h—Kilometers/hour
MU—Multi-user
PL—Path loss
SU—Single-user

Panel 2. Variable MU Resource Allocation Algorithm (S4).

```
sort users descending by their channel gains;
for each user x from sorted list
{
  group A = users y, that can be paired with user x because MU operation
     regions for the pair (x, y) exist;
  group B = users y from group A that form a pair (x, y) where the far user's
     throughput in combination with superposition is at least half of his SU
     throughput;
  group C = users y from group B that have been scheduled fewest up to now;
  Select user y from group C such that the pair (x, y) has highest sum
     throughput;
  Allocate to the pair (x, y) one resource unit;
}
```

with orthogonal SU access, the user is allocated a resource orthogonally (SU access).

The first algorithm, random scheduling (S1, group A1), builds $\frac{N}{2}$ pairs by randomly choosing two users out of the set of N users.

The minimum path loss (PL) difference variation scheduler (S2, group A1) sorts users in descending order by their channel gains $G_{PL1}, G_{PL2}, \ldots, G_{PLN}$ and pairs are built as follows:

$$
\begin{aligned}
&p1 : G_{PL1}, G_{PLN/2+1}\\
&p2 : G_{PL2}, G_{PLN/2+2}\\
&\ldots\\
&pN/2 : G_{PLN/2}, G_{PLN}
\end{aligned}
$$

For each pair, the best possible MCS configuration in terms of sum spectral efficiency is chosen.

The maximum path loss difference scheduler (S3, group A1) also sorts users in descending order by their channel gains $G_{PL}1, G_{PL}2, \ldots, G_{PLN}$ but pairs are built as follows:

$$
\begin{aligned}
&p1 : G_{PL1}, G_{PLN}\\
&p2 : G_{PL2}, G_{PLN-1}\\
&\ldots\\
&pN/2 : G_{PLN/2}, G_{PLN/2+1}
\end{aligned}
$$

Again, for each pair, the best possible MCS configuration in terms of sum spectral efficiency is chosen.

In variable MU resource allocation (S4, group A2), the goal is pairwise allocation of all the users including the ones that in the previous algorithm end up in pairs that are not suited for non-orthogonal MU access (and thus are scheduled with SU access). This problem is linked to the fact that path loss distribution over the cell is not uniform—only a few users experience limited path loss, while there are a significant number of users in high path loss regions. The algorithm presented in **Panel 2** assigns the close end users (small path loss) more often to pairs, i.e., gives them more shared resources, and tries at the same time to keep the number of pairs for each user as balanced as possible for fairness reasons. Additionally, a pair is only scheduled if the far end user has at least half the SU throughput with non-orthogonal MU access.

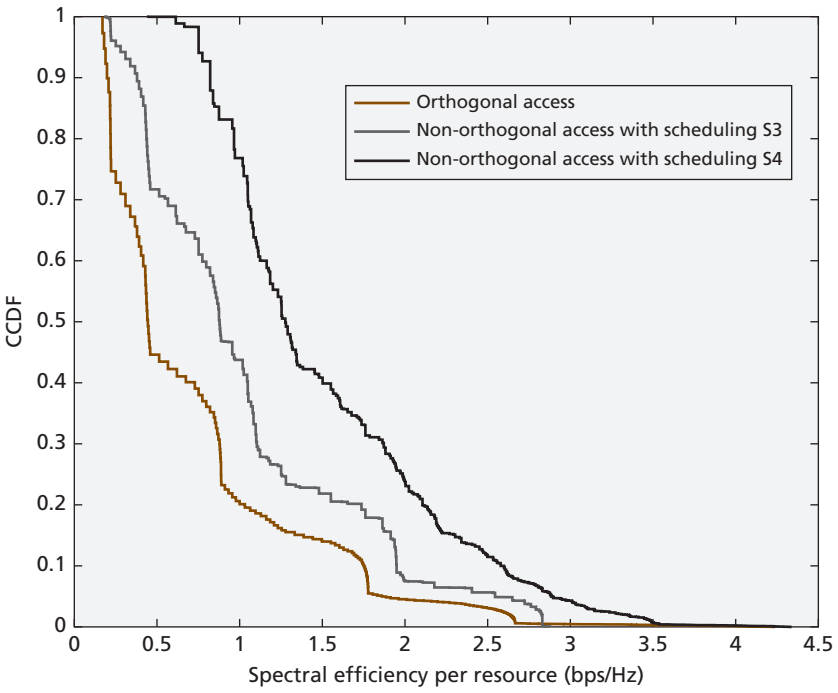
As expected, the highest gain in cell throughput is obtained with S4. The variable MU resource allocation algorithm leads to a cell throughput gain of around 100 percent. Compared to gains of other physical layer transmission techniques, this is significant, especially at that low complexity.

The following provide good insight into the gains and fairness obtained by non-orthogonal MU access:

- Complementary cumulative distribution function (CCDF) of spectral efficiency, shown in **Figure 12**, and
- The CCDF of user throughput, shown in **Figure 13**.

Spectral efficiency values take into account the overhead of the cyclic prefix and the user throughput values additionally consider 12 percent guard subcarriers.

Figure 12 demonstrates that the gains of non-orthogonal MU access are primarily realized by better exploiting the channel for resources where a single



bps—Bits per second
CCDF—Complementary cumulative distribution function
Hz—Hertz
MU—Multi-user
SU—Single-user

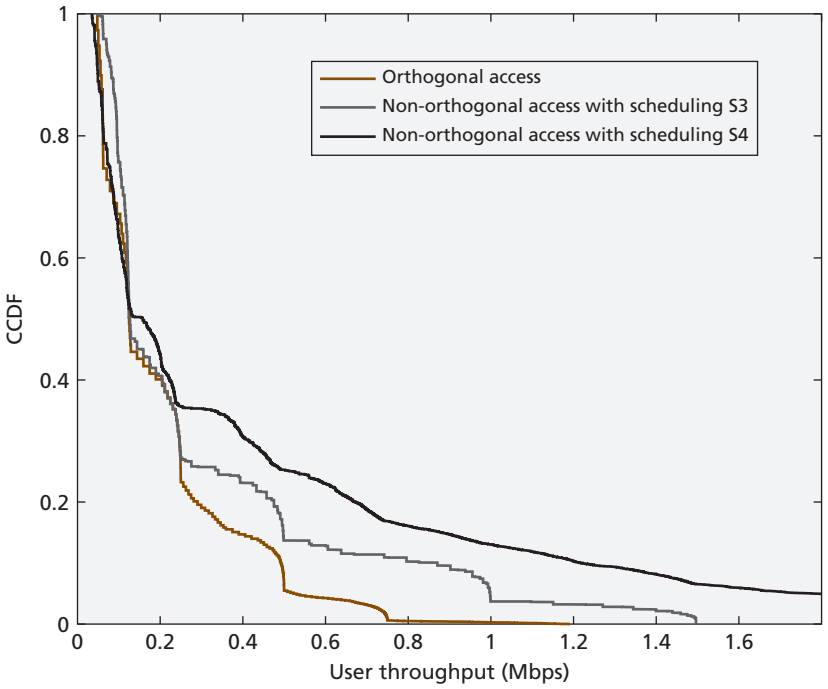
Figure 12.
Spectral efficiency per resource of orthogonal (SU) access and non-orthogonal (MU) access with scheduling algorithms S3 and S4 (Pedestrian B, 3 km/h).

user has low spectral efficiency (e.g., the cell edge). For example, the 25 percent of resources with lowest spectral efficiency in SU access are showing four to five times the spectral efficiency in MU access with scheduling algorithm S4. The related fairness issues can be seen in Figure 13 where user throughput is visualized again by means of a CCDF. It can be observed that with different scheduling algorithms, the gains can be shifted between users with different channel qualities. Algorithm S4 achieves high cell throughput gains—gains of more than 100 percent—and the users who benefit most are the 50 percent with the lowest path loss. In contrast, algorithm S3 increases throughput for the 40 percent of users with the highest path loss and the 25 percent of users with the lowest path loss, by achieving gains around 40 percent in cell throughput.

Obviously, besides the four basic scheduling algorithms presented, a wide range of algorithms is imaginable due to the increased scheduling flexibility in MU access, a consequence of the additional dimension of superposition. The range of scheduling possibilities—balancing between fairness and cell throughput while considering user positions and MU operation regions—means finding optimal scheduling is not a trivial task. Therefore, starting with the algorithms presented here, there is still further space for improvement in fairness and/or cell throughput.

Summary and Conclusions

System throughput in cellular wireless systems can be increased significantly even in a single-antenna setup by introducing non-orthogonal



CCDF—Complementary cumulative distribution function
Mbps—Megabits per second
MU—Multi-user
SU—Single-user

Figure 13.
User throughput of orthogonal (SU) access and non-orthogonal (MU) access with scheduling algorithms S3 and S4 (Pedestrian B, 3 km/h).

signaling and non-linear multi-user detection. Even with low complexity methods like superposition of conventionally coded signals and successive interference cancellation with only two simultaneous users, system throughput can be increased by factors in the range of 1.5 to 2 and above, depending on the scheduling algorithm and the fairness constraints. The proposed system has a substantially higher overall throughput than a system using orthogonal signaling with fairness constraints, and improves fairness considerably compared to a sum rate maximizing system without fairness constraints. Delays can be reduced because more bursts can be scheduled in one frame, or a larger number of different users can be served by a single base station.

The principles described here in the context of single-antenna transmission can be combined with multi-antenna setups. As an example for beamforming, one beam could serve two users at the same time on the same radio resource.

Only small modifications to existing OFDMA based standards like IEEE 802.16/WiMAX or 3GPP LTE are required to use the proposed algorithms. Optimum scheduling and the maximum achievable gain are still open problems.

Acknowledgements

The authors would like to thank Michael Tangemann for supporting this work and for coining the term radial space-division multiple access (RDMA), Elmar Schmidberger for contributing to the simulation work, and Gebhard Thierer, Jürgen Otterbach as well as the anonymous reviewers for corrections and improvements to the manuscript.

***Trademark**

3GPP is a trademark of the European Telecommunications Standards Institute.

References

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory, John Wiley & Sons, New York, 1991.

[2] A. Goldsmith, Wireless Communications, Cambridge University, Cambridge, UK, New York, 2005.

[3] Institute of Electrical and Electronics Engineers, "IEEE Standard for Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE P802.16Rev2/D3, Draft, Feb. 2008.

[4] D. Tse and P. Viswanath, Fundamentals of Wireless Communication, Cambridge University, Cambridge, UK, New York, 2005.

(Manuscript approved September 2008)

JOERG SCHAEPPERLE is a research engineer in the Radio Access domain at Alcatel-Lucent Bell Labs in Stuttgart, Germany, and leader of the Advanced WiMAX Radio System research project which is focused on developing innovative solutions for increasing the throughput in future IEEE 802.16/WiMAX radio networks. He received his Diploma degree in electrical engineering with a focus on telecommunication theory from the University of Stuttgart in Germany, where he subsequently worked as a research assistant. During his tenure with Alcatel-Lucent, he has held different positions on research projects for wireline and wireless access to communication networks. Mr. Schaepperle is a member of the Alcatel-Lucent Technical Academy, and a member of the Institute of Electrical and Electronics Engineers (IEEE). He is also a member of the IEEE 802.16 Working Group on Broadband Wireless Access Standards, and participates actively in the standardization of a new air interface in IEEE 802.16 Task Group m. His research interests include communication theory, radio access, and signal processing with a current focus on multi-user communications in OFDMA-based systems.

ANDREAS RÜEGG is a research engineer in the Bell Labs Radio Access domain at Alcatel-Lucent in Stuttgart, Germany. He received an M.S. degree in communication systems from the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland, and a Master of Research in digital communications from the Nice Sophia Antipolis University in Nice, France. His current work focus is on system design aspects and physical layer techniques for next-generation WiMAX systems including contributions to the IEEE 802.16m standardization group. Prior to joining Alcatel-Lucent, Mr. Rüegg worked on ultra wideband (UWB) systems at STMicroelectronics in Geneva, covering simulator design, performance evaluation, and physical layer algorithm development for data rate enhancement and system co-existence. His current interests include multi-user access techniques for OFDMA-based transmission in next-generation communication systems. ◆