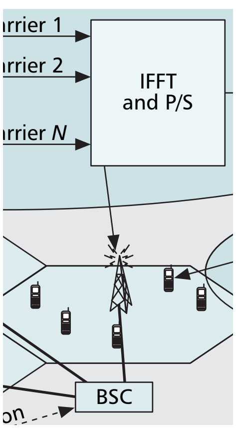
DYNAMIC MULTIUSER RESOURCE ALLOCATION AND ADAPTATION FOR WIRELESS SYSTEMS

KHALED BEN LETAIEF, THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
YING JUN (ANGELA) ZHANG, THE CHINESE UNIVERSITY OF HONG KONG



ABSTRACT

Driven by the increasing popularity of wireless broadband services, future wireless systems will witness a rapid growth of high-data-rate applications with very diverse quality of service requirements. To support such applications under limited radio resources and harsh wireless channel conditions, dynamic resource allocation, which achieves both higher system spectral efficiency and better QoS, has been identified as one of the most promising techniques. In particular, jointly optimizing resource allocation across adjacent and even nonadjacent layers of the protocol stack leads to dramatic improvement in overall system performance. In this article we provide an overview of recent research on dynamic resource allocation, especially for MIMO and OFDM systems. Recent work and open issues on cross-layer resource allocation and adaptation are also discussed. Through this article, we wish to show that dynamic resource allocation will become a key feature in future wireless communications systems as the subscriber population and service demands continue to expand.

To support high-data-rate applications under limited radio resources and harsh wireless channel conditions, dynamic resource allocation, which achieves both higher system spectral efficiency and better QoS, has been identified as one of the most promising techniques.

INTRODUCTION

Over the last decade, mobile and wireless communications have undergone impressive growth worldwide. The forthcoming wireless communication systems are expected to provide a wide variety of new services from high-quality voice to wireless multimedia for anyone, anywhere, anytime, and at the lowest possible cost. In particular, tremendous consumer interest in multimedia applications is fueling the need for very high data rates in future wireless networks. To this end, mobile wireless technologies beyond third generation (3G), which are sometimes labeled next-generation (XG) wireless communications, are desired to satisfy the increasing demands of broadband wireless access. One of the key objectives of XG systems is to support reliable transmissions with high peak data rates ranging from 100 Mb/s for high-mobility applications to 1 Gb/s for low-mobility applications and high spectrum efficiency of up to 10 b/s/Hz. This imposes a major challenge on the design of future wireless communications systems, given the limited radio resources and harsh wireless channel.

In light of the explosive growth of wireless services, radio resources such as frequency spectrum would be far from adequate unless advanced technologies are developed to achieve better efficiency of resource utilization. The traditional approach of statically managing resources results in waste of scarce spectrum and power because an extra margin in the link budget is required to maintain acceptable performance in worst-case fading. Therefore, it is essential to control resource allocation and utilization in a way other than statically to achieve higher spectrum and/or power efficiencies as well as provide better QoS while functioning under bandwidth and power restrictions.

Recently, an attractive multicarrier transmission scheme known as orthogonal frequency-division multiplexing (OFDM) has been identified as one of the leading candidates for supporting broadband and multimedia services in future wireless systems. The scheme is well known for its high spectrum efficiency and robust performance over heavily impaired wireless links. OFDM has already been adopted in HiperLAN, 802.11a, 802.11g, and 802.16 ,and is now considered one of the main air interfaces under consideration for fourth-generation (4G) wireless systems. Meanwhile, recent developments in multiple-input multiple-output (MIMO) techniques have resulted in a significant boost in performance for OFDM systems. Broadband MIMO-OFDM systems with spectrum efficiencies on the order of 10 b/s/Hz are feasible for local/metropolitan area network (LAN/MAN) environments. Besides their robust performance over wireless media, OFDM and MIMO are particularly suitable for adaptive transmission and resource allocation due to the existence of parallel subchannels in the frequency and space domains [1]. This unique feature enables flexible adaptive resource allocation to significantly enhance system capacity and resource utilization.

A key principle of adaptive resource alloca-

This work is supported in part by the Hong Kong Research Grant Council.

1536-1284/06/$20.00 © 2006 IEEE

tion is to exploit inherent system diversities in various domains through intelligent management of bandwidth allocation, multiple access, scheduling, and rate and power adaptation. Inherent system diversities typically result from the time variation and frequency selectivity of wireless channels, independent fading of multiple users, parallel channels in the space and frequency domains, random traffic arrival, user mobility, as well as the interaction between different layers of the network protocol stack. Link adaptation (LA) is a good example of a technique that can be used to exploit the diversity due to wireless channel variations. Being one of the elementary adaptation techniques, LA adapts modulation, coding, transmit power, and/or other signal transmission parameters to the instantaneous channel condition, aiming to increase spectrum efficiency and reliability of wireless systems. In particular, distributing power and information bits according to the water-filling principle maximizes the capacity of a system with parallel Gaussian channels as in OFDM and MIMO. When there are multiple users sharing the same wireless media, independent fading experienced by the users can be exploited to achieve additional capacity gain on top of the link adaptation gain. This is typically referred to as multiuser diversity gain. For example, in OFDM systems subcarriers in deep fade for a given user may not be faded for other users. Dynamically allocating the subcarriers to users according to their channel conditions ensures that each subcarrier is allocated to the user with high channel gains, hence effectively improving spectrum utilization. Recent work on this topic can be found in [1–3].

Most existing adaptive resource allocation algorithms have focused on a single layer of the overall network protocol stack, thereby ignoring the important interdependencies between the layers. Each layer optimizes its own goal, and the design can hardly be optimal from a system point of view. For example, LA algorithms mainly deal with the channel variation in the physical (PHY) layer. The selected transmission parameters may not be optimal as far as the higher-layer QoS requirements are concerned. In contrast, a cross-layer or integrated approach can implement more efficient protocols by exploiting direct coupling between the layers. The nature of cross layer design is to provide an innovative insight into the vertical integration of different protocol layers with the ultimate goal of achieving efficient management of system resources. As a result, various mechanisms that utilize the interconnections among different layers have recently emerged to significantly improve overall system performance [4–6].

The organization of this article is as follows. We begin with a brief review of the potential diversities and dynamics that can be exploited in adaptive resource allocation. We then explore recent advances of dynamic multiuser resource allocation in OFDM and MIMO systems with and without perfect transmitter-side channel state information (CSI). After this, we discuss various challenges and opportunities regarding the concept of cross-layer design and adaptation. Finally, we provide our conclusions.

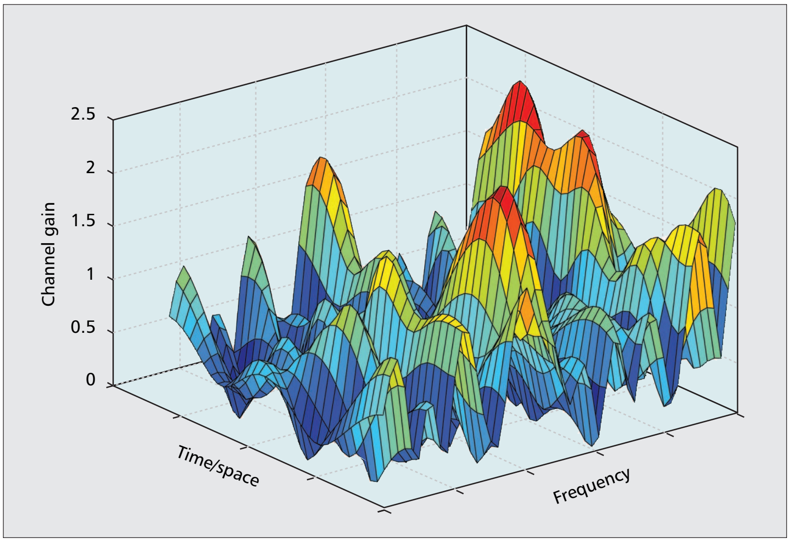


Figure 1. *Time and frequency selective attenuation of wireless channels.*

Dynamics in Wireless Communications Systems

In this section we briefly review the dynamics in wireless communications systems that can be exploited to enhance the utilization of scarce radio resources and optimize system performance.

Channel Variation and Link Adaptation

The variation of wireless channels used to be regarded as a fundamental limitation on the performance of wireless systems, as bit error rate (BER) is typically dominated by the worst-case channel condition. With adaptive resource allocation, however, the time- and frequency-varying characteristics of wireless channels can be exploited and indeed considered an advantage.

A snapshot of a wireless channel is shown in Fig. 1, which illustrates the variation of the channels in different domains. We can see that due to the presence of multiple propagation paths as well as the motion of mobile users and surroundings, a wireless channel varies substantially in time, frequency, and space. The primary goal of LA is to ensure that the most efficient transmission mode is always used regardless of the underlying channel variation. This is achieved by dynamically adjusting the key transmission parameters such as modulation, coding, and transmit power to the changing channel states. The water-filling principle gives a theoretical explanation of this idea. It states that under a certain power constraint, the overall information rate of an arbitrary channel is maximized by transmitting more power where the attenuation and noise are smaller. In other words, a higher transmission rate should be used when the channel is under good condition and vice versa. For example, in case of M-QAM, the number of bits transmitted on subchannel n in an OFDM system can be approximated as

$$\log_2(1 + \gamma_n P_n) \qquad (1)$$

where

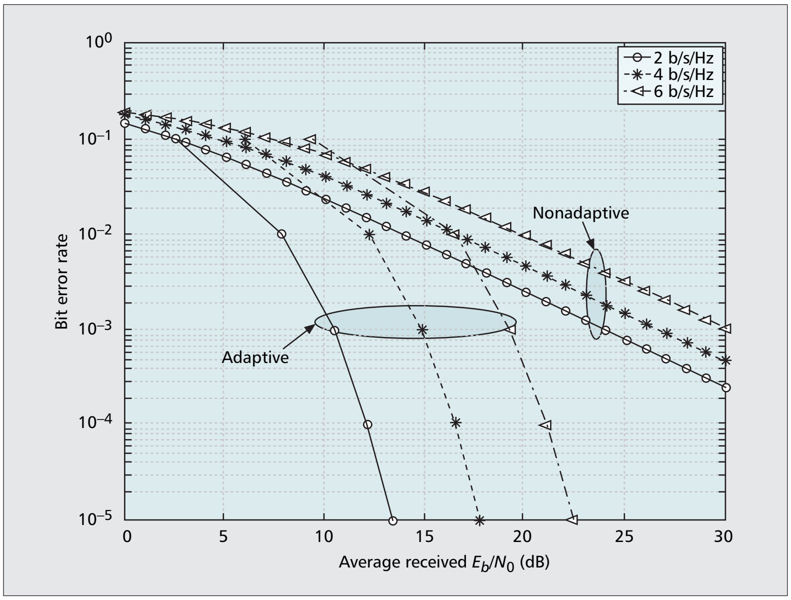
$$\gamma_n = \frac{1.5|h_n|^2}{-\ln(5BER)\sigma^2}$$

denotes the normalized channel-to-noise ratio (CNR) normalized by BER on subchannel n, and h_n denotes the channel coefficient on this subcarrier. Likewise, BER denotes the target BER, and σ^2 is the power of the additive white Gaussian noise. According to the water-filling criterion, the optimal transmit power on subchannel n, denoted P_n, is given by

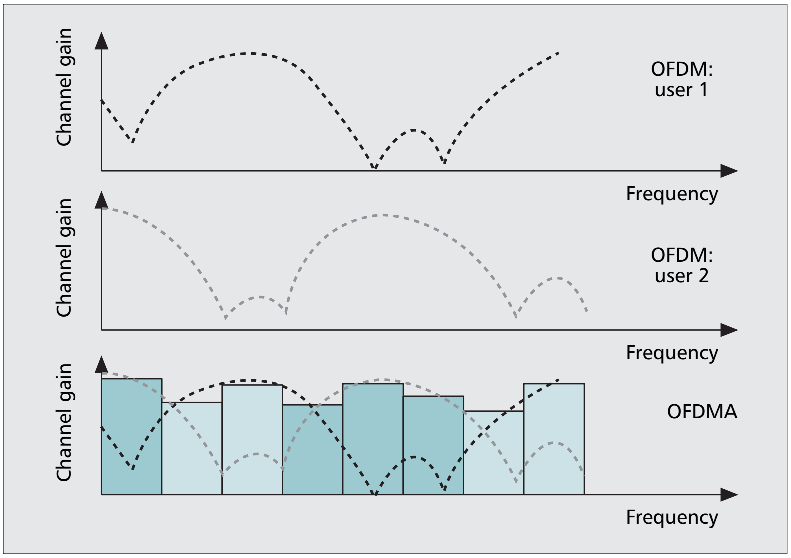
$$P_n = \begin{cases} \lambda - \frac{1}{\gamma_n} & \text{if } \gamma \geq \frac{1}{\gamma_n} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where λ denotes the water level that satisfies the total power constraint

$$\sum_{n=1}^{N} P_n = P_{total}.$$



■ **Figure 2.** *Performance improvement due to link adaptation.*



■ **Figure 3.** *Principle of multi-user diversity and OFDMA. The dark and light slots denote the subchannels allocated to user 1 and user 2, respectively.*

The notion of adaptive modulation in the context of OFDM was proposed as early as 1989 by Kalet [7]. This was further refined for duplex wireless links, for example in [8]. A similar concept is also applicable to MIMO systems where parallel subchannels with different and time-varying channel gains can be created by singular value decomposition (SVD)-based space-time vector coding. In Fig. 2 we demonstrate the significant improvement in system spectrum and power efficiency brought forth by link adaptation. We simulate an OFDM system with 64 subcarriers in a typical urban channel using the Hughes-Hartogs' optimal bit-and-power loading algorithm [9]. It can be seen that while maintaining the same reception quality, resource utilization efficiency is greatly improved. There is a power gain of around 15 dB when BER is 10^{-3}, for example.

The curves plotted in Fig. 2 are for uncoded systems. If channel coding is involved, the gap between adaptive and nonadaptive systems will be narrowed. In essence, if infinite decoding delay is affordable, single-user single-antenna adaptive and nonadaptive systems achieve approximately the same ergodic capacity with similar complexity when the channel sequence is independent and identically distributed (i.i.d.) [10]. In practice, most wireless channels exhibit correlated fading in both the time and frequency domains. In this case, not only does the capacity of nonadaptive systems degrade, but the encoding and decoding complexity to achieve capacity increases dramatically as well. In contrast, adaptive systems yield both higher capacity and lower complexity [10].

Multiuser Diversity

Thanks to the independent fadings of multiple users that share the same wireless link, significant gain in resource utilization efficiency can be achieved by exploiting so-called multiuser diversity. For example, in an OFDM system, different users experience mutually independent attenuations on the subcarriers. In Fig. 3 the dark and light dashed curves denote the channel gains of a two-user system. This figure shows that both users may experience deep fades on a number of subcarriers. If a user is allocated deep-faded subcarriers due to fixed multi-access schemes such as time-division multiple access (TDMA) or frequency-division multiple access (FDMA), these subcarriers will not be efficiently utilized as they can only carry a few information bits. Fortunately, it is quite unlikely for a subcarrier to be in deep fade for all users. If we dynamically allocate the subcarriers according to the instantaneous channel conditions, spectrum efficiency can be effectively improved. This idea, which is often referred to as orthogonal FDMA (OFDMA) or multiuser OFDM, is illustrated in Fig. 3. In this case the positions of the bars indicate the subcarriers that are allocated to the users, while the amplitudes of the bars indicate the data rate that can be transmitted on the subcarriers.

Apart from subcarrier allocation, opportunistic scheduling is another effective method to exploit multiuser diversity [11, 12]. As in subcarrier allocation, the concept of opportunistic scheduling is to give some form of priority to users with (temporarily) better channels. A major problem for multiuser adaptation schemes is how to design practical algorithms that achieve the multiuser diversity gain while supporting diverse QoS requirements under realistic channel scenarios.

INTERACTION BETWEEN THE PHY LAYER AND HIGHER LAYERS

Besides PHY-layer channel variation, the time-varying nature of higher-layer user activities also leads to dynamics that can be taken advantage of by dynamic resource allocation. Such dynamics include user mobility, random packet arrival, time-varying topology of wireless ad hoc networks, and so on. Unlike wired networks, the dynamics at multiple layers of wireless networks are correlated. For example, the physical channel variation could result in a time-varying topology of ad hoc networks. Likewise, the routing cost on each link is affected by the underlying channel condition. Besides, QoS provisioning at each layer is also interrelated. For instance, exploiting multiuser diversity results in an increase in total system capacity. However, this must be balanced with network layer issues such as fairness and delay. All these coupling effects motivate us to consider the higher-layer QoS issues such as throughput, delay, and fairness jointly with PHY-layer issues such as channel fading and modulation. This demonstrates the need for cross-layer dynamic resource allocation, which we discuss in more detail later.

MULTIUSER RESOURCE ALLOCATION IN SISO-OFDM SYSTEMS

Following a variety of valuable contributions to adaptive rate-and-power transmission, Wong *et al.* were the first to consider dynamic resource allocation to achieve multiuser diversity gain in single-input single-output (SISO) OFDM systems [2]. In this work subcarrier, bit, and power allocation are jointly optimized to minimize the total transmit power under the constraints of each user's minimal data rate. This work was followed by numerous research investigations of how to appropriately formulate the optimization problem for various communications scenarios, how to balance the overall system performance and QoS constraints, and how to implement the joint optimization algorithms in real time. In general, two major objective functions have been considered in the literature:

- Maximizing data rate for a given power budget and a target BER (referred to as the bit rate maximization problem) [3]
- Minimizing total transmit power for a given data rate and BER (referred to as the power margin maximization problem) [2]

Before introducing the formulation of the optimization problems, we introduce some notations. Assume that there are K users and N subcarriers in an OFDM system. The CSI at the base station (BS) is typically obtained either by estimating it at the mobile stations and sending it to the BS via a feedback path or through channel estimation of the uplink in a time-division duplex (TDD) system. Using this information, the BS assigns a set of subcarriers to each user and determines the number of bits and amount of power to be transmitted on each subcarrier. Let $r_{k,n}$ and $p_{k,n}$ denote the number of bits and amount of power transmitted on subcarrier n if the subcarrier is assigned to user k. Given a fixed BER, $r_{k,n}$ and $p_{k,n}$ are related as follows:

$$p_{k,n} = \frac{f_k(r_{k,n})}{|h_{k,n}|^2} \text{ for all } k,n, \tag{3}$$

where $f_k(r_{k,n})$ is the required received power in the nth subcarrier for reliable reception of $r_{k,n}$ bits/symbol when the channel gain is equal to unity, and $h_{k,n}$ is the channel coefficient of user k on subcarrier n. Likewise, let $c_{k,n}$ be the subcarrier allocation indicator. $c_{k,n} = 1$ implies that subcarrier n is assigned to user k. Otherwise, $c_{k,n} = 0$. To avoid severe co-channel interference (CCI), a subcarrier can be occupied by at most one user. That is,

$$\sum_{k=1}^{k} c_{k,n} = 1 \text{ for all } n. \tag{4}$$

In the case of maximizing the data rate, the problem is formulated as follows:

$$\max_{p_{k,n},c_{k,n},r_{k,n}} \sum_{n=1}^{N}\sum_{k=1}^{K} c_{k,n} r_{k,n} \tag{5.1}$$

subject to

$$\sum_{n=1}^{N}\sum_{k=1}^{K} c_{k,n} p_{k,n} \le P_{total}, \tag{5.2}$$

$$\sum_{n=1}^{N} c_{k,n} = 1 \text{ for all } n, \tag{5.3}$$

$$c_{k,n} \in \{0,1\}, \tag{5.4}$$

$$p_{k,n} \ge 0 \text{ and } r_{k,n} \ge 0, \text{ integer}, \tag{5.5}$$

where Eq. 5.1 corresponds to the objective of maximizing the total data rate, and Eq. 5.2 imposes a limit on the total transmit power.

For the power margin maximizing problem, the mathematical formulation is given as follows:

$$\min_{r_{k,n},p_{k,n},c_{k,n}} \sum_{n=1}^{N}\sum_{k=1}^{K} c_{k,n} p_{k,n} \tag{6.1}$$

subject to

$$\sum_{n=1}^{N} c_{k,n} r_{k,n} \ge R_k, \tag{6.2}$$

$$\sum_{k=1}^{K} c_{k,n} = 1 \text{ for all } n, \tag{6.3}$$

$$c_{k,n} \in \{0,1\}, \tag{6.4}$$

$$p_{k,n} \ge 0 \text{ and } r_{k,n} \ge 0, \text{ integer}, \tag{6.5}$$

where Eq. 6.1 corresponds to the objective of minimizing the total power consumption and Eq. 6.2 represents the minimum data rate requirement, denoted by R_k, for each user.

The above formulated problems are typically

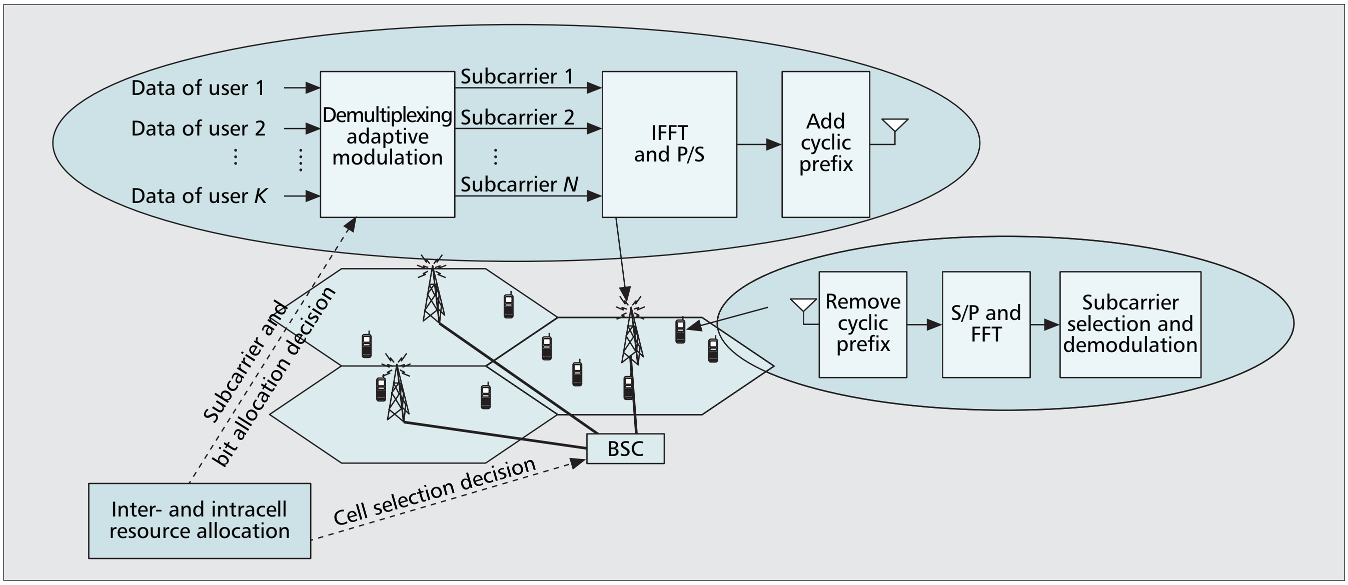


Figure 4. *System diagram of inter- and intracell resource allocation and optimization.*

nonlinear since $f_k(r_{k,n})$ is nonlinear. Moreover, the integer constraints on $c_{k,n}$ and $r_{k,n}$ makes the optimization complexity grows exponentially with the number of variables. As a result, the computational complexity of Wong's algorithm in [2] is prohibitively high. To reduce the complexity, a variety of practical algorithms, including sorting-search subcarrier assignment, greedy bit loading and power allocation, as well as objective aggregation, are proposed in [5]. In [13], the authors considered a weighted rate-sum maximization problem instead of imposing constraints on individual rate requirements explicitly. By allowing subcarrier sharing among multiple users, the optimization problem is converted to a convex one and is solved using duality methods. Though the above algorithms lead to suboptimal resource allocation, computational complexity is greatly decreased compared to the original problem.

In [3] Zhang *et al.* proposed a multiuser resource allocation algorithm that involves both intracell and intercell optimization. The system diagram is shown in Fig. 4. For the intracell level, their main contribution is to introduce a reduced-complexity multiuser subcarrier bit and power allocation scheme. Specifically, the subcarriers are first allocated to maximize the total data rate without considering the individual user's data rate constraints. Afterward, subcarrier allocation is adjusted step by step to satisfy the individual data rate constraints. As to the intercell level, the major contribution is to design a cell selection scheme to exploit the intercell diversity and reduce the outage probability, which is inevitable when there is a transmit power limit at the BS. Unlike conventional cell selection schemes, the adaptive cell selection algorithm proposed in [3] is particularly designed for adaptive physical-layer transmission, and is based on not only the traffic density, but also the QoS requirements and received power levels of all users. In particular, the cell is assigned based on the following factors:

- Resource availability in the candidate cells
- Amount of resources that have already been assigned to other users
- Amount of resources necessary to satisfy the investigated user's QoS

This algorithm effectively solves the outage problem not addressed in previous adaptive OFDM contributions. Moreover, the proposed algorithm can considerably mitigate the destructive effect of nonuniform traffic density, which is a conspicuous issue in wireless LAN or hotspots.

Dynamic Resource Allocation in MIMO-OFDM Systems

Multiuser MIMO-OFDM systems benefit from the combined frequency and space domain freedom as well as multiuser diversity. Moreover, allowing intracell bandwidth reuse by means of space-division multiple access (SDMA) is an advantage of multi-antenna systems over other systems. It greatly enhances the spectrum efficiency if the bandwidth is shared by spatially separable users. However, resource allocation in MIMO-OFDM systems is much more challenging than that in SISO-OFDM systems due to the following reasons. First of all, CCI caused by subcarrier reuse makes the optimization problem combinatorial and non-convex. Adapting the transmission of one user affects the interference to other co-channel users, which in turn changes the optimal transmission schemes of all users. Second, the achievable signal-to-interference ratio (SIR) is a function of the set of users that share the subcarrier. To maximize spectrum efficiency while achieving a sufficient SIR, optimal sets of co-channel users should be identified for every subcarrier based on their spatial correlations and power distributions. Third, MIMO-OFDM systems are able to multiplex the users in both the space and frequency domains. As a result, we have to decide which dimension should be occupied by which set of users. Finally, QoS requirements impose additional constraints on the optimization problem.

Multiuser adaptive transmission in multiple antenna systems has recently been reported in [14, 15] to exploit space and multiuser diversity from an information theory point of view. In particular, [14] examined the optimal power allocation strategy that maximizes the sum ergodic capacity for a non-asymptotic case when there are finite transmit and receive antennas. This previous work shows that in a system with n transmit antennas for each user and m receive antennas, the optimal power control strategy allows up to $1/2m(m + 1)$ degrees of freedom to be used at any time, with each user contributing up to $1/2n(n + 1)$ degrees of freedom. In [15] the authors studied the joint problem of intracell resource allocation and transmit beamforming in the context of an OFDM-SDMA system. The proposed algorithm allocates spatially separable users to the same subcarriers, while appropriately adjusting the beam pattern of individual users at the transmitter. The objective is to maximize the receiver signal-to-interference-plus-noise ratio (SINR), increase resource utilization, guarantee user separability, and ultimately enhance system capacity. Reference [15] also considered a constrained resource allocation problem, in which each user has a minimum data rate requirement. In this case subcarrier reuse is not allowed to simplify the problem. By doing so, the bandwidth reuse capability of MIMO systems is not exploited, and hence the spectrum resource is not fully utilized.

In [14, 15] instantaneous QoS provisioning is not guaranteed when the spectrum reuse capability of MIMO is utilized. In other words, although the data rate sum is maximized, some users' data rate might be very low for certain time intervals due to poor channel conditions. In [1] Zhang *et al.* proposed multiuser resource allocation algorithms with QoS provisioning for MIMO-OFDM systems. In particular, a user-grouping-based resource allocation algorithm is proposed in [1]. The idea is to group users according to their spatial separability, which is quantified by the correlation between spatial signatures. For example, in Fig. 5 mobile station (MS) 1 and MS 2 are most likely to be highly correlated in the space domain as their direction of arrivals (DOAs) at the BS are close.[1] Hence, they are classified in one group. Likewise, MS 3 and MS 4 are in one group. MS 5 and MS 6 are spatially separable from any other stations, and hence they are not grouped with the other mobile units. The proposed scheme only allows the users that are sufficiently separated in the space domain (and hence in different groups) to share the same subcarriers. By doing so, CCI due to frequency reuse is negligible at the receiver, given that an appropriate multiuser detector is used at the BS. As a result, the intractability of the original optimization problem due to the existence of CCI is broken and the computational complexity of the algorithm is greatly reduced.

In Fig. 6 the performance improvement of the proposed system with a parallel interference cancellation (PIC) detector is demonstrated. Assume that there are 64 subcarriers and two users in the system, with each user having a rate requirement of 128 b/OFDM symbol interval. Likewise, assume that each mobile unit is equipped with two antennas, and the BS is equipped with four antennas. For comparison, the performance of the conventional nonadaptive SDMA system (i.e., conventional SDMA system without adaptive subcarrier

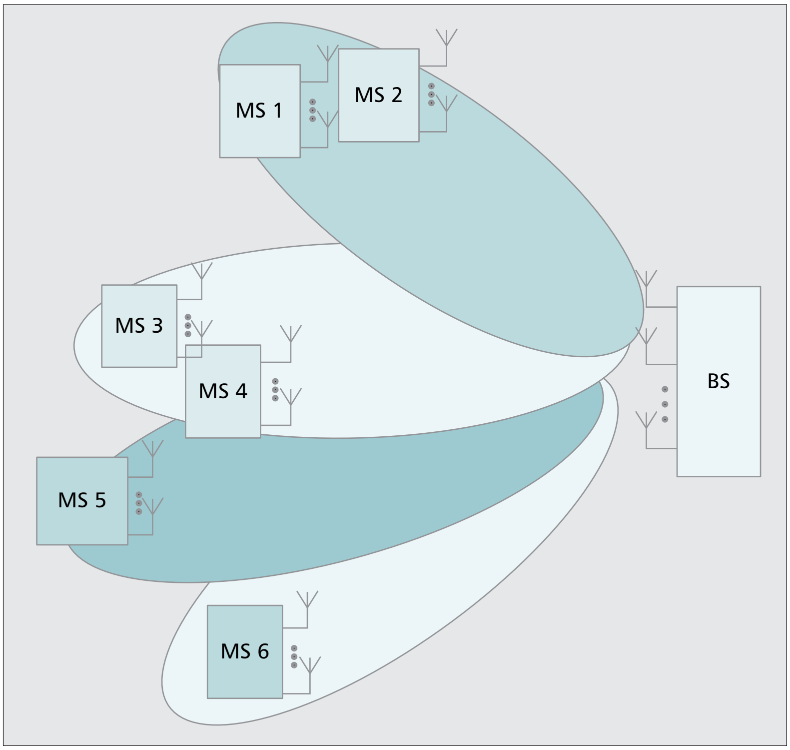


Figure 5. *User grouping in the MIMO-OFDM multiuser resource allocation algorithm proposed in [1].*

power and rate adaptation) is also plotted. The figure shows that the proposed adaptive resource allocation algorithm is able to provide tremendous diversity gain and power gain compared to the nonadaptive system due to the successful exploitation of the system diversity in the frequency, space, and user domains. There is more than 8 dB gain when BER is 10^{-2}, for instance. The performance is also compared to the single user bound, in which there is no CCI. It can be seen that the diversity order is almost the same as that of a single-user adaptive MIMO-OFDM system where single-user rate and power adaptation is adopted. In particular, note that the power gap is less than 1 dB. This implies that the proposed algorithm is able to multiplex users in the space domain without sacrificing antenna diversity.

Dynamic MIMO-OFDM Resource Allocation with Imperfect Channel State Information

A critical element in all of the dynamic resource allocation schemes is the CSI knowledge. Most adaptation schemes assume perfect CSI at the transmitter. However, perfect CSI is rarely available in real systems. The imperfection may result from feedback delay, channel estimation or prediction errors, and quantized feedback. Very recently, the problem of how to use practical CSI to optimize the resource allocation in MIMO-OFDM systems has been considered (e.g., [6, 16, 17].)

In [16] Xia *et al.* assume that only channel mean feedback is available at the transmitter. Assuming that the transmitter obtains an unbiased channel estimate $\widetilde{\mathbf{H}}$. The true MIMO channel can then be modeled as

[1] *This applies to an outdoor environment where the BS is located higher than the surrounding scatterers, and the user signals arrive at the BS from a single DOA with a small angular spread. When the angular spread is large, users are grouped by directly calculating the correlation between their spatial signatures.*

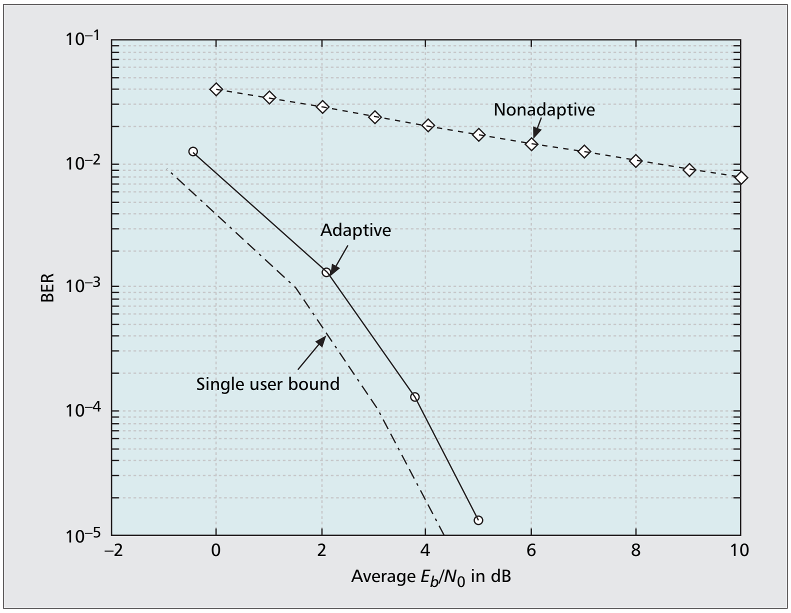


Figure 6. *Performance comparison between the proposed and nonadaptive systems with PIC detector.*

$$\mathbf{H} = \tilde{\mathbf{H}} + \hat{\mathbf{I}} \quad (7)$$

where $\hat{\mathbf{I}}$ is a random matrix Gaussian distributed according to $CN\,(0_{N_t \times N_{rt}}, N_r\sigma_\varepsilon^2 \mathbf{I}_{Nt})$. The variance σ_ε^2 encapsulates the CSI reliability. At the transmitter, a 2D coder-beamformer, which combines space-time block coding and eigen-beamforming, is adopted. Adaptation will then take place in three levels at the transmitter:

- Rate and power adaptation of information symbols
- Power splitting among space-time coded information symbol substreams
- The basis beams of multidimension beamformers that are used to steer the transmission over the MIMO subchannels corresponding to each subcarrier

The proposed algorithm maximizes the transmission rate for a fixed transmit power and prescribed BER performance.

In contrast to [16], which deals with a single user scenario, Huang *et al.* proposed an adaptive resource allocation scheme for multiuser MIMO-OFDM systems by utilizing channel mean feedback [6]. Similar to [16], a 2D beamformer is used at the base station. Subcarrier assignment, beamforming vector adjustment, and power allocation over both subcarriers and basis beamforming vectors are then jointly optimized to maximize the system spectrum efficiency. In addition, packet scheduling is adapted to the PHY-layer resource allocation results so as to better utilize the channel variation in various domains.

Quantized feedback is another class of CSI model that imposes a bandwidth constraint on the feedback channel. Single-user adaptive modulation and transmit beamforming design based on quantized feedback was studied in [17], for instance, under the assumption that channels are flat faded. Specifically, the transmit power, signal constellation, beamforming direction, and feedback strategy are jointly optimized. Due to the difficulty of solving the joint optimization problem, a suboptimal nested iterative approach that iteratively adjusts the transmission modes and feedback strategies was then proposed. Although [17] considers a system with multiple transmit antennas and a single receive antenna, the extension to a MIMO system is straightforward. However, more efforts are required to extend the work to a multiuser and multicarrier system.

So far, there is little published work on MIMO-OFDM resource allocation with partial CSI. How to optimally allocate resources without the exclusive subcarrier allocation constraint is still an open problem. The topic is of significant interest to enhancing resource utilization of practical systems that adopt MIMO and OFDM, such as 802.11n, and hence is worthy of further investigation.

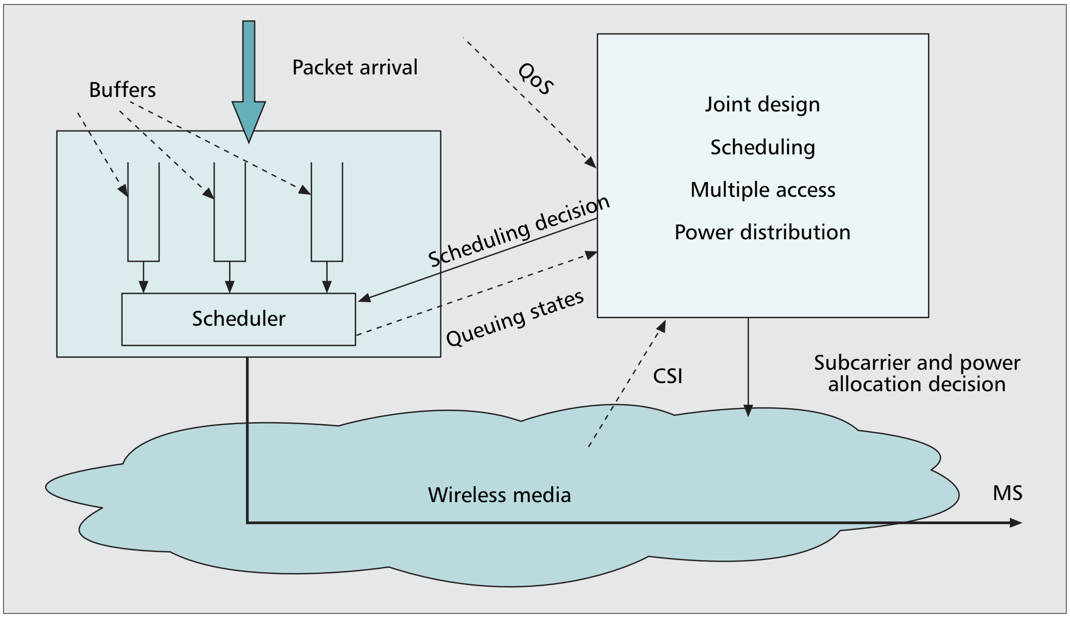
Cross MAC-PHY Layer Resource Allocation

Adaptive cross-layer resource allocation that exploits interdependencies and interactions across the PHY, MAC, and higher layers has recently attracted extensive research interests [4]. The basic idea of cross-layer resource allocation is to adapt bandwidth and power allocation as well as transmission strategies jointly across the protocol stack in order to significantly improve the resource utilization efficiency and guarantee a predetermined QoS at the receiver.

One example of the cross MAC-PHY layer resource allocation is opportunistic scheduling, such as that proposed in [11, 12]. Opportunistic scheduling endeavors to maximize the system capacity by scheduling one or more users with the best instantaneous channels in each time slot. In [12] Viswanath *et al.* proposed the use of multiple transmit antennas to induce large and fast channel fluctuations to increase multiuser diversity in the system. To ensure proportional fairness, users' transmissions are scheduled at times when the channel SINR is near its own peak. In [11] the resource sharing constraint in terms of the fraction of time assigned to each user is explicitly considered to strike a balance between efficiency and fairness. The authors illustrate via simulation that opportunistic scheduling can significantly outperform the channel-independent scheduling policies such as round-robin.

The above mentioned work assumes that there are always sufficient data waiting to be transmitted. In particular, dynamic queuing behavior in the MAC layer is not considered. This assumption was not made in recent work in [18, 19], in which the dynamic features observed in the MAC layer such as random traffic arrival and finite buffer size are considered. In [18] the authors developed an analytical procedure to study the performance for transmissions over wireless links, where finite-length queuing at the link layer is coupled with adaptive modulation at the PHY layer. A cross-layer scheme was proposed to minimize the packet loss rate and maximize the average throughput. This work, however, was focused on a single-user scenario.

In [19] Zhang and Letaief proposed a joint MAC-PHY layer resource management algorithm for wireless OFDM networks with random packet arrivals. The proposed resource allocation



■ **Figure 7.** *Cross MAC-PHY layer resource allocation.*

The joint MAC-PHY optimization is based on the users' QoS requirements, queuing states observed in the MAC layer, and CSI observed in the PHY layer. The resource allocation decisions are fed back to the two layers to adjust the scheduling and transmission parameters.

scheme involves the joint optimization of three parts: packet scheduling, subcarrier allocation, and power control. The idea is illustrated in Fig. 7. Assume that there are K users and N subcarriers, and the time axis is divided into frames. On arriving at the BS, the packets from different users are buffered in separate queues, which are assumed to have infinite lengths. Within one queue, packets are served in a first-in first-out (FIFO) order. Across the queues, packets are served according to a proposed resource allocation discipline. The joint MAC-PHY optimization is based on the users' QoS requirements, queuing states observed in the MAC layer, and CSI observed in the PHY layer. The resource allocation decisions are fed back to the two layers to adjust the scheduling and transmission parameters. The basic idea behind this algorithm is to serve the packets in a way that approximates a fair queuing system in wired networks, which is referred to as *reference system*. Leads or lags of the flows are introduced in a controlled manner to explore the time domain diversity on top of the diversities in the frequency and multiuser domains. The authors proved through analysis that the proposed algorithm is able to guarantee QoS in terms of throughput and maximum delay as well as fairness to users over wireless media.

Cross-Layer Adaptation, Design, and Optimization

In the previous section, we have introduced cross MAC-PHY adaptive resource allocation for broadband OFDM systems. In this section, we will further provide a state-of-the-art review of cross-layer adaptation and design as well as discuss the challenges and open issues in this emerging research field.

To satisfy different QoS requirements in future wireless communication systems, a multitude of advanced techniques have been studied in each layer of the network protocol stack. For example, multiple antennas, coding, adaptive modulation, and power control at the PHY layer, scheduling and aggressive multiple access at the MAC layer, delay-constrained and energy-constrained routing and mobility management in the network layer, and adaptive QoS in the application layer. These research efforts have mainly focused on isolated components of the overall network design. However, tasks would be too complex for a single layer to support the heterogeneous traffic and stringent QoS requirements. As a result, a cross-layer design that supports adaptivity and optimization across multiple layers of the protocol stack is needed. This represents a paradigm shift in the design of wireless networks where the protocols at each layer are not developed in isolation, but rather within an integrated and hierarchical framework to take advantage of the interdependencies between different layers. In particular, the available resources are dynamically optimized in response to the variation in the link quality and QoS requirements through an integrated design framework where only necessary information is exchanged between the different layers and selected parameters are jointly optimized.

Being new and vibrant, the cross-layer approach, as shown in Fig. 8, has already shown various optimization opportunities through the co-design of the different protocol layers. For example, in wireless networks, packets may be lost not only due to collisions, but also because of channel fading. Traditional MAC protocols, which are designed independent of the PHY layer, typically regard each packet loss as a collision. Whenever a mobile senses a packet loss, it starts a backoff process to avoid successive collisions. As a result, the traditional MAC protocols end up being over-conservative when packet losses can be caused by fading. As such, it is clear that there are potential benefits to be obtained by designing MAC operations with feedback from the PHY layer, which indicates loss probability due to channel fading. Addition-

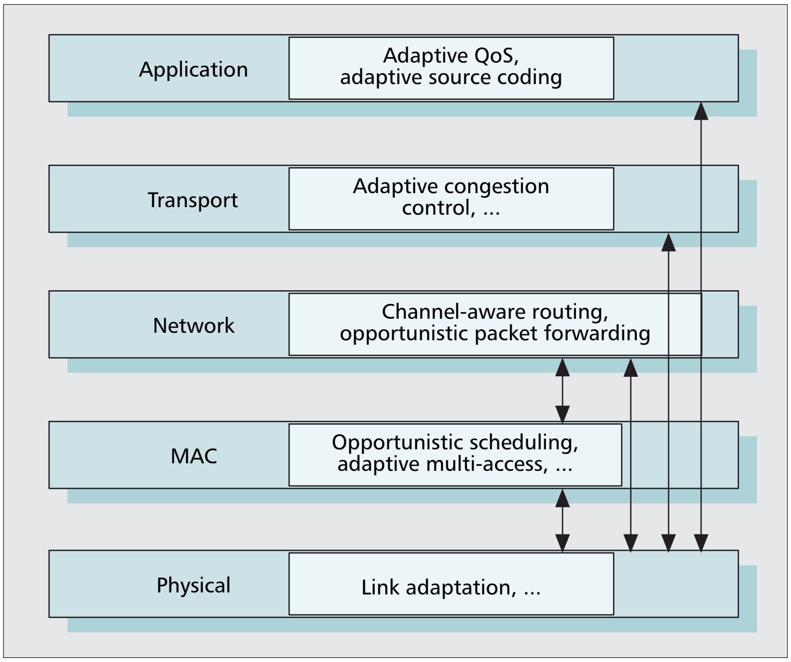


Figure 8. *Cross-layer design, adaptation, and optimization.*

al performance improvement can also be obtained by leveraging the use of the latest performance enhancing tools in cross-layer design. For instance, multiuser detection is a powerful PHY-layer scheme that allows a wireless receiver to decode multiple packets simultaneously. Such a scheme, if used, will have a great impact on the performance of the MAC protocol in wireless networks. This is because multiple packets can now be reliably and correctly detected at the same time, and as a result, conventional MAC protocols that endeavor to avoid simultaneous packet transmissions in order to minimize collisions may not be efficient in this scenario.

Wireless links not only affect the design of MAC protocols, but also have a large impact on the network, transport, and application layers. For example, the connectivity and topology of a network can change with the PHY-layer link adaptation and power control. So does the routing cost on each link. In [20] Chiang proposed a distributed algorithm for joint optimal end-to-end congestion control and per-link power control. The algorithm utilizes the coupling between the transport and physical layers to increase the end-to-end throughput and energy efficiency in a wireless multihop network. The overall communication network is modeled by a generalized network utility maximization problem, with each layer corresponding to a decomposed subproblem. In [21] Cui *et al.* proposed energy-efficient joint routing, scheduling, and link adaptation strategies that maximize the life time of sensor networks. They modeled the cross network-MAC-PHY layer problem into a convex optimization problem, which leads to significant energy saving compared to traditional MAC and routing schemes that are based on isolated layered approaches. The interaction between the application and PHY layers can also be exploited. Cross application-MAC-PHY layer approaches were recently studied to enhance multimedia services, including video streaming and multimedia conferencing. For instance, [22] proposed to jointly design the application-layer forward error correction (FEC), maximum MAC retransmission limit, and packet size to realize robust scalable video transmission over 802.11 WLANs. Likewise, in [23] the selection of source coding parameters is jointly considered with the PHY power and rate adaptation. The scheme maximizes the battery life of mobile devices while guaranteeing the fulfillment of delay and quality constraints of video transmissions over unreliable wireless links.

Clearly, cross-layer design and adaptation represents a new and exciting methodology, but also poses significant technical challenges with numerous open problems that are broad and deep. These open problems are broadly classified into two categories: theoretical modeling and algorithm design. As to theoretical modeling, critical issues include: how to properly take advantage of the interdependencies among the multiple layers; which set of parameters should be exchanged across the multiple protocol layers; which layers should be jointly optimized and designed; and how to deal with network complexity in terms of analysis, performance limits, and protocol design. Algorithm design also poses many challenges. For example, one major issue is that most cross-layer schemes require centralized control and optimization. Although good for cellular systems, centralized optimization may be infeasible for distributed wireless networks such as ad hoc networks and 802.11 WLANs with a distributed coordination function (DCF) mode, due to the lack of a central coordinator. It is therefore essential to develop decentralized cross-layer design and optimization algorithms for practical implementation.

Another issue relates to the enormous overhead due to frequent control signal transmission and CSI feedback, once the PHY channel is considered in the design of higher-layer protocols. This is because the wireless links vary on a timescale that is much faster than the variation in higher-layer behaviors, such as traffic load and network topology. In particular, the shadowing effect varies on the order of seconds, and the multipath fading typically varies on the order of milliseconds to tens of milliseconds. This overhead may significantly cancel out the spectrum efficiency enhancement brought by cross-layer optimization. To solve this problem, it is necessary to develop resource allocation algorithms that require minimum CSI and control signaling.

Finally, computational complexity is yet another critical issue. Typically, the cross-layer design is formulated into an optimization problem, which is usually combinatorial and nonconvex, due to the QoS constraints, the interference between different users, and the interaction between the protocol layers. Prohibitively high computational complexity is therefore typically needed to solve such problems. In particular, the optimal system design solution may need to be updated once the wireless channel changes, and this may make the complexity unaffordable to real-world systems. As a result, it is important to build up a simple model for cross-layer optimization and propose reduced-complexity algorithms.

Conclusions

Dynamic resource allocation considerably improves resource utilization efficiency by exploiting multiuser diversity gain as well as system dynamics in various domains. As the subscriber population and service demand continue to expand, the advantages of dynamic resource allocation will be increasingly important in future broadband and ubiquitous wireless communications systems. This article has attempted to give an overview of the recent advances on dynamic multiuser resource allocation. In particular, the latest work on resource allocation for SISO-OFDM, MIMO-OFDM and joint MAC-PHY design was introduced. We also discussed the state-of-art research on cross-layer design, adaptation, and optimization. This overview is not meant to be exhaustive, but rather to show the significance of dynamic resource allocation in the next-generation wireless systems.

References

[1] Y. J. Zhang and K. B. Letaief, "An Efficient Resource Allocation Scheme for Spatial Multiuser Access in MIMO/OFDM Systems," *IEEE Trans. Commun.*, vol. 53, no. 1, Jan. 2005, pp. 107–16.

[2] C. Y. Wong *et al.*, "Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation," *IEEE JSAC*, vol. 17, no. 10, Oct. 1999, pp. 1747–58.

[3] Y. J. Zhang and K. B. Letaief, "Multiuser Adaptive Subcarrier-and-Bit Allocation with Adaptive Cell Selection for OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, Sept. 2004, pp. 1566–75.

[4] A. J. Goldsmith and S. B. Wicker, "Design Challenges for Energy-Constrained Ad Hoc Wireless Networks," *IEEE Wireless Commun. Mag.*, vol. 9, no. 4, Aug. 2002, pp. 8–27.

[5] G. Song and Y. Li, "Cross-Layer Optimization for OFDM Wireless Networks — Part II: Algorithm Development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, Mar. 2005, pp. 625–34.

[6] W. Huang and K. B. Letaief, "A Cross-Layer Resource Allocation and Scheduling for Multiuser Space-Time Block Coded MIMO/OFDM Systems," *Proc. IEEE ICC '05*, vol. 4, May 2005, pp. 2655–59.

[7] I. Kalet, "The Multitone Channel," *IEEE Trans. Commun.*, vol. 37, Feb. 1989, pp. 119–24.

[8] T. Keller and L. Hanzo, "Adaptive Modulation Techniques for Duplex OFDM Transmission," *IEEE Trans. Vehic. Tech.*, vol. 49, Sep. 2000, pp. 1893–1906.

[9] D. Hughes-Hartogs, "Ensemble Modem Structure for Imperfect Transmission Media," U.S. Patents nos. 4 679 227, 4 731 816, 4 833 706, July 1987.

[10] A. J. Goldsmith and P. P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Trans. Info. Theory*, vol. 43, no. 6, Nov. 1997, pp. 1986–92.

[11] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic Transmission Scheduling with Resource-Sharing Constraints in Wireless Networks," *IEEE JSAC*, vol. 19, no. 10, 2001, pp. 2053–64.

[12] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Trans. Info. Theory*, vol. 48, June 2002, pp. 1277–94.

[13] J. Huang *et al.*, "Downlink Scheduling and Resource Allocation for OFDM Systems," CISS '06, Princeton, NJ, Mar. 2006.

[14] W. Yu, W. Rhee, and J. M. Cioffi, "Optimal Power Control in Multiple Access Fading Channels With Multiple Antennas," *IEEE Proc. ICC '01*, vol. 2, 2001, pp. 575–79.

[15] I. Koutsopoulos and L. Tassiulas, "Adaptive Resource Allocation in SDMA-Based Wireless Broadband Networks with OFDM Signaling," *IEEE Proc. INFOCOM '02*, vol. 3, 2002, pp. 1376–85.

[16] P. Xia, S. Zhou, and G. B. Giannakis, "Adaptive MIMO-OFDM Based on Partial Channel State Information," *IEEE Trans. Sig. Processing*, vol. 52, no. 1, Jan 2004, pp. 202–13.

[17] P. Xia, S. Zhou, and G. B. Giannakis, "Multiantenna Adaptive Modulation with Beamforming based on Bandwidth-Constrained Feedback," *IEEE Trans. Commun.*, vol. 53, no. 3, Mar. 2005, pp. 526–36.

[18] Q. Liu, S. Zhou, and G. B. Giannakis, "Queuing with Adaptive Modulation and Coding over Wireless Links: Cross-Layer Analysis and Design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, May 2005, pp. 1142–53.

[19] Y. J. Zhang and K. B. Letaief, "Cross-layer Adaptive Resource Management for Wireless Packet Networks with OFDM Signal," to appear in *IEEE Trans. Wireless Commun.*; www.ie.cuhk.edu.hk/~yjzhang/Research.htm

[20] M. Chiang, "Balancing Transport and Physical Layers In Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control," *IEEE JSAC*, vol. 23, no. 1, Jan. 2005, pp. 104–16.

[21] S. Cui *et al.*, "Joint routing, MAC, and Link Layer Optimization in Sensor Networks with Energy Constraints," *IEEE Proc. ICC '05*, vol. 2, May 2005, pp. 725–29.

[22] M. V. D. Schaar *et al.*, "Adaptive Cross-Layer Protection Strategies for Robust Scalable Video Transmission over 802.11 WLANs," *IEEE JSAC*, vol. 21, no. 10, Dec. 2003.

[23] C. E. Luna *et al.*, "Joint Source Coding and Rate Adaptation for Energy Efficient Wireless Video Streaming," *IEEE JSAC*, vol. 21, no. 10, Dec. 2003, pp. 1710–20.

Biographies

Khaled Ben Letaief [F] (eekhaled@ee.ust.hk) from 1990 to 1993 was a faculty member at the University of Melbourne, Australia. Since 1993 he has been with the Hong Kong University of Science and Technology where he is currently chair professor and head of the Electronic and Computer Engineering Department. He is also the Director of the Hong Kong Telecom Institute of Information Technology as well as the Director of the Center for Wireless Information Technology. His current research interests include wireless and mobile networks, broadband wireless access, OFDM, CDMA, and beyond 3G systems. In these areas he has published over 280 journal and conference papers and given invited talks as well as courses all over the world. He served as a consultant for different organizations and is currently the founding Editor-in-Chief of *IEEE Transactions on Wireless Communications*. He has served on the editorial board of other prestigious journals including *IEEE Journal on Selected Areas in Communications — Wireless Series* (as Editor-in-Chief) and *IEEE Transactions on Communications*. He has been involved in organizing a number of major international conferences and events. These include serving as Technical Program Chair of the 1998 IEEE Globecom Mini-Conference on Communications Theory, Sydney, Australia, as well as Co-Chair of the 2001 IEEE ICC Communications Theory Symposium, Helsinki, Finland. In 2004, he served as Co-Chair of the IEEE Wireless Communications, Networks and Systems Symposium, Dallas, Texas, as well as the Co-Technical Program Chair of the 2004 IEEE International Conference on Communications, Circuits and Systems, Chengdu, China. He was Co-Chair of the 2006 IEEE Wireless Ad Hoc and Sensor Networks Symposium, Istanbul, Turkey, as well as General Chair of the 2007 IEEE Wireless Communications and Networking Conference (WCNC '07), Hong Kong. In addition to his active research and professional activities, he has been a dedicated teacher committed to excellence in teaching and scholarship. He received the Mangoon Teaching Award from Purdue University in 1990; the Teaching Excellence Appreciation Award by the School of Engineering at HKUST (four times); and the Michael G. Gale Medal for Distinguished Teaching. He is an elected member of the IEEE Communications Society Board of Governors and an IEEE Distinguished Lecturer of the IEEE Communications Society. He also served as Chair of the IEEE Communications Society Technical Committee on Personal Communications as well as a member of the IEEE ComSoc Technical Activity Council.

Yin Jun (Angela) Zhang (yjzhang@ie.cuhk.edu.hk) [S'02, M'05] received her B.S. degree (with honors) in electronics engineering from Fudan University, Shanghai, China, in 2000, and a Ph.D. degree in electrical and electronics engineering from the Hong Kong University of Science and Technology (HKUST) in 2004. From 2000 to 2004 she was a research assistant in the Center for Wireless Information Technology at HKUST, where she worked on adaptive resource allocation for wireless OFDM networks. Since January 2005 she has been with the Chinese University of Hong Kong, where she is currently an assistant professor of theDepartment of Information Engineering. Her research interests include broadband multicarrier/OFDM techniques, adaptive coding and modulation, optimal resource allocation for wireless networks, cross-layer design and optimization, MIMO antenna techniques, and wireless LAN/PAN.

As the subscriber population and service demand continue to expand, the advantages of dynamic resource allocation will be increasingly important in future broadband and ubiquitous wireless communications systems.