Resource Allocation and Scheduling in MIMO-OFDMA with Time-Varying Channel and Bursty Arrivals

Kai Sun,Wei Huang,Yong Tie
College of Electronic Information Engineering
Inner Mongolia University
Hohhot, China
Email: sunkai1501@gmail.com

Yong Zhang
Wireless Technology Innovation Institute
Beijing University of Posts and Telecommunications
Beijing, China
Email: xiazibin@gmail.com

***Abstract*—A cross-layer resource allocation and scheduling scheme in the downlink of MIMO-OFDMA systems is proposed to provide quality of service (QoS) guarantees to real-time traffics and non-real-time traffics in case of heavy-traffic. The proposed algorithm aims to maximize the efficient system throughput with the constraints of packet error rate, delay, packets dropping rate and buffer size. At the medium access control (MAC) layer, the proposed algorithm dynamically assigns subcarriers to the highest priority user based on the channel state information (CSI) and the queue state information (QSI). After initial subcarrier assignment, the subcarrier re-assignment is applied to approve the system performance. At the physical (PHY) layer, a low complexity rate-and-power adaptation scheme is used to the predetermined subcarriers. It is demonstrated in the numerical results that the proposed algorithm significantly improves the system performance in terms of system throughput and packet loss rate.**

I. INTRODUCTION

Spatial Multiplexing offers high channel capacities or transmission rate for the same bandwidth and with no additional power requirements by employing multiple antennas at the transmitter and receiver [1]. However, high data transmission is limited by inter-symbol-interference (ISI). Orthogonal frequency division multiplexing (OFDM) uses the spectrum efficiently by spacing the channels closer together as well as it gives the ability of reducing ISI. The combination of these two technologies has been researched for the most promising candidate technique for the next generation wireless systems [2]. Recently, channel-aware scheduling has received much attention since it can exploit the variations of fading channels and earn multiuser diversity gain to dramatically improve the spectral efficiency in code division multiple access (CDMA) systems [3]. Moreover, utility function has also been widely used. That is because fairness criterion plays a more and more important role in resource scheduling, and utility theory can conveniently do tradeoff between efficiency and fairness [4]. The proportional fair properties have been studied empirically in [5], and a OFDM and multiple antenna implementation of such an algorithm over slowly time-varying channels has been proposed in [6], [7].

However, the channel-aware-only schedulers work with two implicit assumptions that there is no constraint on the maximum size of the transmission buffer (or queuing delay) and the scheduled users always have data to transmit. This is true for the application is delay-tolerant and the buffer size is relative large. When the application is delay-sensitive and when the system buffer size is limited, delay performance and stability are important considerations. Ideally, a well designed cross-layer scheduler should provide a good match between the randomness of the traffic patterns and that of environment while being subject to QoS constraints. A seminal work for incorporating randomness and stability issues has been provided in [8] where a system is said to be stable if queue sizes do not grow to infinity during normal operation. And it has been shown that allocating resources to maximize a queue-length-weighted sum of the rates (which are feasible in the current time slot) is a stability policy. This idea has been further developed into a general of queue length based or delay based scheduling schemes that focus on stability and throughput optimality, such as modified largest weighted delay first (M-LWDF) [9], exponential scheduler (EXP) [10] and max delay utility (MDU) [4]. However, the aforementioned works assume that the arrival rate is aware and lies inside capacity region of the network. In a real wireless network, the base station (BS) could not exactly know each mobile equipment's (ME) arrival rate. Furthermore, in heavy-traffic or deep-fading scenario, the arrival rate would lie out the capacity region, and this causes the packet dropping due to the buffer overflow .

In this paper, we focus on the cross-layer resource allocation and scheduling scheme in the downlink of multiuser MIMO-OFDMA systems. Allocating limited resources at MAC layer and PHY layer among users are formulated as optimization problem, where the objective is to maximize efficient system throughput under PHY layer constrains and QoS constraints on the MAC layer. After the analysis of the average packet dropping probability and average packet delay performance, a low complexity scheme, namely joint space-frequency-power scheduling algorithm (JSFP), is proposed. In JSFP, a unified

CHINACOM 2010, August 25-27, Beijing, China

DOI 10.4108/chinacom.2010.18

scheduling priority is first given by taking into account bit error rate (BER) or packet error rate (PER), estimated packet delay, queue length, estimated arrival rate and user fairness. The subcarriers are dynamically allocated to the user with the highest priority. Then, a subcarrier reassignment is carried out to increased the system performance based on the buffer size and estimated packet dropping rate. Finally, a low complexity adaptive bits and power loading algorithm is applied to the predetermined subchannels for each user in MIMO-OFDMA system. Simulation results show that the proposed algorithm can achieve good performance in terms of throughput and packet dropping rate.

The remainder of this paper is organized as follows. Section II presents the system model and problem formulation. Section III.A firstly analyzes the user's packet level performance. Then the proposed scheduling and resource allocation algorithm is given in section III.B. The simulation results and analysis are given in section IV. Finally, Section V concludes this paper.

II. System Model and Problem Formulation

A. System Model

We consider the downlink of the network with total number of K mobile Users. The block diagram of the MIMO-OFDMA downlink system is depicted in Fig. 1. The BS is equipped with N_t transmit antennas and the kth ME is equipped with N_r receive antennas.

Information packets from upper layer, such as the application layer, destined to the users arrive randomly at the base station, and are stored into K finite size buffers (queues), where buffer k is associated to user k and serves in a first-in-first-out manner. The data link layer is responsible for the final encapsulation of higher-level messages into frames that are sent over the network at the physical layer. At data link layer, a time-frame model with $A[t]$ being the number of packets arriving at time frame t is considered. For real-time traffic, we assume constant packet rate traffic with constant packet arriving rate λ_1 packets per time frame. The Poisson distributed bursty traffic with average arrival rate λ_2 packets per time frame is used as non-real-time traffic, that is, $A[t]$ is Poisson distributed with parameter $\lambda_2 T$

$$P\{A[t]=a\}=\frac{(\lambda_2 T)^a e^{-\lambda_2 T}}{a!} \tag{1}$$

which leas to $E\{A[t]\}=\lambda_2 T$. T is the frame duration and the packet length is fixed with L bits. The packet scheduler mainly operates at MAC layer and collects the cross-layer information (i.e., CSI and QSI) and select different users packets for transmission.

At the PHY, we deal with frame by frame transmissions, where each frame contains a fixed number of symbols. Each frame at the physical layer may contain one or more packets from the data link layer. We denote the total number of orthogonal subcarriers by N_c. The channel between the BS and ME k on subcarrier n is flat fading and is denoted by a $N_r \times N_t$ matrix $\mathbf{H}_{k,n}$. It is assumed that the channel is constant during one OFDM frame transmission time and may change frame by frame. The CSI is perfectly known by the receiver, and each ME feedbacks a certain form of channels information correctly to the BS. In this paper, the V-BLAST algorithm implementation based on zero-forcing detection combined with symbol cancellation is applied to improve the performance while maintaining low implemental complexity. For the case of $N_t \geq N_r$, the subset of N_r transmit antennas can be determined by some selection algorithms operating at the receiver. Let $\mathbf{H}_{k,n}^a$ denote the $N_r \times N_r$ channel matric corresponding to the selected transmitted antennas subset. After multiplying post-processing vector by the received signal vector of the ME k on subcarrier n, we can obtain

$$\mathbf{z}_{k,n}=\mathbf{u}_{k,n}\mathbf{y}_{k,n}=\sqrt{\frac{P_{k,n}}{N_t}}\mathbf{u}_{k,n}\mathbf{H}_{k,n}^a\mathbf{s}_{k,n}+\mathbf{u}_{k,n}\mathbf{n}_{k,n} \tag{2}$$

where $\mathbf{u}_{k,n}$ is the post-processing vector and is the pseudo-inverse of the channel $\mathbf{H}_{k,n}^a$. $\mathbf{y}_{k,n}$ is the received signal vector, $\mathbf{s}_{k,n}$ is the complex transmitted signal vector and $P_{k,n}$ is the average transmit power on carrier n for user k. $\mathbf{n}_{k,n}$ is the noise vector and its elements are independent identically distributed (i.i.d) circularly symmetric complex Gaussian variables with zero-mean and variance of σ^2.

Without loss of generality, we adopt uncoded M-QAM modulation with BER constraint. The achievable transmission rate (bit/symbol) of ME k's ith spatial subchannel on subcarrier n, can be expressed as

$$r_{k,n,i}[t]=\log_2\left(1+\Gamma\frac{p_{k,n,i}\lambda_{k,n,i}^2}{\sigma^2}\right) \tag{3}$$

where $p_{k,n,i}$ is the allocated power, $\lambda_{k,n,i}$ is the singular value of $\mathbf{H}_{k,n}^a$ and Γ is a constant related to a targeted BER by [11]

$$\Gamma=\frac{-1.5}{\ln(5B_k)} \tag{4}$$

In realistic systems, only a discrete finite constellations set is available. So the integer value $b_{k,n,i}$ is obtained by rounding the value of $r_{k,n,i}$ to either 0,1,2,4 or 6. Let $\mathcal{N}_k[t]$ denote the set of subcarrier indices that are assigned to user k at frame t. Then the data transmission rate of user k at frame t, $R_k[t]$, is given by (bits/s)

$$R_k[t]=\sum_{k\in\mathcal{N}_k[t]}\sum_{i=1}^{M}\frac{b_{k,n,i}[t]}{T} \tag{5}$$

where M is the number of spatial subchannels of ME k.

B. Problem Formulation

The packet loss rate is affected by both the queuing overflow, and the packet reception error. If we assume that all bits are independent and are equally protected in a packet, then PER due to the packet reception error can be written as a function of BER

$$P_k^0=1-(1-B_k)^L\approx LB_k \tag{6}$$

In practice, the queue length (buffer size) is finite and is assumed to be Q_k packets. When the buffer is full and

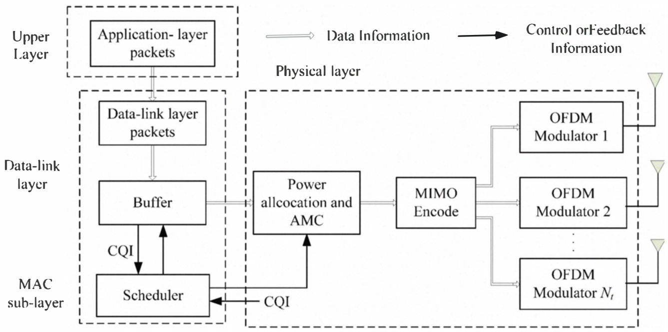

Fig. 1. Block diagram of MIMO-OFDMA downlink system

overflow occurs, excess packets have to be dropped. Let P_k^d denote the packet dropping rate due to the buffer overflow. Then the effective system throughput can be written as

$$C = \sum_{k=1}^{K} \lambda TL \left(1 - P_k^0\right)\left(1 - P_k^d\right) \tag{7}$$

In this paper, we mainly concentrated on the effect of overflow on the overall system performance. Automatic Return reQuest (ARQ) is applied to recover the packet errors, and a packet is retransmitted until it is received successfully or is dropped due to the buffer overflow. We would like to find the scheduling algorithm so that the effective system throughput is maximized under a total power constraint. The optimal scheduling algorithm can be expressed as

$$\max \sum_{k=1}^{K} \lambda TL \left(1 - P_k^0\right)\left(1 - P_k^d\right) \tag{8}$$

subject to

$$\bigcup_{k \in \mathcal{U}[t]} \mathcal{N}_k \subseteq \mathcal{N} \tag{8a}$$

$$\mathcal{N}_k \cap \mathcal{N}_j = \emptyset, \quad k \neq j \ \ \forall k, j \in \mathcal{K}[t] \tag{8b}$$

$$\sum_{n=1}^{N_c} \sum_{i=1}^{M} p_{k,n,i} \leq P, \quad n \in \mathcal{N}_k \tag{8c}$$

where $\mathcal{N} = \{1, \ldots, N_c\}$ is the subcarrier index set, $\mathcal{K}[t]$ is the non-empty queue user index set at the beginning of the frame t. Constraints (8a) and (8b) are determined by the characteristics of OFDM systems. The constraint (8a) means that each user selects subcarriers from $\mathcal{N}$ and the constraint (8b) means that each subcarrier can only be assigned to one user. The constraint (8c) shows the total power P constraint.

III. Queuing Analysis and Proposed Algorithm

In this section, we first analyze the predicted average dropped packets due to buffer overflow and the predicted average packet waiting time. Then a low complexity algorithm based on utility function is given.

A. Queuing Analysis

Let $Q_k[t]$ be the amount of packets in the queue of user k at the beginning of the frame t. During time frame t, the BS services user k at rate $R_k[t]$, and the number of packets left in the queue is

$$L_k[t] = \max\left\{0, Q_k[t] - \frac{R_k[t]T}{L}\right\} \tag{9}$$

Then the queue length of user k at the beginning of the frame $t+1$ or at the end of the frame t can be expressed as

$$Q_k[t+1] = \min\left\{Q_k, L_k[t] + A_k[t]\right\} \tag{10}$$

The amount of dropped bits due to the buffer overflow at the end of the frame t or at the beginning of the frame $t+1$ can be expressed as

$$D_k[t+1] = \max\left\{0, L_k[t] + A_k[t] - Q_k\right\} \tag{11}$$

In order to avoid wasting the resource, the BS controls service data rate so that

$$R_k[t] \leq \frac{Q_k[t]L}{T} \tag{12}$$

According to service rate constraint in (12), (10) and (11) can be rewritten as

$$Q_k[t+1] = \min\left\{Q_k, Q_k[t] - \frac{R_k[t]T}{L} + A_k[t]\right\} \tag{13}$$

$$\begin{aligned} D_k[t+1] &= \max\left\{0, Q_k[t] - \frac{R_k[t]T}{L} + A_k[t] - Q\right\} \\ &= \max\left\{0, Q'_k[t+1] - Q\right\} \end{aligned} \tag{14}$$

where $Q'_k[t+1]$ is the queue length of user k without of the buffer limit at the beginning of the frame t. Inspired by the analysis method in [4], we denote the average queue over the a time window of user k without of the buffer limit at the beginning of the frame t as

$$\bar{Q'}_k[t] = \left(1 - \frac{1}{t_c}\right)\bar{Q'}_k[t-1] + \frac{1}{t_c}Q'_k[t] \tag{16}$$

where t_c is the time window which is the integer multiple of T. At the beginning of frame t, given the average queue length $\bar{Q'}_k[t]$ and actual queue length $Q_k[t]$, the predicted average queue length without queue limit at the beginning of the frame $t+1$, $\hat{Q'}_k[t+1]$, can be written as

$$\begin{aligned} \hat{Q'}_k[t+1] &= E_{A_k[t]}\left\{\bar{Q'}_k[t+1]\right\} \\ &= \left(1 - \frac{1}{t_c}\right)\bar{Q'}_k[t] + \frac{1}{t_c}E_{A_k[t]}\left\{Q'_k[t+1]\right\} \\ &= \left(1 - \frac{1}{t_c}\right)\bar{Q'}_k[t] + \frac{1}{t_c}\left\{Q'_k[t] - \frac{R_k[t]T}{L} + \lambda_k T\right\} \end{aligned} \tag{17}$$

According to (14) and (17), the predicted average dropped packets and predicted average dropped rate at the beginning of the frame $t+1$ can be expressed as

$$\hat{D}_k[t+1] = \max\left\{Q_k, \hat{Q'}_k[t+1]\right\} = \mathrm{F}\left(R_k[t]\right) \tag{18a}$$

$$\hat{P}_k[t+1] = \frac{\hat{D}_k[t+1]}{\lambda_k} \tag{18b}$$

Since $\hat{Q'}_k[t+1]$ is a non-increasing function related to $R_k[t]$, $F(R_k[t])$ is also a non-increasing function. The use's data rate should meet $\hat{P}_k \leq P_k^E$. P_k^E is the maximum allowable packet dropping rate of user k. We can express this data rate by

$$R_k[t] \geq \mathrm{F}^{-1}\left(\lambda_k \cdot P_k^E\right) \tag{19}$$

According to Little's law [12], the predicted average waiting time at the end of frame $t+1$ can be expressed as

$$\hat{W}_k[t+1] = \frac{\hat{Q}_k[t+1]}{\lambda_k} \tag{20}$$

where $\hat{Q}_k[t+1] = \max\{\hat{Q'}_k[t+1], Q\}$.

B. Packet Scheduling and Resource Allocation

The algorithm is composed of three parts: initial subcarrier assignment with constraint (12), subcarrier reassignment and adaptive bit and power loading.

In initial subcarrier assignment part, subcarrier n is assigned to the user with the largest scheduling priority which is described as follows

$$\mu_k[t] = \frac{\left(\hat{W}_k[t]\right)^{\gamma-1}}{\hat{\lambda}_k} r_{k,n}[t] \tag{21}$$

where $\hat{\lambda}_k$ is the estimated packets arrival rate due to the BS does not know the real value arrival rate before scheduling operation finished. $\hat{\lambda}_k$ can be estimated through an exponentially smooth low-pass window similar to the definition in (16). γ is used to provide different degrees of delay fairness for users. $r_{k,n}$ is the data rate of user k on subcarrier n. Obviously, the users with longer waiting time and/or higher instantaneous data rates will be given higher priority.

To get the simple decoding at the receiver side and reduce the complexity of the algorithm, we use the same modulation scheme on all spatial sub-channels to user k on subcarrier n. Moreover, the performance of spatial multiplexing with linear receivers depends on the minimum post-detection SNR [13]. For the ZF receiver, the post-processing SNR of the worst sub-channel can be expressed as [13]:

$$\Lambda_{k,n}^{min} \geq \lambda_{min}^2\left(\mathbf{H}_{k,n}^a\right)\frac{P}{N_c M \sigma^2} \tag{22}$$

where $\lambda_{min}\left(\mathbf{H}_{k,n}^a\right)$ is the minimum singular value of channel $\mathbf{H}_{k,n}^a$. We use the maximum minimum singular value antenna subset selection introduced in [13] to construct $\mathbf{H}_{k,n}^a$. That is, for every subset of transmit antennas compute $\lambda_{\min}$ and choose the subset with the largest $\lambda_{\min}$. Therefore, the user's data rate $R_k[t]$ in (21) can be estimated by substituting (22) into (3) and (5). When the subcarrier n is allocated to user k, this user's data rate is updated by $R_k[t] = R_k[t] + b_{k,n}/T$. $b_{k,n}$ is the amount of bits that user k can transmit on subcarrier n. If $R_k[t] \geq Q_k[t]L/T$, it means the queue will be empty after this service, and then this user should be removed from the scheduling user set as not to waste resource.

After initial subcarrier assignment, let $\mathcal{U}$ be the user set where the users' rates meet (19), that is $\mathcal{U} = \left\{k : R_k[t] \geq \mathrm{F}^{-1}\left(\lambda_k \cdot P_k^E\right)\right\}$. Let $\mathcal{U}^c = \mathcal{K} - \mathcal{U}$ be the user set where each user's packet dropping performance is not satisfied. If $\mathcal{U} = \mathcal{K}$, each user's packet dropping performance is satisfied. Then the adaptive bit and power loading is applied under the predetermined subcarrier assignment; otherwise, subcarrier reassignment is needed to enhance the performance.

For a user and subcarrier pair (i, j, n) $\left(i \in \mathcal{U}, j \in \mathcal{U}^c, n \in \mathcal{N}_\mathcal{U} = \bigcup_{i\in\mathcal{U}} \mathcal{N}_i\right)$, if user i's data rate can still satisfy the QoS performance constrain in (19) and the corresponding total utility is the largest among all reassignment attempts, then the subcarrier n of user i is allocated to user j. Here the total utility corresponding user and subcarrier pair (i, j, k) is denoted by

$$\mu_{i,j,n} = \frac{\left(\hat{W}_i\hat{W}_j\right)^{\gamma-1}}{\hat{\lambda}_i\hat{\lambda}_j}\left\{R_i - b_{i,n}\right\}\left\{R_j + b_{j,n}\right\}\prod_{k\neq i,j}\mu_k$$

The subcarrier reassignment will continue until $\mathcal{U}^c = \emptyset$ or there are not enough subcarriers that could be allocated to the users in $\mathcal{U}^c$.

Given the subcarrier allocation and $b_{k,n}$, (3) is used to calculate the transmit energy for each sub-channel. And then, the transmit power is distributed to the space and frequency according to the result of $p_{k,n,i}$ and under the constraint of total transmit power. The method ensures that the current total power allocated to the space and frequency is equal to the total transmit power of the system.

IV. Simulation Results

A. Simulation Design

In the simulation, 5 MHz bandwidth is divided into 300 subcarriers. The cyclic prefix length is 8.33 s. The subcarrier spacing is 15 kHz and 12 successive subcarriers are grouped as a sub-band. We group 70 OFDM symbols into an $T = 5$ ms OFDM frame. The multipath fading channel model is the tapped delay-line spatial channel model (SCM) [14]. We assume that the numbers of real-time users and non-real-time users are 10, respectively. All users have independent fading channel characteristics. Each ME moves randomly with a velocity of 3 km/h. The carrier frequency is 2 GHz. Assume that each MS is equipped with $N_r = 2$ antenna and the BS is equipped with $N_t = 4$ antennas. Let the acceptable BER, PER and fairness factor γ be 10^{-5}, 10^{-2} and 2, respectively. The available modulation orders are constrained to 0, 2, 4, 16, 64, where 0 means that no data is transmitted. We assume the buffer size is $Q_k = 300$ packets and the packet length is $L = 840$ bits/packet. For comparison, the PF [6] algorithm is also evaluated.

B. Simulation Results

Fig. 2 and Fig. 3 show the average efficient system throughput and average packet dropping rate versus the arrival rate for JSFP and PF when the average transmitted SNR of one subcarrier is 10 dB. In the simulation, all users have same path gain ($pl_k = 1$). As we have expected that JSFP achieves good performance compared to PF because of that PF only considers the CSI and ignores the QSI. For example, when the arrival rate is 1428 Kbps per user, the efficient system throughput of JSPF can reach 28.54 Mbps, which is 10.89% larger than that of PF. When The packet the arrival rate is 1512 Kbps per user, the dropping rate of JSPF is only 2.373%, which is 6 times shorter that of PF.

Fig. 4 shows the average throughput and packet dropping rate of users with different path gains and different arrival rates. In the simulation, the path gains are $pl_k = 10 - 0.3 \times (k-1))$ dB for RT or NRT users. The users' arrival rates are $\lambda_k T = 9 \times 0.95^{(k-1)}$ for RT or NRT. Although PF works well under the assumption that the buffer size is infinite and there are always data to transmit, the performance can not be guaranteed when the application is delay-sensitive application or buffer size is finite.

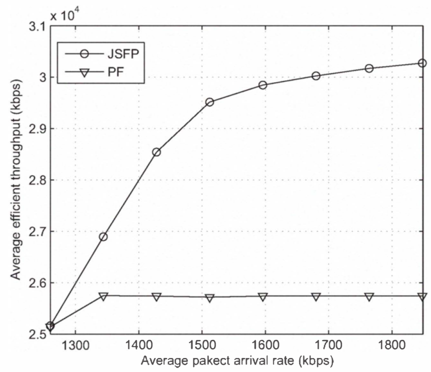

Fig. 2. Average efficient system throughput versus average packet arrival rate

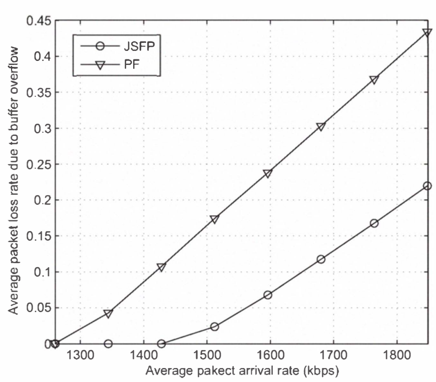

Fig. 3. User average packet dropping rate versus average packet arrival rate

V. Conclusion

In this paper, a cross-layer resource allocation and packet scheduling scheme is proposed for multiuser MIMO-OFDMA systems based on users' CSI and QSI. Simulation results show that the proposed algorithm has a good balance between the system throughput and the QoS requirements, and it is important for the cross-layer scheduling design to exploit the channel state information and queue status.

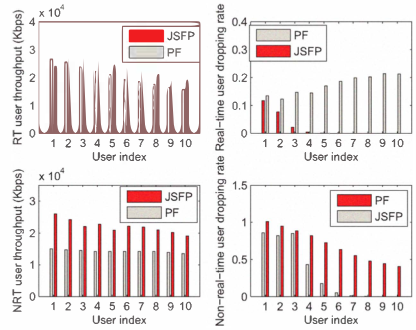

Fig. 4. Performance comparison of JSFP and PF with different arrival rate and path loss.

Acknowledgment

This work is supported by Foundation of NeiMongol Educational Committee (NJ10004) and Program of Higher talents ofInnerMongoliaUniversity(Z20090136).

References

[1] G. Caire, and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. on Inf. Theory*, vol. 49, no. 7, pp. 1691-1706, July 2003.
[2] Y. J. Zhang and K. B. Letaief, "An efficient resource allocation scheme for spatial multiuser access in MIMO/OFMD systems," *IEEE Trans. Commun.*, vol. 53, no.1, pp. 107-116, Jan. 2005.
[3] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Networking*, vol. 13, no. 3, pp. 636C647, June 2005.
[4] G. Song and Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband network," *IEEE Communications Magazine*, vol. 43, no. 12, PP. 127-134, Dec. 2006.
[5] R. Padovani, A. Jalali, and R. Pankaj, "Data throughput of cdma hdr a high efficiency-high data rate personal communication wireless system," in *proc. VTC2000-Spring*, July 2000, pp.1854-1858.
[6] Z. Shen, J. G. Andrews, B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional fairness," *IEEE Trans. Wireless Commun.*, vol. 4, Nov. 2005, pp. 2726-2737.
[7] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inf. Theory*, vol. 48, no. 6, pp. 1277-1294, June 2002.
[8] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput inmultihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 1936-1948, Dec. 1992.
[9] M. Andrews, K. Kumaran, K. Ramanan, A.Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commu. Mag.*, vol. 39, no. 6, pp. 150-154, Feb. 2001.
[10] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: the exponential rule," *Translations of the AMS, Series 2, A volume in memory of F. Karpelevich*, vol. 207, Sep. 2002.
[11] A. J. Goldsmith and S. Chua. "Variable-rate vaiable-power M-QAM for fading channels," IEEE Trans. Wireless Commun., vol. 45, pp.1218-1230, Oct. 1997.
[12] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed. New York: Wiley,1998.
[13] R. W. Heath, S. Sandhu, and A. Paulraj. "Antenna slection for spatial multiplexing systems with linear receivers," *IEEE Communi. Letters*, vol. 5, no. 4, pp. 142-144, Apr. 2001.
[14] 3GPP TR 25.996 v 6.1.0. Spatial channel model for multiple input multiple output (mimo) simulations (release 6). 2003.