Resource Allocation in Wideband Wireless Systems

Marc Realp[1], Raymond Knopp[2] and Ana I. Pérez-Neira[3]

[1]marc.realp@cttc.es; Telecommun. Technological Center of Catalonia (CTTC); Barcelona (Catalonia-Spain)

[2]knopp@eurecom.fr; Institut Eurécom; Sophia Antipolis (France)

[3]anuska@gps.tsc.upc.es; Technical University of Catalonia (UPC); Barcelona (Catalonia-Spain)

***Abstract*— We consider the problem of resource allocation in multiuser single-antenna wideband OFDM(A) systems. The key advantage of such systems with respect to narrow band systems is the possibility of considering frequency as an additional resource to be allocated. Although the maximum normalized average throughput is not increased with respect to that in a narrowband system, a more efficient use of resources is possible especially if the bandwidth is considerably larger than the coherence bandwidth of the channel and the channel is varying slowly with respect to the scheduling updates. This is mainly because randomness in the system is increased by the wideband resources. The work presented in this paper analyzes the effects of bandwidth on the delay characteristics. To this end, the relationship between ergodic information rates, stability and delay in multiuser communications systems is studied and candidate resource allocation policies are presented and simulated.**

I. Introduction

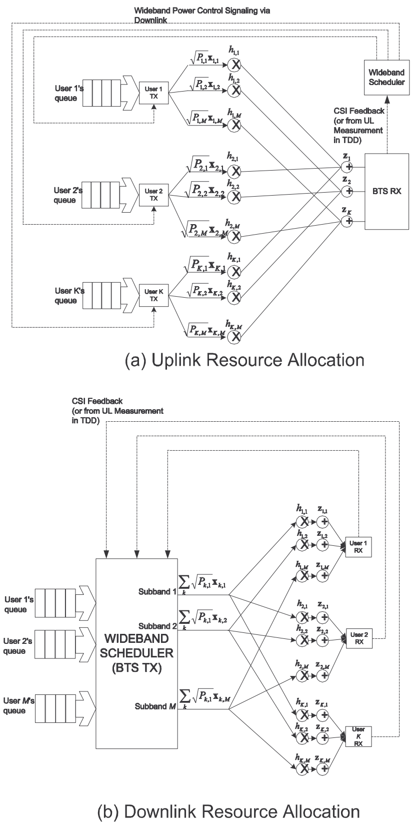
We consider resource allocation strategies for multiuser single-antenna systems transmitting over *wideband frequency-selective channels*. This could represent the case of any wideband OFDM system, such as evolving Mobile Broadband Wireless Access (MBWA) systems, for example the IEEE 802.16 standard where an Orthogonal Frequency Division Multiple Access (OFDMA) technique is proposed. Another example of such system model could be the UTRAN HSDPA (high-speed data packet-access) 3GPP proposal using an OFDM(A) physical layer instead of WCDMA for the downlink channel[1]. We can also conceive the use of these techniques in extensions of IEEE802.11a/g (e.g. 802.11n), Hiperlan2 or multiband-OFDM UWB systems. In this context, the algorithms proposed in this paper would be used to allocate the different frequency subbands and appropriate power levels on each sub-band to users.

We are interested in resource allocation strategies which exploit multiuser diversity by means of accurate channel state information at the transmitter. The gains offered by multiuser diversity techniques with respect to constant power allocation over all resources can be seen as either a significant increase in spectral efficiency as the user population grows (which amounts essentially to a factor of two for low signal-to-noise ratios) or equivalently a transmit power savings of around 6dB for [2].

However, the key advantage of wideband OFDM(A) systems with respect to narrowband systems is the possibility of performing multiuser scheduling both in time and frequency. Although the attainable average throughput (normalized with respect to the bandwidth) is not increased by the wideband resources [3], [4], the additional dimensions potentially allow for a more efficient use of the channel due to the increased randomness in the system. This randomness can be beneficial if constraints are placed in order to guarantee a certain instantaneous bandwidth or maximum delay. The latter are particularly important for today's circuit-switched applications (e.g. voice, real-time video) if they are to be run effectively on wireless packet networks. In the context of such cross-layer views, one could pose the problem of finding wideband resource allocation strategies guaranteeing the peak queue length as opposed to average queue length for a given link. To this end, [5] considers orthogonal allocation and power control strategies guaranteeing a deterministic channel use (i.e. guaranteed instantaneous bit-rate) for parallel (e.g. OFDMA) slowly fading channels with multiple-antennas. Although clearly sub-optimal from the point-of-view of the delay-limited capacity region [6] (from the point-of-view of short-term fading), which to-date remains an open-problem for frequency-selective multi-antenna channels, it is shown that reasonably simple orthogonal allocation strategies can yield both multiuser diversity and spatial multiplexing. The achievable rates of these strategies approach those of the ergodic sum-rate, however with strict guarantees on channel use.

Average packet delays and channel rates were considered by Neely and Modiano in [7]. The authors analyze the stability and delay of power and rate allocation in a multibeam satellite downlink which transmits data to K different ground locations over K time varying channels. They present a resource allocation algorithm that, according to the queue lengths and the channel state allocates power and rate in order to achieve system stability. The work of Neely and Modiano was extended by Yeh and Cohen to the multiple access and broadcast wireless channels in [8]. In that work, the authors presented a resource allocation policy that allocates power and rate considering the queue length as a reward. Hence, the queue length establishes a priority order in the allocation of resources. Both works, [7] and [8], assume a block fading channel and rely on the variability of the channel by performing the scheduling updates once every channel realization. Hence, this delay is determined by the channel coherence time. In [9], Boche and Wiczanowski considered the Multiple Input Multiple Output (MIMO) multiple access channel ending up with the same resource allocation policy as in [8]. Furthermore, in order to decrease the average

[0]the research reported in this work was supported in part by the European Commission Framework Programme 6 Information System Technology collaborative research projects NEWCOM and WIDENS. The work of R. Knopp is also partially supported by Eurecom and its industrial partners: Hasler Stiftung, Swisscom, France Telecom, La Fondation Cegetel, Bouygues Telecom, Thales, ST Microelectronics, Hitachi Europe, Sharp and Texas Instruments.



(a) Uplink Resource Allocation

(b) Downlink Resource Allocation

Fig. 1. System Scenarios

delay, they considered the possibility of reallocating resources many times during one channel period. In [10], Kobyashi et al. considered a SDMA/TDMA system with feedback delay where the variability of the channel was increased by using opportunistic beamforming techniques. The latter was also considered in [11] where the limitations in terms of delay when performing opportunistic beamforming for single user channels is assessed and the use of multiple channels in order to consider multiuser communications is introduced. In this work, we follow the approach of [7], [8], [9], [10] to study resource allocation in a wideband frequency selective slow fading channel in order to assess improvements in terms of average packet delay.

The outline of this work is as follows: Section II provides the underlying wideband channel models and system architecture. Section III presents the relationship between ergodic throughput, stability and delay in multiuser communications as well as candidate resource allocation policies. In section IV we present numerical examples showing the effect of system bandwidth on the delay characteristics. Finally, in section V we present our conclusions.

II. Wideband System and Channel Model

In order to model wideband channels we assume M parallel discrete-time channels. This is a typical way of discretizing a waveform channel [12]. Moreover, in the context of digitally sampled OFDM systems, the use of a cyclic prefix allows the channel to be considered as a memoryless system in the discrete Fourier transform-domain (DFT) and the parallel channels just represent frequency samples over the system bandwidth. Each channel is characterized by a complex channel gain $h_{k,m}$, $k = 1, 2, \cdots, K$, $m = 1, 2, \cdots, M$, which corresponds to the amplitude and phase of the m^{th} channel for user k.

The systems under consideration are shown in Figure 1 and are the downlink and uplink channels in cellular or Wireless LAN network topologies. Channel access is assumed to be time-slotted, with slot duration T sec. The scheduler will update the power and rate allocations of the user streams every scheduling period $T_{sch} = N_{sch}T$ sec. Note that in a real system there will necessarily be a signaling delay of a few slots for uplink scheduling which is neglected here for simplicity. Interference between slots is neglected by using appropriate guard-times of duration greater than the typical delay-spread of the propagation channels and significantly shorter than T so that information rate loss is negligible. The channels are assumed to remain constant during one frame duration $T_{fr} = N_{fr}T$ and to vary across frames in a time-uncorrelated manner. Information bits for user m are retrieved from a queue which buffers packets of L_p bits each and arriving in the queue with average rate ρ_k packets/sec, $k = 1, 2, \cdots, K$. Packets are assumed to arrive synchronized to the slot-time. The average information rate across the wireless channel is $\rho_k \overline{L_p}$.

Information for user k is encoded with rate R_k bits/dimension in the form of codewords of length NM dimensions to be transmitted every slot, where N is the number of dimensions per channel in a transmission slot. Codewords are made up of symbols $x_{k,m,n}$ with unit average energy. The goal of the resource allocation strategy is to select the code rates R_k and the appropriate transmit powers $P_{k,m}$ for each user every T_{sch} slots. After rate and power allocation, the number of backlogged bits remaining in queue k in scheduling period l is

$$B_{k,l} = [B_{k,l-1} - N_{sch}NMR_k]^+ + \sum_{i=1}^{N_p} L_{p,i}, \qquad (1)$$

where $L_{p,i}$ is the length of the i^{th} packet arrival and N_p is the number of packets which have arrived in the scheduling period. In section IV we will examine the performance of different resource allocation algorithms in terms of average queue length per user $\overline{B_k}$. Notice that applying little's theorem, the average delay is given by $\overline{D_k} = \overline{B_k}/(\rho_k \overline{L_p})$.

A. Multiuser Channel Models

For the case of an uplink channel (multiple-access channel) the signal at the receiver is given by

$$r_{m,n} = \sum_{k=1}^{K} \sqrt{P_{k,m}} h_{k,m} x_{k,m,n} + z_{m,n}$$
$$m = 1, 2, \cdots, M, n = 1, 2, \cdots, N, \qquad (2)$$

where $P_{k,m}$ is the instantaneous transmit energy used by user k on channel m and $z_{m,n}$ is additive white complex circularly-symmetric Gaussian random sequence with variance σ_z^2 and mean zero. It is assumed that the receiver (basestation) can adjust the $P_{k,m}$ based on channel state information (CSI) measurements, and moreover that these are signalled (via the downlink) and received without error at the user terminals. The basestation estimates the CSI for each user from received pilots which are known signals transmitted *over the entire bandwidth over which power allocation is performed*. Note that for slowly-varying channels this is reasonably simple to accomplish and consumes little signaling bandwidth since the allocation remains invariant across several slots. The considered power constraint is

$$\mathrm{E}\left(\sum_{m=1}^{M} P_{k,m}\right) \le P_k, k = 1, 2, \cdots, K \tag{3}$$

where the expectation is over the random channels. For the downlink (broadcast channel), the signal at receiver k is given by

$$r_{k,m,n} = h_{k,m} \sum_{k'=1}^{K} \sqrt{P_{k',m}} x_{k',m,n} + z_{k,m,n}$$
$$m = 1, 2, \cdots, M, n = 1, 2, \cdots, N, k = 1, 2 \cdots, K \tag{4}$$

where $z_{k,m,n}$ is additive white complex circularly-symmetric Gaussian random sequence with variance σ_z^2 and mean zero. The considered power constraint is

$$\mathrm{E}\left(\sum_{m=1}^{M} \sum_{k=1}^{K} P_{k,m}\right) \le P, \tag{5}$$

where the expectation is over the random channels. Note that this is the general non-orthogonal broadcast channel.

III. Ergodic Information Rates and Resource Allocation Policies

A complete characterization of the ergodic capacity region of wideband fading multiple-access channels was found [4]. The ergodic sum rate was found in [3]. In our discrete sub-band case, the ergodic capacity region is a solution to the optimization problem (for each of the parallel channels)

$$\max_{\mathbf{R},\mathbf{P}} \boldsymbol{\mu} \cdot \mathbf{R} - \boldsymbol{\lambda} \cdot \mathbf{P} \text{ s.t. } \mathbf{R} \in \mathcal{C}(\mathbf{h}, \mathbf{P})$$

where

$$\mathcal{C}(\mathbf{h}, \mathbf{P}) = \left\{ \mathbf{R} : \sum_{k \in S} R_k \le \mathrm{E}\left[\log_2\left(1 + \frac{1}{\sigma_z^2} \sum_{k \in S} |h_{m,k}|^2 P_{m,k}\right)\right], \forall S \subseteq \{1, 2, \cdots, K\} \right\},$$

$\boldsymbol{\mu}$ is a vector of rate rewards (priorities for each user) and $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers reflecting the total average power constraints for each user. The optimal information rate on each subband and $P_{m,k}$ are readily found by generalizing the results of [4] to the discrete-subband case. A particular user will be assigned power on a given sub-band if it yields the maximum increase in the objective function, and in general more than one user will be allocated power on a particular subband. As a result, in the general case, a multiuser receiver (e.g. using interference cancellation) is required to detect each user's signal because of the non-orthogonal channel access. A more practical realization of this wideband power allocation is to allocate constant power to the best user. The proportional fair scheduling (PFS) policy applied to a wideband channel does precisely this in the long-term [13]. Wideband PFS attempts to maximize $\sum_k \log(R_{m,k})/T_k$ where T_k is the long-term average throughput of user k and $R_{m,k}$ is the allocated rate on channel m. In the symmetric case, T_k is the same for each user and since $R_{m,k}$ is monotonic in $|h_{m,k}|^2$, wideband PFS amounts to choosing the user with the best channel on each subband with power KP_k/M. The added gain in performing additional waterfilling on the channel gain is quite negligible in Rayleigh fading except at very low signal-to-noise ratios [14].

For the broadcast channel (downlink channel) it was recently discovered that there exists a complete duality with the multiple-access channel which allows its characterization to be applied with a slight modification stemming from the power constraints[15]. For the symmetric broadcast channel case, user k is allocated to channel m if $|h_{m,k}|^2$ is largest and wideband PFS will allocate power P/M to each subband. The asymmetric case yields a non-orthogonal multiple-access scheme where multiuser receivers (i.e. interference cancellation) are required at the user terminals.

A. Random Packet Arrivals and Stability

In order to assess the impact on the average delay in a packet data system we follow the approach taken in [7], [8], [9], [10]. Neely *et al* introduced the notion of the stability region of a multiuser system with time-varying channels. It is a measure relating the properties of source arrival processes with the underlying channel capacity. A system is said to be stabilizable if there exist rate and power allocation policies such that the average queue length (in our case $\overline{B_k}$) remains bounded. Strong theorems exist for the case of finite-state arrival processes and channels with respect to the existence and construction of stabilizing allocation policies for multiple-access channels [8]. Specifically, any stationary power and rate allocation policy designed for arrival rates lying inside the multiple-access capacity region stabilizes a multiple-access queuing system over time-varying channels. Moreover, for a power and rate allocation policy operating over i.i.d. realizations of a channel (i.e. each time a new allocation is performed, the channel is an independent realization of a random channel) a blind policy taking advantage of the instantaneous backlogs of each of the queues (i.e. without having explicit knowledge of the arrival rates but assuming they lie within the capacity region) can be shown to stabilize the multiple-access queuing system. This policy simply associates the rate rewards from the Tse and Hanly characterization of the multiple-access channel capacity region with the instantaneous queue backlogs (i.e. $\mu_{k,l} = B_{k,l}$)

IV. Numerical Comparison of Practical Resource Allocation Policies

In this section, some numerical results are presented with two main purposes. First, to present the advantages of taking

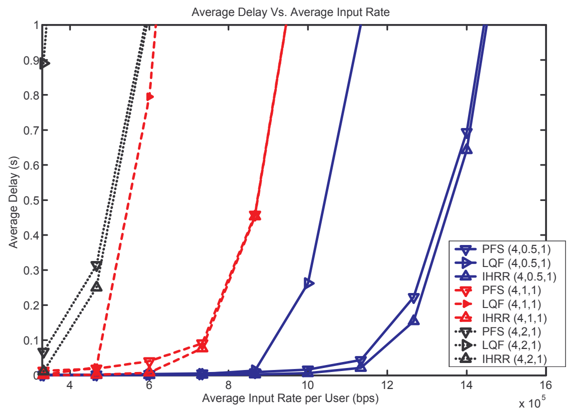


Fig. 2. Average delay for PFS, LQF and IHRR scheduling policies and different number of users per frequency subband.

a cross-layer approach when performing resource allocation in wireless systems. And second, to show the advantages of performing resource allocation in wideband wireless systems.

Three different resource allocation policies have been simulated. One policy is the PFS policy. As described in the previous section, in the case of symmetric channels, the PFS policy allocates resources on a per subband basis. Then, each subband is considered as a single user channel where the selected user is the one that shows the highest achievable rate. We have considered that the power allocation in each subband is such that the average received SNR is $0dB$.

The second resource allocation policy is a modification of the approach considered in [16], where an optimal scheduling policy for a K queue and a single server system was derived. The channel state of each queue was modelled as an ON-OFF process. At state ON, packets were transmitted without error and at state OFF no packet transmission was possible. Under such conditions, Tassiulas and Ephremides showed that serving the longest queue among those that are ON minimizes the total number of packets in the system when traffic is Bernoulli distributed. This is called the Longest Queue First (LQF) policy. In our case, the channel is not modelled by an ON-OFF process but is Rayleigh distributed. A modified LQF policy consists of performing the following iterative algorithm at the beginning of each scheduling period:

1) Choose the user $k = 1, 2, ..., K$ with the longest queue $B_{k,l}$
2) From those subbands that have not been already allocated, allocate the best subband m to that user.
3) Update the user queue length as $B_{k,l} = [B_{k,l} - N_{sch}NMR_{k,m}]^+$ where $R_{k,m}$ is the rate corresponding of allocating subband m to user k.
4) If the total number of allocated subbands is strictly less than M, go to step 1.

It is worth recalling that in scheduling period $l+1$, the kth user queue will follow the drift presented in equation (1) where $R_k = \sum_m R_{k,m}$. Furthermore, it has been considered constant power allocation per subband such that the average received SNR is 0dB.

The third resource allocation policy is a cross-layer policy based on those in [7], [8], [9], [10]. To be consistent with the two previous resource allocation policies, we considered also constant power allocation and single user detectors. Then, our policy is such that, at scheduling period l and for each channel m, choose the user that maximizes $B_{k,l} \cdot R_{k,m}$. In order to exploit the rate granularity offered by the wideband channel, we use the following algorithm:

1) Choose the user $k = 1, 2, ..., K$ with the maximum value $B_{k,l} \cdot R_{k,m}$ for m belonging to the set of non allocated subbands.
2) Update the user queue length as $B_{k,l} = [B_{k,l} - N_{sch}NMR_{k,m}]^+$ where $R_{k,m}$ is the rate corresponding of allocating subband m to user k.
3) If the total number of allocated subbands is strictly less than M, go to step 1.

We will name this policy Iterative Highest Rewarded Rate (IHRR).

The average queue length (and hence, average delay) as a function of the average input rate have been simulated for different scenarios. Packet arrivals follow a Poisson process with average input rate of $\rho_k \overline{L_p}$ bits per second (bps). A system is defined by the tupple $(M, \frac{K}{M}, N_{sch})$. Following the system specifications provided in [17], the following simulation values have been considered: slot duration $T = 300\mu s$ and frame duration $T_{fr} = 64T = 19.2ms$. Hence, N_{sch} ranges from 1 to 64. Furthermore, the bandwidth for each frequency subband is equal to $0.5MHz$.

In figure 2, we observe that by reducing the ratio of users per frequency subband $\frac{K}{M}$, there are more resources per user and hence, average delay can still be bounded at higher average input rates. We also observe that the best policy is the IHRR policy which takes a cross-layer approach by considering both the queue state and the channel state in the resource allocation policy. For instance, we observe that for a $(4, 1, 1)$ system the IHRR policy outperforms the PFS policy in approximately $50ms$ at $600kbps$ per user. On the other hand, we observe that results for the LQF scheduling policy are far from those concerning PFS and IHRR policies.

In figure 3 and 4 we show results obtained on the average delay and delay variance in slow fading wideband frequency selective channels when the number of carriers in the system is increased. We observe that although the number of users per subband is kept constant, diversity in the system is increased by increasing the number of subbands and hence, resources can be exploited more efficiently. Average delay results, show that IHRR policy better exploits resources than PFS and, for instance, for a $(8, 2, 1)$ system a gain of about $100ms$ can be achieved at $400kbps$. From figure 3 one could expect that ideally, with $M = \infty$, the curve of delay would be a step function of $0s$ of delay at any average rate inside the stability region and infinite delay otherwise. For the delay variance, we observe that increasing the system bandwidth also allows bounding the delay variance for either the PFS policy or the IHRR policy. Particularly, we observe that for average input rates that give reasonable average delays, delay variance is very low indicating that average delay approximates to the peak delay. This is a very interesting result for real time applications where peak delays are very harmful.

It is also interesting to see the impact of duration of the scheduling period N_{sch} on the average delay. Clearly, from equation (1) one could expect that shorter scheduling periods

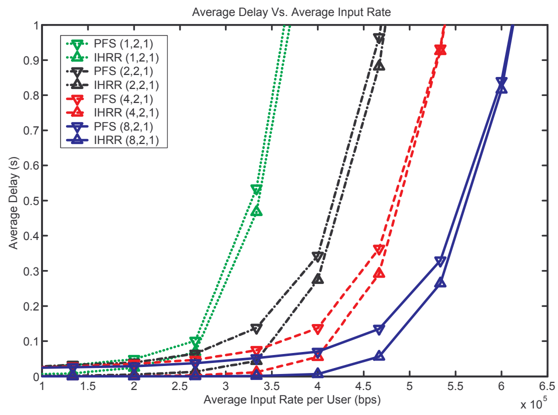


Fig. 3. Average delay for PFS and IHRR scheduling policies.

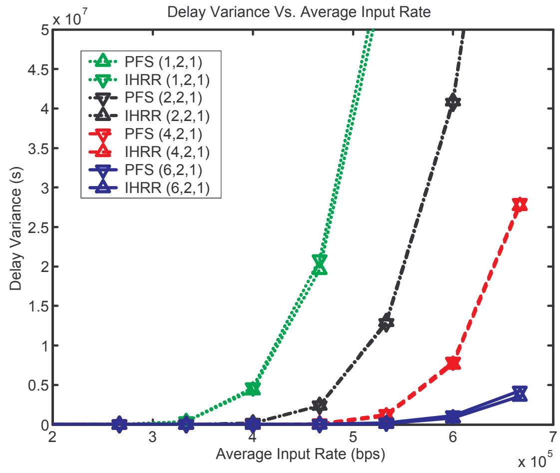


Fig. 4. Average delay for PFS and IHRR scheduling policies

permit the scheduler to keep a better track of the buffer state. This is shown in figure 5, where we observe that a shorter scheduling period is particularly beneficial at high average input rates. Then, for instance, we observe that for an average delay of $300ms$, the average input rate can be increased by $40kbps$ per user, i.e., a total of $640kbps$, when N_{sch} changes from 64 to 1.

V. Conclusion

In this paper the problem of resource allocation in multiuser single-antenna wideband OFDM(A) systems has been considered. It has been shown that if the availability of multiple channels is exploited the average packet delay can be considerably decreased. This is because the use of multiple channels increases rate granularity and introduces randomness that can be exploited specially if the channel changes very slowly in time and resource allocation is performed many times per channel realization. We have compared different resource allocation policies taken from different research communities. One that allocates resources considering the states of the queues only, another that only exploits channel information only and a cross-layer policy that by considering queue and channel state information offers the best average delay results.

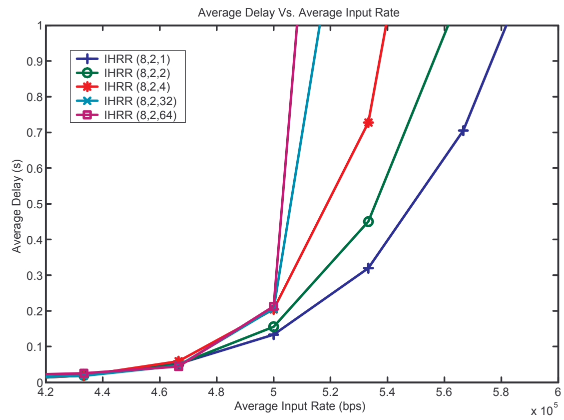


Fig. 5. Average delay for PFS and IHRR scheduling policies when the scheduling period duration is changed.

VI. Acknowledgments

The authors would like to thank prof. G. Caire for his valuable help.

References

[1] “Feasibility Study For OFDM for UTRAN enhancement” 3GPP TR 25.892 V1.1.0 (2004-03)
[2] R. Knopp and P. A. Humblet, “Information capacity and power control in single-cell multiuser communications” *In Proc. IEEE ICC'95, Seattle, WA.*, June 1995.
[3] R. Knopp and P. A. Humblet, “Multiple-Accessing over Frequency-Selective Fading Channels,” *In Proc. IEEE PIMRC'95, Toronto, Ont.*, Sept. 1995.
[4] D. N. C. Tse and S. D. Hanly, “Multiaccess fading channels. I: Polymatroid structure, optimal resource allocation and throughput capacities” *IEEE Transactions on Information Theory,* Vol. 44, pp. 2796-2815, Nov. 1998.
[5] I Toufik, R. Knopp, "Multiuser channel allocation algorithms achieving hard fairness," in proc. IEEE GLOBECOM '04, Vol. 1, 29 Nov.-3 Dec. 2004 pp. 146 - 150.
[6] S. D. Hanly and D. N. C. Tse, “Multiaccess fading channels. II: Delay-limited capacities” *IEEE Transactions on Information Theory,* Vol. 44, pp. 2816-2831, Nov. 1998.
[7] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power Allocation and Routing in Multi-Beam Satellites with Time Varying Channels," IEEE Transactions on Networking, vol. 11, no. 1, pp. 138-152, Feb. 2003.
[8] E.M. Yeh, A.S. Cohen, "Throughput and delay optimal resource allocation in multiaccess fading channels," in proc. IEEE International Symposium on Information Theory, 29 June-4 July 2003 pp. 245 - 245
[9] H. Boche; A. Jorswieck and T. Haustein, “Channel Aware Scheduling for Multiple Antenna Multiple Access Channels” *Proc. 2003 The Thirty-Seventh Asilomar Conference,* Volume: 1 , Nov. 2003 pp. 992 - 996
[10] M. Kobayashi, G. Caire, D. Gesbert, "Opportunistic Beamforming vs. Space-Time Coding in a Queued Downlink," to appear in Proc. of IST Mobile Summit, Dresden, June 2005.
[11] R. Laroia, L. Junyi, S. Rangan, M. Srinivasan, "Enhanced opportunistic beamforming," Vehicular Technology Conference (Fall), 6-9 Oct, 2003, Vol. 3 pp. 1762-1766.
[12] R.G. Gallager, "Information Theory and Reliable Communication", J. Wiley and Sons, New York, 1968.
[13] P. Viswanath; D. N. C. Tse and R. Laroia,“Opportunistic beamforming using dumb antennas,” *IEEE Transactions on Information Theory*, Vol. 48, pp. 1277 -1294, June 2002.
[14] R. Knopp, *Coding and Multiple-Access Over Fading Channels*, PhD Thesis, École Polytechnique Fédérale de Lausanne, 1997.
[15] P. Viswanath; D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," IEEE Transactions on Information Theory, Vol. 49, no. 8, Aug. 2003 pp. 1912-1921
[16] L. Tassiulas, A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity, IEEE Trans. on Information Theory, Vol. 39, Issue 2, March 1993, pp. 466 - 478
[17] Deliverable D4.1 "WIDENS Specifications", IST project WIDENS.