ITC19/ Performance Challenges for Efficient Next Generation Networks 1507-1516
LIANG X.J. and XIN Z.H.(Editors)
V.B. IVERSEN and KUO G.S.(Editors)
Beijing University of Posts and Telecommunications Press

On the Traffic Capacity of Cellular Data Networks

T. Bonald [1,2], A. Proutière [1,2]

[1] France Telecom
Division R&D, 38-40 rue du Général Leclerc, 92794 Issy-les-Moulineaux, France
{thomas.bonald, alexandre.proutiere}@francetelecom.com
[2] École Normale Supérieure
45 rue d'Ulm, Paris, France
{thomas.bonald, alexandre.proutiere}@ens.fr

Abstract. We introduce a notion of traffic capacity useful for dimensioning cellular data networks, where the dynamic nature of traffic is explicitly taken into account. Specifically, the traffic capacity is based on the stable behavior of a dynamic system where the number of active users varies at random as new data transfers are initiated and other cease. We apply this notion to both time-multiplexed and frequency-multiplexed cellular networks, and compare the results to the maximum capacity predicted by the information theory.
Keywords: Cellular data networks, traffic capacity, broadcast channel.

1 Introduction

Since Shannon's pioneer work in 1948 [1], the capacity of wireless systems has mostly been evaluated in static scenarios with a fixed number of users. One of the most famous results of Shannon gives the capacity C of a Gaussian channel:

$$C = W \log_2 \left(1 + \frac{P}{N}\right),$$

where W denotes the bandwidth, P the average transmission power and N the white noise power. This result applies to point-to-point communications only but has since been extended to point-to-multipoint communications (the broadcast channel), multipoint-to-point communications (the multiple access channel) and multipoint-to-multipoint communications (the interference channel) [2]. In such multiuser systems the channel capacity actually corresponds to a *set* of data rates that can be used simultaneously. In particular, the capacity set of the broadcast channel gives the maximum rates at which a base station can simultaneously send data to a fixed set of users as a function of their radio conditions. This result is hardly applicable to network dimensioning, however, where one of the key issues is the trade-off between cell capacity and cell size. The cell capacity must indeed be characterized by a *single* parameter and not by a set of parameters that depends on user locations.

A key observation is that in cellular networks, the number of users sharing the same radio resources is not fixed but varies at random as new data transfers are initiated and other cease. In particular, some cells may well be unstable if not properly dimensioned, in the sense that the number of ongoing data transfers increases continuously [3]. This calls for a new notion of capacity, corresponding to the maximum traffic that can be supported by the network. Specifically, we refer to the capacity of a cell as the maximum traffic intensity compatible with the stability of this cell. This notion has already been applied to time-multiplexed systems such as CDMA 1xEV-DO and UMTS HSDPA systems [4]. Here we apply this notion to both time-multiplexed and frequency-multiplexed systems like OFDM systems [5] and compare the results to the maximum capacity predicted by the information theory.

In the next section, we recall the usual capacity set of Gaussian broadcast channels in a static scenario with a fixed number of users. We then introduce the notion of capacity in a dynamic system with a randomly varying number of users. Finally, we apply the results to the evaluation of the downlink capacity of cellular time-division and frequency-division multiplexed networks and make a brief conclusion.

2 Static Scenario

Consider the broadcast channel with one sender and K receivers. The signal Y_k at receiver k is the sum of the signal X at the sender and the noise Z_k:

$$Y_k = X + Z_k.$$

The signal X has the power constraint $\mathrm{E}[X^2] \leq P$ and the noise Z_k is drawn i.i.d. from a Gaussian distribution with variance N_k.

TDMA channel. For a time-division multiple access, the sender transmits to each receiver one at a time. Let τ_k be the fraction of time the sender transmits to receiver k. The capacity set is the set of data rates $R_1, R_2, \ldots, R_K$ such that:

$$R_k = \tau_k W \log_2\left(1 + \frac{P}{N_k}\right) \tag{1}$$

subject to the constraint:

$$\sum_{k=1}^{K} \tau_k \leq 1.$$

FDMA channel. For a frequency-division multiple access, the sender transmits to each receiver on a specific frequency band. Let σ_k be the fraction of bandwidth allocated to receiver k and let P_k be the associate transmission power. The noise power is assumed to be proportional to the allocated bandwidth so that receiver k has noise power $\sigma_k N_k$. The capacity set is the set of data rates $R_1, R_2, \ldots, R_K$ such that:

$$R_k = \sigma_k W \log_2\left(1 + \frac{P_k}{\sigma_k N_k}\right) \tag{2}$$

subject to the constraints:

$$\sum_{k=1}^{K} \sigma_k \leq 1 \quad \text{and} \quad \sum_{k=1}^{K} P_k \leq P.$$

Remark 1 *Any TDMA rate allocation is achieved in FDMA by allocating bandwidth and transmission power in the same proportions (i.e. $\sigma_k = P_k/P = \tau_k$ for all k). In particular, the TDMA capacity set is included in the FDMA capacity set.*

We have the following useful property.

Proposition 1 *The TDMA (resp. FDMA) capacity set obtained by adding a receiver with noise power N' equal to N_j for some $j \in \{1, 2, \ldots, K\}$ is the set of data rates R' and $R'_1, R'_2, \ldots, R'_K$ such that:*

$$R' + R'_j = R_j, \quad R'_k = R_k \quad k \neq j,$$

for some rates $R_1, R_2, \ldots, R_K$ in the original TDMA (resp. FDMA) capacity set with K receivers.

Proof. The property trivially holds for TDMA channels. For FDMA channels, let σ' be the fraction of bandwidth allocated to the new receiver and let P' be the associate transmission power. By the concavity of the function $t \mapsto \log(1 + t)$, we have:

$$(\sigma' + \sigma_j) \log_2 \left(1 + \frac{P' + P_j}{(\sigma' + \sigma_j)N'}\right) \geq \sigma' \log_2 \left(1 + \frac{P'}{\sigma' N'}\right) + \sigma_j \log_2 \left(1 + \frac{P_j}{\sigma_j N_j}\right),$$

so that the total data rate R_j obtained by merging the radio resources allocated to the new receiver and to receiver j is higher than the sum of the individual rates. In view of Remark 1, any rate allocation R', R'_j such that $R' + R'_j = R_j$ is feasible. □

Broadcast channel. The maximum capacity set as predicted by the information theory requires the joint coding of user data flows [2]. We assume without loss of generality that $N_1 \leq N_2 \leq \ldots \leq N_K$. The broadcast channel capacity set is the set of data rates $R_1, R_2, \ldots, R_K$ such that:

$$R_k = W \log_2 \left(1 + \frac{P_k}{N_k + \sum_{j<k} P_j}\right), \tag{3}$$

subject to the constraint:

$$\sum_{k=1}^{K} P_k \leq P.$$

We have the analogue of Proposition 1.

Proposition 2 *The broadcast channel capacity set obtained by adding a receiver with noise power N' equal to N_j for some $j \in \{1, 2, \ldots, K\}$ is the set of data rates R' and $R'_1, R'_2, \ldots, R'_K$ such that:*

$$R' + R'_j = R_j, \quad R'_k = R_k \quad k \neq j,$$

for some rates $R_1, R_2, \ldots, R_K$ in the original broadcast channel capacity set with K receivers.

Proof. The result simply follows from the fact that, denoting by P' the power allocated to the new receiver,

$$\begin{aligned}
R' + R'_j &= W \log_2 \left(1 + \frac{P'}{N' + \sum_{k<j} P_k}\right) + W \log_2 \left(1 + \frac{P_j}{N_j + P' + \sum_{k<j} P_k}\right) \\
&= W \log_2 \left(\frac{N_j + P' + \sum_{k<j} P_k}{N_j + \sum_{k<j} P_k}\right) + W \log_2 \left(\frac{N_j + P' + \sum_{k\le j} P_k}{N_j + P' + \sum_{k<j} P_k}\right) \\
&= W \log_2 \left(\frac{N_j + P' + \sum_{k\le j} P_k}{N_j + \sum_{k<j} P_k}\right) \\
&= W \log_2 \left(1 + \frac{P' + P_j}{N_j + \sum_{k<j} P_k}\right).
\end{aligned}$$

Thus the capacity set is the same as that obtained if the new receiver and receiver j are considered as a single receiver with allocated power $P' + P_j$. □

In the rest of the paper, we define K receiver classes. The noise power of class-k receivers is equal to N_k. We denote by R_k the total rate of class-k receivers. In view of Propositions 1 and 2, the set of feasible total rates coincides with the capacity sets (1), (2) and (3) for the TDMA channel, the FDMA channel and the broadcast channel, respectively. Figure 1 gives an example of such capacity sets for $K = 2$ classes. In all figures, data rates are expressed in bit/s/Hz, equivalently for a unit bandwidth W. As expected, the TDMA capacity set is included in the FDMA capacity set (cf. Remark 1) and both are included in the broadcast channel capacity set. The three capacity sets coincide if and only if $N_1 = N_2$.

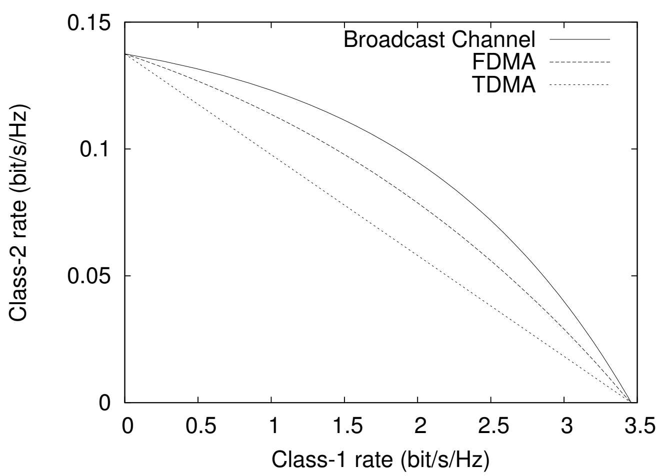

Figure 1: Capacity sets in a static scenario with $K = 2$ classes ($P/N_1 = 10$dB, $P/N_2 = -10$dB).

3 Dynamic Scenario

We now consider a dynamic scenario where the number of class-k receivers varies at random as new data transfers are initiated and other cease. We do not make any specific traffic assumptions except that the arrival times of new transfers and the corresponding transfer sizes (in bits) form a stationary and ergodic marked point process [6]. We denote by ρ_k the *traffic intensity* of class-k receivers (in bit/s). Thus the sender must transmit an average of ρ_k bits to class-k receivers every second. In particular, a necessary condition for the dynamic system to be stable, in the sense that the total number of data transfers remains finite, is that the traffic intensities belong to the capacity set.

We have the following key results.

Proposition 3 *If the traffic intensities belong to the interior of the TDMA capacity set, i.e.*

$$\rho_k = \tau_k W \log_2\left(1 + \frac{P}{N_k}\right) \tag{4}$$

for some $\tau_1, \tau_2, \ldots, \tau_K$ *such that:*

$$\sum_{k=1}^{K} \tau_k < 1,$$

there exists a rate allocation such that the dynamic TDMA channel is stable.

Proposition 4 *If the traffic intensities belong to the interior of the FDMA capacity set, i.e.*

$$\rho_k = \sigma_k W \log_2\left(1 + \frac{P_k}{\sigma_k N_k}\right) \tag{5}$$

for some $\sigma_1, \sigma_2, \ldots, \sigma_K$ *and* $P_1, P_2, \ldots, P_K$ *such that:*

$$\sum_{k=1}^{K} \sigma_k < 1 \text{ and } \sum_{k=1}^{K} P_k \le P \quad \text{or} \quad \sum_{k=1}^{K} \sigma_k \le 1 \text{ and } \sum_{k=1}^{K} P_k < P,$$

there exists a rate allocation such that the dynamic FDMA channel is stable.

Proposition 5 *If the traffic intensities belong to the interior of the broadcast channel capacity set, i.e.*

$$\rho_k = W \log_2\left(1 + \frac{P_k}{N_k + \sum_{j<k} P_j}\right) \tag{6}$$

for some $P_1, P_2, \ldots, P_K$ *such that:*

$$\sum_{k=1}^{K} P_k < P,$$

there exists a rate allocation such that the dynamic broadcast channel is stable.

Proof of Propositions 3, 4, 5. If the traffic intensities belong to the interior of the capacity set, there exist rates $R_1, R_2, \ldots, R_K$ in the capacity set such that $\rho_k < R_k$ for all classes k. The system then behaves as a set of K independent stable queues. □

We can now define the traffic capacity as follows. Let

$$\rho = \sum_{k=1}^{K} \rho_k$$

be the overall traffic intensity and $\alpha_k = \rho_k/\rho$ be the fraction of traffic intensity due to class-k receivers. Assume that the traffic distribution among classes $\alpha_1, \alpha_2, \ldots, \alpha_K$ is fixed but that the overall traffic intensity ρ may vary. We refer to the traffic capacity C as the maximum traffic intensity ρ such that the system is stable for some rate allocation.

TDMA channel. In view of Proposition 3, the TDMA traffic capacity satisfies:

$$\alpha_k C = \tau_k W \log_2\left(1 + \frac{P}{N_k}\right)$$

with

$$\sum_{k=1}^{K} \tau_K = 1,$$

from which we deduce:

$$C = \left(\sum_{k=1}^{K} \frac{\alpha_k}{W \log_2\left(1 + \frac{P}{N_k}\right)}\right)^{-1}.$$

FDMA channel. There is no such a closed-form expression for the FDMA traffic capacity since the capacity set then depends both on the bandwidth allocation *and* the power allocation. We numerically evaluate the FDMA traffic capacity in the rest of the paper.

Broadcast channel. For the broadcast channel, it follows form Proposition 5 that there exists a unique power allocation $P_1, P_2, \ldots, P_K$ such that:

$$\alpha_1 C = W \log_2\left(1 + \frac{P_1}{N_1}\right), \quad (7)$$

$$\alpha_2 C = W \log_2\left(1 + \frac{P_2}{N_2 + P_1}\right), \quad (8)$$

$$\ldots$$

$$\alpha_K C = W \log_2\left(1 + \frac{P_K}{N_K + P_1 + P_2 + \ldots + P_{K-1}}\right) \quad (9)$$

and

$$\sum_{k=1}^{K} P_K = P.$$

By the monotonicity of C with respect to P, the traffic capacity can be evaluated using the following simple iterative algorithm:

1. Start from:
$$C_{\min} = 0, \quad C_{\max} = W \log_2\left(1 + \frac{P}{N_1}\right).$$
2. Deduce from (7)–(9) the power allocation $P_1, P_2, \ldots, P_K$ for
$$C = \frac{1}{2}(C_{\min} + C_{\max}).$$
3. Choose $C_{\min} = C$ if $P_1 + P_2 + \ldots + P_K < P$, $C_{\max} = C$ otherwise.
4. Repeat steps 2 and 3 until the difference $C_{\max} - C_{\min}$ is sufficiently small.

Figure 2 gives the traffic capacities of the TDMA channel, the FDMA channel and the broadcast channel for the two classes considered in Figure 1. The capacity is plotted against α_1, the fraction of traffic intensity due to class-1 receivers. Note that the traffic capacities are very close although the corresponding capacity sets differ significantly in a static scenario (refer to Figure 1). This is simply due to the fact that most load is generated by class-2 receivers (which have lower data rates) and shows how misleading conclusions may be drawn from the analysis of a static scenario.

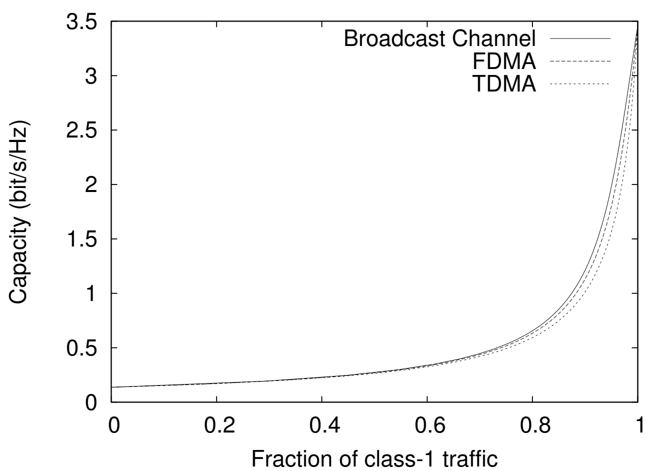

Figure 2: Traffic capacity in a dynamic scenario with $K = 2$ classes (P/N_1 =10dB, P/N_2 =-10dB).

4 Application to Cellular Networks

Consider a base station serving users in a given cell of a cellular network. We assume that the communication channel from the base station to the users may be represented as a Gaussian broadcast channel with power constraint P. We use previous results to evaluate the maximum downlink capacity of the base station as follows.

We define an arbitrarily large number of user classes K such that class-k users experience approximately the same radio conditions. We denote by Γ_k the path loss from the base station to class-k users and by I_k the interference power due to other base stations. When the signal X is sent to a class-k receiver, the received signal is:

$$Y_k = \Gamma_k X + Z_k,$$

where Z_k is drawn i.i.d. from a Gaussian distribution with variance $N + I_k$, where N denotes the thermal noise power. The model then corresponds to a Gaussian broadcast channel with equivalent noise power:

$$N_k = \frac{N + I_k}{\Gamma_k}.$$

For illustrative purposes, we consider infinite linear and hexagonal networks as shown in Figure 3. We take $P = 40$ dB, $N = -100$ dB, and assume the path loss between any two points is a function Γ of their distance d only:

$$\Gamma(d) = 130 + 35 \log_{10}(d) \text{ dB},$$

where d is expressed in km. Note that this corresponds to a path loss exponent of 3.5. The interference term is the sum of the powers received from the other base stations. Figure 4 gives the cell capacity with respect to the cell radius when the traffic intensity is uniformly distributed in the cell. For linear networks, the cell radius r refers to the maximum distance at which a user is served; for hexagonal networks, the cell radius r refers to the radius of a disk with the same area as the hexagon.

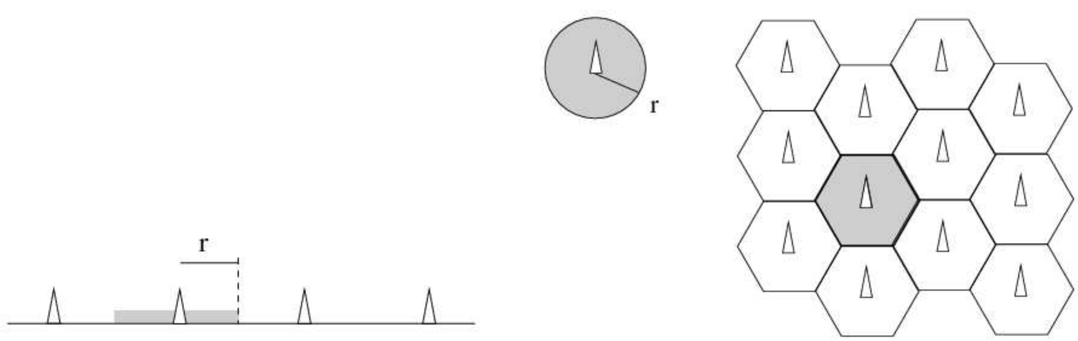

Figure 3: Linear and hexagonal networks.

As expected, the cell capacity is a decreasing function of the cell radius (because of the path loss). The maximum capacity is obtained when the cell radius r tends to zero, i.e., for infinitely dense networks (in view of Figure 4, it is in fact sufficient for r to be less than 500 m, say). The cell capacity is then limited by interference only. In particular, it depends on the radio model through the network topology and the path loss exponent only. The table below shows how the path loss exponent impacts the maximum cell capacity.

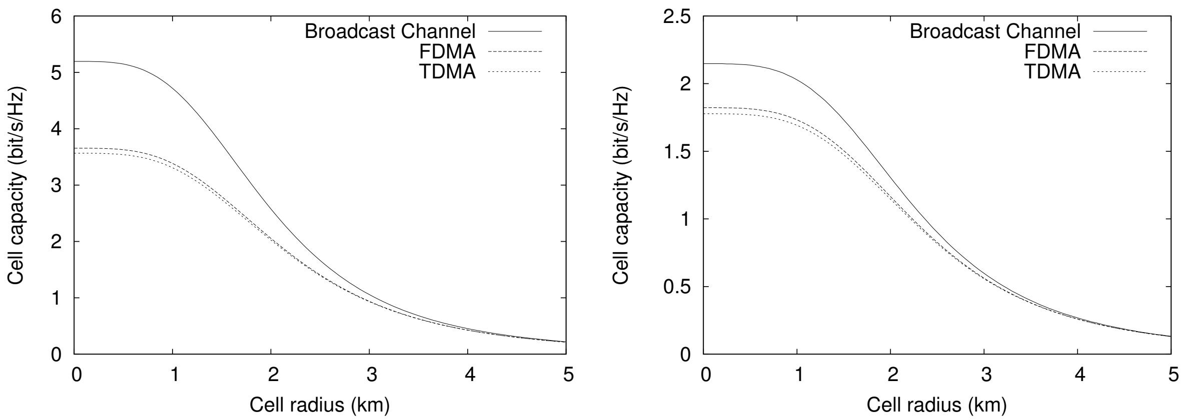

Figure 4: Cell capacity for linear networks (left) and hexagonal networks (right).

The difference between TDMA and FDMA is relatively small while the broadcast channel capacity is much higher than both TDMA and FDMA capacities in all cases. This suggests that significant capacity gains may be achieved my more complex multiplexing schemes than TDMA and FDMA, e.g. by a joint coding the user data flows.

Path loss exponent	3	3.5	4
Broadcast Channel	4.34	5.17	5.96
FDMA	3.08	3.61	4.02
TDMA	3.17	3.51	3.90

Path loss exponent	3	3.5	4
Broadcast Channel	1.54	2.15	2.71
FDMA	1.36	1.82	2.21
TDMA	1.34	1.78	2.15

Table 1: Maximum cell capacity (in bit/s/Hz) for linear networks (left) and hexagonal networks (right).

For large cells, on the other hand, TDMA and FDMA achieve the maximum broadcast channel traffic capacity. Most cell load is indeed concentrated at the cell edge in this case. Users at the cell edge experience the same (bad) radio conditions and suffer from noise rather than interference. In such a homogeneous setting, both TDMA and FDMA have the same capacity set as the broadcast channel, independently of the number of users (cf. the static scenario of Section 2). No capacity gain can be achieved by more complex multiplexing schemes than TDMA and FDMA.

5 Conclusion

We have introduced a new notion of capacity for Gaussian broadcast channels and used this notion to evaluate the maximum downlink capacity of cellular networks. For a path loss exponent of 3.5 and a bandwidth of 3.84 MHz for instance, corresponding to the UMTS band, our results show that the cell capacity cannot exceed 19 Mbit/s for linear networks (around 13 Mbis/s in TDMA or FDMA), 7 Mbit/s for hexagonal networks (around 6 Mbis/s in TDMA or FDMA). Though we have considered symmetric network topologies only, the results readily apply to more realistic, heterogeneous networks.

We have observed a significant difference between the broadcast channel capacity and the TDMA and FDMA capacities, especially for dense networks, which suggests that time-multiplexed systems such as CDMA 1xEV-DO and UMTS HSDPA systems and frequency-multiplexed systems will never achieve the maximum capacity promised by the information theory, whatever the future improvements of these systems. Such systems may take advantage of fast fading by means of opportunistic scheduling, however, a phenomenon that has not been taken into account in the present paper. Thus it is necessary to consider random fading channels to fully assess the performance of time-multiplexed and frequency-multiplexed systems compared to the broadcast channel.

Though the broadcast channel provides a theoretical limit for the downlink capacity of an individual cell, the cell capacity may be further increased by considering the whole network as a multipoint-to-multipoint communication channel. We are currently investigating the capacity gain that may result from the coordination of base stations in time-multiplexed cellular networks [7]. The use of MIMO technics is another potential source of capacity gain [8]. In all cases, the benefits of a given technology should be assessed in the realistic, dynamic scenario described in the present paper.

References

1. C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal, 1948.
2. T. M. Cover and J. A. Thomas, Elements of information theory, Wiley, 1991.
3. S. Ben Fredj, T. Bonald, A. Proutière, G. Regnié and J. Roberts, Statistical bandwidth sharing: A study of congestion at flow level, Proc. of ACM SIGCOMM, 2001.
4. T. Bonald and A. Proutière, Downlink data channels: User performance and cell dimensioning, Proc. of ACM MOBICOM, 2003.
5. R. van Nee, R. Prasad, OFDM for wireless multimedia communications, Artech House, 2000.
6. F. Baccelli and P. Brémaud, Elements of queueing theory: Palm-Martingale calculus and stochastic recurrences, Springer-Verlag, 1994.
7. T. Bonald, S. Borst and A. Proutière, Inter-cell scheduling in wireless data networks, to appear in European Wireless, 2005.
8. A. Amraoui, G. Kramer and S. Shamai, Coding for the MIMO Broadcast Channel, Proc. of IEEE International Symposium on Information Theory, 2003.