MULTIPLE-ANTENNA SYSTEMS AND WIRELESS NETWORKS

Babak Hassibi

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125
hassibi@systems.caltech.edu, (626)395-4810

ABSTRACT

The use of multiple antennas, at the transmitter, receiver or both, promises both higher data rates as well as greater reliability in point-to-point communication channels. While this is now well understood, it is only recently that the impact of multiple antenna systems in a wireless network is being systematically assessed. In this paper, using the example of of the downlink of a cellular system, we review some of the issues related to throughput, delay and scheduling that arise when mutliple antennas are deployed at the base station or at the user terminals.

1. INTRODUCTION

Multiple antenna communication systems have recently generated a great deal of interest in industry and academia since they promise to greatly increase the rate and reliability of point-to-point wireless communication links. In fact, the whole field goes now by the acronym MIMO (multiple-input multiple-output). While this is now well recognized, it is only very recently that researchers have begun to systematically assess the impact of multiple antenna systems on the multi-user setting of a wireless network. In some respects this is surprising, since multiple-antenna systems were, in fact, first proposed for use in multi-user settings, where the idea was referred to as SDMA (space division multiple access) and was based on using array signal processing techniques to distinguish between users in different spatial locations. So in some sense the field has now come full circle.

In this paper we will focus on the downlink of a cellular system as a means to review some of the issues that come up when multiple antennas are employed in a wireless network. We will, in particular, emphasize some key differences with the familiar point-to-point case, thereby demonstrating that many of the results obtained there do not carry over to the network setting. We will study the total aggregate throughput of the network as a function of the number of transmit antennas at the base station, the number of antennas employed at each user terminal, the number of users in the cell, and the total transmit power. It will turn out that the results are highly sensitive to the amount of channel information available at the base station. We will then focus on a delay analysis of the system, in the sense of determining *how long* it takes for all users in the cell to receiver information. Finally, we shall touch upon the issue of scheduling, i.e., what users should be transmitted to at each channel use, which is an important question in any multi-user wireless system.

2. BASIC MODEL

Consider a cellular system with M antennas at the transmitter and n users, each with N receive antennas.[1] The channels to each user are assumed to be block fading with a coherence interval of T; in other words, the channels remain constant for T channel uses after which they change to independent values.[2] Furthermore, over different users the fading is assumed to be independent

Thus, during any coherence interval, the signal to the i-th user, $i = 1, 2, \ldots, n$, can be written as

$$x_i(t) = \sqrt{P}H_i s(t) + w_i(t), \quad t = 1, \ldots, T \qquad (1)$$

where $H_i \in \mathcal{C}^{1 \times M}$ is constant during the coherence interval and has iid $\mathcal{CN}(0, 1)$ entries, $w_i(t)$ is additive white noise with distribution $\mathcal{CN}(0, 1)$, $s(t) \in \mathcal{C}^{M \times 1}$ is the transmit symbol satisfying $E\|s(t)\|^2 = 1$ and P is the total transmit power.

2.1. The Point-to-Point Case

It turns out that in a point-to-point link with M transmit antennas and N receive antennas, assuming that the envi-

---

[1]Although we will generally assume $N \geq 1$, we shall often focus on $N = 1$ which is very reasonable from a practical point of view.

[2]We should remark that, although the assumption of a constant channel for T channel uses is critical, the requirement that the channels vary independently from one coherence interval to the next is not.

ronment is rich scattering, the channel capacity is not very sensitive to channel knowledge at the transmitter and or receiver. In fact,

1. If the transmitter and receiver know the channel:

$$C = \min(M, N) \log P + O(1), \qquad (2)$$

where the $O(1)$ term is with respect to increasing P.

2. If the receiver only knows the channel [1, 2]:

$$C = \min(M, N) \log P + O(1). \qquad (3)$$

3. If neither the transmitter or receiver knows the channel [3, 4]

$$C = \min(M, N) \left(1 - \frac{\min(M, N)}{T} \right) \log P + O(1), \qquad (4)$$

where, as before, T is the coherence interval of the channel.

2.2. The Multi-User Case

In point-to-point multi-antenna systems the throughput scaling is often equivalent to the "multiplexing gain" defined as $\lim_{P \to \infty} \frac{C}{\log P}$. However, in the multi-user setting two different throughput scaling laws can be envisioned.

Theorem 1 (Large Power Regime) *Consider the downlink cellular system of Section 2 and assume that the channels to all n users are known at the transmitter. Then for fixed M and n, we have*

$$\lim_{P \to \infty} \frac{C_{sum}}{\log P} = M, \qquad (5)$$

where C_{sum} refers to the maximum possible sum of the rates to all n users.

Theorem 2 (Large Number of Users Regime) *[5] Consider the downlink cellular system of Section 2 and assume that the channels to all n users are known at the transmitter. Then for fixed M and P, we have*

$$\lim_{n \to \infty} \frac{C_{sum}}{\log \log n} = M, \qquad (6)$$

where C_{sum} refers to the maximum possible sum of the rates to all n users.

These are clearly two very different regimes. We argue that, from a practical perspective, the latter regime may be more interesting. There are three reasons that come to mind.

1. Many practical systems operate with a large number of per-cell users (n could be in the hundreds, whereas M may be no more than two, three or four).

2. Significant rates can be obtained even at low to moderate transmit powers P.

3. The first gain requires channel knowledge with very high fidelity at the transmitter (indeed a fidelity that grows with the transmit power) [6], whereas, as we shall see below, the latter requires very little CSI (channel state information).

Another interesting fact is that the number of receive antennas N plays very little role in the sum-rate capacity of the downlink of the cellular system (a result which is in stark contrast to the point-to-point case). In fact, we can give a tighter result on the sum-rate, which explicitly involves N, in the next theorem.

Theorem 3 *[7] Consider the setting of Theorems 1 and 2. For fixed P and M*

$$C_{sum} = M \log \log nN + M \log \frac{P}{M} + o(1), \qquad (7)$$

where C_{sum} refers to the maximum possible sum of the rates to all n users and $o(1)$ is with respect to growing n.

Thus, the capacity grows only doubly logarithmic in N.

Finally, the next result shows what happens when the transmitter has no knowledge of the channels to the different users.

Theorem 4 *[7] Consider the setting of Theorems 1 and 2, however, assume that the transmitter has no knowledge of the channels to the n users.*

1. *For fixed M and n:*

$$\lim_{P \to \infty} \frac{C_{sum}}{\log P} = \min(M, N). \qquad (8)$$

2. *For fixed P and M*

$$\lim_{n \to \infty} \frac{C_{sum}}{\log \log n} = 0. \qquad (9)$$

Thus, unlike the point-to-point case, lack of channel knowledge at the transmitter signicantly reduces the sum rate of the system.

2.3. Opportunistic Beamforming

We now briefly describe a simple scheme that achieves most of the broadcast sum-rate in the large n regime, yet requires very little CSI at the transmitter. The idea is based on transmitting M random beams and is described in [8] (a similar construction, albeit with little analysis, appears in the appendix of [9]).

Basically, during any coherence interval the transmitter chooses M random orthonormal vectors $\phi_m \in \mathcal{C}^{M \times 1}$ according to an isotropic distribution and then transmits the vector

$$s(t) = \sum_{m=1}^{M} \phi_m s_m(t), \qquad (10)$$

where each $s_m(t)$ is a scalar signal intended for one of the users. Assuming the users know their own channel coefficients (a much more reasonable assumption than the transmitter knowing *all* the channel gains to the different users), each user can compute its signal-to-interference-plus-noise-ratio (SINR) for every beam as

$$\text{SINR}_{m,i} = \frac{|H_i \phi_m|^2}{\frac{M}{P} + \sum_{l \neq m} |H_i \phi_l|^2}. \qquad (11)$$

If each user (or, in fact, only those users who have favorable SINRs) feeds back its best SINR and corresponding beam index to the transmitter, the transmitter can assign each beam to the user that has the best SINR for that beam. (This is the jist of the idea—for more details see [8].)

For this scheme the following result can be shown.

Theorem 5 *[7] Consider the setting of Theorem 3. Then if C_{bo} is the sum rate obtained by opportunistic beamforming*

$$\lim_{n \to \infty} (C_{ob} - C_{sum}) = 0. \qquad (12)$$

Thus, asymptotically, opportunistic beamforming has no loss compared to a scheme that has full CSI at the transmitter.

3. DELAY ANALYSIS

The sum rate capacities computed in the previous section are generally obtained by *opportunistic transmission*, i.e., the transmitter transmits to the user(s) with the best channel conditions. While this is optimal for throughput, it will inevitably lead to unfairness in the system as users with poor channel conditions may incur large delays.

To formalize the idea and to obtain some explicit results, assume that transmission over each coherence interval is packetized (i.e., consists of a single packet). Thus, during any coherence interval only M packets are transmitted to M users (one packet per user). We will define the delay of the system as $D_{m,n}$, the number of channel uses it takes until all n users have successfully received m packets. Clearly, one delay optimal scheme would be round-robin scheduling which clearly incurs a delay of $D_{m,n} = \frac{mn}{M}$. Now the throughput optimal scheme will inevitably incur a larger delay and the following result characterizes the resulting delay penalty.

Theorem 6 *[10] Consider the downlink of a cellular system as in Theorems 1 and 2, and assume that opportunistic beamforming is employed. Then*

1. *For n fixed and $m \to \infty$,*

$$E(D_{m,n}) = \frac{mn}{M} + O(n \log m), \qquad (13)$$

which clearly states that the system is long-term fair.

2. *For m fixed and $n \to \infty$, we have*

$$E(D_{m,n}) = \frac{n \log n}{M} + O(n \log \log n). \qquad (14)$$

3. *For $m = \log n$ and $n \to \infty$, we have*

$$E(D_{m,n}) = 3.126 \frac{mn}{M} + O(n \log \log n). \qquad (15)$$

4. *For $m = (\log n)^r$ where $r > 1$ is fixed and $n \to \infty$,*

$$E(D_{m,n}) = \frac{mn}{M} + O(n \log n), \qquad (16)$$

which demonstrates how long it takes for the system to become fair.

For a more comprehensice discussion of the implications of these results see [10].

4. DIFFERENTIATED RATE SCHEDULING

In homogenous networks, the sum-rate point is a symmetrical point on the boundary of the capacity region and so treats all the users equally. In systems which are provisioned to provide differentiated services to different users, the transmitter has to give different services (or rates) to different subsets of receivers, and yet at the same time, maximize the throughput (see e.g., [11] for a discussion of the SISO case). Giving differentiated rates to users clearly means operating at non-symmetrical boundary points of the capacity region.

Therefore an important problem is to come up with *simple* schemes, that require very *little* CSI, give *differentiated rates* to the users, and that operate *close* to the boundary of the capacity region. In particular, it will be interesting to quantify the rate loss, compared to the sum rate, for various differentiated rate schemes.

Problem 1 *Consider the cellular system of Section 2. Assume that the n users are divided into K groups, each with $\alpha_k n$ users ($\sum_{k=1}^{K} \alpha_k = 1$) and each require a different rate. Let R_i denote the rate to i-th user and R^k denote the rate to all users in group k. Then construct a transmission scheme such that*

$$\max \sum_{i=1}^{n} R_i$$

$$\text{subject to } \frac{R^k}{R^K} = \beta_k, \quad k = 1, \ldots, K-1$$

4.1. Time-Division Opportunistic (TO) Beamforming

Assume we divide each coherence interval into K slots of duration t_k each, $k = 1, \ldots, K$. During the k-th subinterval the transmitter performs opportunistic beamforming to *only* the $\alpha_k n$ users in the k-th group. It is not hard to convince oneself that to satisfy the rational rate constraints, we must have

$$\frac{t_k}{T} = \frac{\alpha_k \beta_k}{\sum_{l=1}^{K} \alpha_l \beta_l}, \quad k = 1, \ldots, K \qquad (17)$$

We now have the following result.

Theorem 7 *[7] Consider the cellular system of Section 2. Let M, P, and the α_k and β_k's be fixed and let the subintervals be chosen as (17). Then the rational rate constraints are met and*

$$\lim_{n \to \infty} (C_{sum} - C_{tdob}) = 0, \qquad (18)$$

where C_{tdob} represents the sum-rate for the time-division opportunistic scheme.

Thus, asymptotically in n, TO beam-forming achieves the unconstrained sum-rate capacity while also satisfying the rational rate constraints.

4.2. Weighted Opportunistic (WO) Beamforming

Here we weigh the SINR of each user according to its group by a weight μ_k, $k = 1, \ldots, K$. Then during each coherence interval, the transmitter assigns the M random beams to the M users that have the largest *weighted* SINR.

In the WO beamforming scheme we first must determine the weights such that the rational rate constraints are met. Here, unlike the TO case, the answer is not trivial.

Theorem 8 *[7] Consider the cellular system of Section 2. Consider the WO beamforming scheme with*

$$\mu_k = 1 + \frac{\log \beta_k}{\log n + (M - 1)(1 - \log \log n)}. \qquad (19)$$

Assuming, M, P, α_k's and β_k's are fixed, we have

$$\lim_{n \to \infty} \frac{R^k}{R^K} = \beta_k, \quad k = 1, \ldots, K. \qquad (20)$$

The final result shows that, as in the case of TO beamforming, WO beamforming achieves the sum-rate of the unconstrained broadcast channel as $n \to \infty$.

Theorem 9 *[7] Consider the setting of Theorem 8 and let C_{wob} denote the sum of the rates obtained by the weighted opportunistic beamforming scheme. Then.*

$$\lim_{n \to \infty} (C_{sum} - C_{wob}) = 0. \qquad (21)$$

5. REFERENCES

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[2] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[3] L. Zheng and D.N.C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel", *IEEE Trans Info Thy*, pages 359-383, Feb 2002.

[4] B. Hassibi and T.L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form", *IEEE Trans Info Thy*, pages 1473-1484, June 2002.

[5] M. Sharif and B. Hassibi, "Scaling laws of sum rate using time-sharing, DPC, and beamforming for MIMO broadcast channels", *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, 27 June-2 July 2004 Page(s):175

[6] A. Lapidoth, S. Shamai and M.A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information", *Proceedings of the 43rd Allerton Conference on Communication, Control and Computing*, September 2005

[7] A. Vakili, A.F. Dana, M. Sharif and B. Hassibi, "Differentiated rate scheduling for MIMO broadcast channels", *Proceedings of the 43rd Allerton Conference on Communication, Control and Computing*, September 2005

[8] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information", *Information Theory, IEEE Transactions on*, Volume 51, Issue 2, Feb. 2005 Page(s):506 - 522

[9] P. Viswanath, D.N.C. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas", *Information Theory, IEEE Transactions on* Volume 48, Issue 6, June 2002 Page(s):1277 - 1294

[10] M. Sharif and B. Hassibi, "A delay analysis for opportunistic transmission in fading broadcast channels", *24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2005*, vol. 4, pages 2720-2730.

[11] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates, *IEEE Trans. Info. Theory*, vol. 49, no. 11, pp. 2895–2909, 2003