Throughput and Delay Analysis of MIMO Broadcast Channels with Partial CSI

Babak Hassibi and Masoud Sharif

Department of Electrical Engineering

California Institute of Technology, Pasadena, CA 91125

Intel Workshop on Wireless and Smart Antennas

Santa Clara, CA, February 27, 2004

Most of the research on multi-antenna systems has focused on

- **Channel models and information theory**
  - known channels, unknown channels, spatial correlation, wide-band, etc.
- **Space-time codes**
  - BLAST, orthogonal designs, USTM, differential schemes, etc.
- **Algorithms**
  - nulling and canceling, sphere decoding, iterative detection/decoding, etc.

All of this is from a point-to-point perspective. Only recently has there been a focus on the role of multiple antennas in a wireless network.

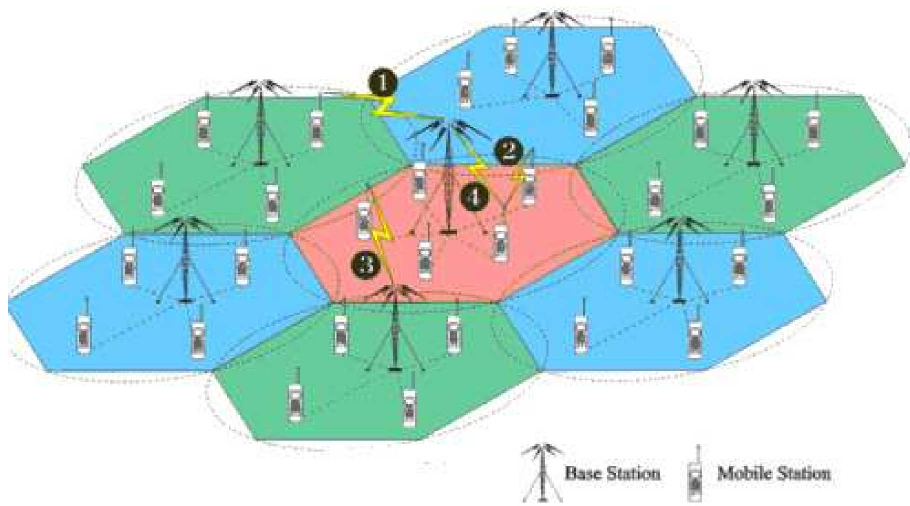
In this talk we will study the role of multiple antennas in broadcast channels, i.e., the *downlink problem*.

Outline

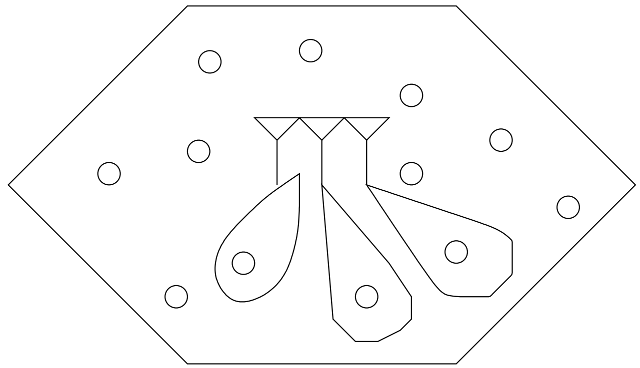
- **Introduction**
- **MIMO broadcast channels**
  - asymptotic sum-rate capacity with and without CSI
  - propose a scheme that requires very little CSI
  - asymptotic analysis
  - fairness issues
  - delay analysis
- **Conclusion**

Introduction to Broadcast Channels

- motivated by down-link scheduling in cellular systems
- to increase network capacity cell sizes are shrinking
- multiple transmit/receive antennas have proven useful in single link communications — what is their role in multi-user networks?



Increasing Cell Capacity with Multiple Antennas



- idea is to exploit *spatial diversity* to transmit simultaneous beams to more than one user (SDMA)
- often requires knowledge of spatial location of different users at base station (also line-of-sight environment)

Assume a system with n users, where the base station has M antennas and each user has N receive antennas.

- The first question we are interested in is how capacity scales with M, N and n.

Multiple Antennas in Single User Links

Consider a single user link with M transmit and N receive antennas. Assume that the environment is rich-scattering.

- Transmitter and receiver know the channel:

$$C = \min(M, N) \log \mathrm{SNR} + O(1)$$

- Only receiver knows the channel (Foschini '97, Telatar '97):

$$C = \min(M, N) \log \mathrm{SNR} + O(1)$$

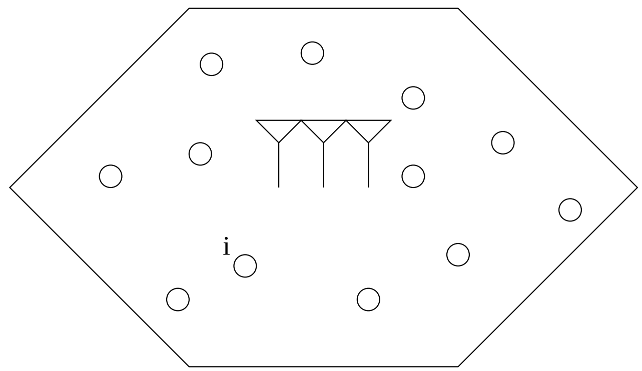
- Neither receive nor transmitter knows the channel (Zheng and Tse '01, Hassibi and Marzetta '01):

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

where T is the coherence interval of the channel

Multiple Antennas in Broadcast Channels

Suppose now we have M transmit antennas at the base station.



Assuming $N = 1$, the channel to the i-th user is given by an M-dimensional vector:

$$h_i = \left[\begin{array}{cccc} h_{i1} & h_{i2} & \ldots & h_{iM} \end{array} \right]$$

The h_i are independent random vectors, whose distribution depends on the fading environment.

For Rayleigh fading the entries of h_i are iid complex Gaussian.

Question: *What is the capacity of such a multi-antenna broadcast channel?*

Answer depends on what the transmitter knows about the channels.

Broadcast Channels with full CSI in Tx/Rx

Capacity with full CSI in the transmitter (Caire and Shamai '03, Viswanath and Tse '03, Goldsmith et al. '03):

$$C^{DP} = E\left\{ \max_{\{P_1,\ldots,P_n,\sum P_i = P\}} \log\det\left(I + \sum_{i=1}^{n} h_i^* P_i h_i \right) \right\}.$$

- Achieved by what is known as *dirty paper coding* (Costa '80).
- We have analyzed this for large number of users, fixed M and N, and shown that

$$\lim_{n\to\infty} \frac{C^{DP}}{M \log\log nN} = 1$$

Dirty paper coding presents two problems:

- It requires full channel knowledge at the transmitter and can be sensitive to channel errors
- It is very computationally intensive (VQ at both Tx/Rx)

Capacity of MIMO BC with no CSI in Tx

What if we assume no CSI in Tx?

- It is straightforward to show that the Gaussian MIMO BC with no CSI in the transmitter is *degraded* no matter whether the receivers have CSI or not (Amraoui et al '03, Sharif and Hassibi '03).
- Therefore, when the users have the same number of antennas, superposition coding is the same as time sharing.
- Furthermore, assuming $N = 1$, the sum rate capacity is

  $$C = \log M + O(1)$$

  independent of n! (Assuming a fixed transmit power per antenna.)

Thus, when $N = 1$, lack of knowledge of the channel coefficients brings us down from $M \log \log n$ to $\log M$.

- full channel-state-information (CSI) unreasonable
- multiple antennas do not buy us anything if no CSI available

What to do?

- is there any critical side information in the transmitter that can fill this gap?

Send Random Beams

Choose M random orthonormal vectors ϕ_m, $m = 1, \ldots, M$ (according to an isotropic distribution). At time t, the m-th vector is multiplied by the signal $s_m(t)$ ao that the transmitted signal is

$$S(t) = \sum_{m=1}^{M} \phi_m s_m(t), \qquad t = 1, \ldots, T$$

where T is less than the coherence interval of the channel.

After T channel uses we independently choose another isotropic set of orthonormal vectors $\{\phi_m\}$, and so on.

In other words, we are transmitting M random beams.

Each receiver $i = 1, \ldots, n$ therefore can compute the following M SINRs

$$\text{SINR}_{im} = \frac{|h_i \phi_m|^2}{1/\text{SNR} + \sum_{n \neq m} |h_i \phi_n|^2}$$

Of course, on average the SINR is (roughly) $\frac{1}{1/\text{SNR}+M-1}$, and so if we randomly assign beams to users we get

$$C \approx M \log \left(1 + \frac{1}{1/\text{SNR} + M - 1} \right) < \frac{M}{1/\text{SNR} + M - 1},$$

which is pretty lousy.

So what is the point?

Exploit Multi-User Diversity

Suppose now each user (or, in fact, only those who get favorable SINRs) **feeds back to the transmitter its best SINR**.

Rather than randomly assign the beams, the transmitter assigns signal s_m to the user with the best SINR for that signal. Therefore

$$C = E \sum_{m=1}^{M} \log \left(1 + \max_{i=1,\dots,n} \frac{|h_i \phi_m|^2}{1/\text{SNR} + \sum_{n \neq m} |h_i \phi_n|^2} \right)$$

Due to the symmetry of all the random variables involved:

$$C = ME \log \left(1 + \max_{i=1,\dots,n} \frac{|h_i \phi_1|^2}{1/\text{SNR} + \sum_{n \neq 1} |h_i \phi_n|^2} \right)$$

Note that the random variables

$$\frac{|h_i \phi_1|^2}{1/\text{SNR} + \sum_{n \neq 1} |h_i \phi_n|^2}$$

are ratios of independent Gaussians, and themselves are independent.

Maximum of n IID Random Variables

Theorem 1 (Frechet, 1927, v. Mises 1947, Uzgoren 1954) *Let x_i, $i = 1, \ldots, n$ be iid random variables with distribution $p(x)$. Then if,*

$$\lim_{x \to \infty} -\frac{p(x)}{p'(x)} = c > 0,$$

then

$$Prob\left(\frac{|x_{\max} - \frac{1}{c}\log n|}{\log\log n} > \epsilon\right) < \frac{\delta}{(\log n)^{\alpha}}.$$

In our case, the condition of the theorem is met with $c = 1/\mathrm{SNR}$.

Therefore, for M fixed and $n \to \infty$

$$C = M \log\log n + O(\log\log\log n)$$

Thus, multi-user diversity buys us a lot!

Some Remarks

- Using partial CSI, our scheme achieves $M \log \log n$, the same asymptotic capacity as having full CSI.
- If we have log-normal shadowing

$$C = M\sqrt{\log n} + O(\sqrt{\log \log n}).$$

- Compared to the full CSI case where each user had to feedback $2M$ complex numbers, here each user need only feed back its *best* SINR and the corresponding index.
  - In fact, not all users need to do this. If only users whose best SINR exceeds a threshold, η, say, feed it back then

$$C \geq \left(1 - F^{Mn}(\eta)\right) M \log \log n + O(\log \log \log n),$$

    where $F(\cdot)$ is the CDF of the SINR. This reduces the feedback by a factor of $\frac{1}{1-F(\eta)}$.

How Large can M Be?

Our result required that M be fixed and $n \to \infty$.

In this case, $\mathrm{SINR}_{\max}$ was obtained when $\sum_{j \neq m} |h_j \phi_m|^2 \approx 0$ and $|h_i \phi_m|^2 \approx \log n$.

However, in practice, we will have a large (but finite n) and so it is useful to know how large M can be to retain a linear increase in capacity.

Result: Let $M = \alpha \log n$. Then as $n \to \infty$, we have $\mathrm{SINR}_{\max} \to 1/\alpha$ and

$$C = M \log\left(1 + \frac{1}{\alpha}\right) + O(1)$$

Moreover, if $\frac{M}{\log n} \to \infty$, then $\frac{C}{M} \to 0$.

Sum Rate Throughput of Random Beamforming

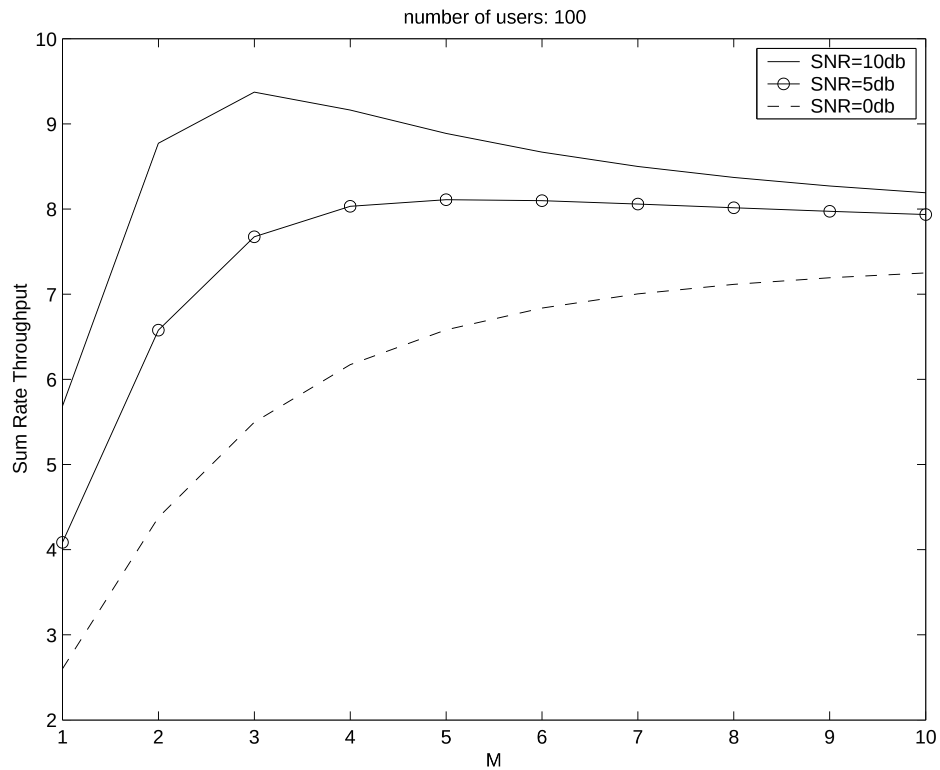


Figure 1: Sum rate throughput for a system with $n = 100$ users and for different number of transmit antennas

What if We Add Multiple Receive Antennas?

Suppose we have $N > 1$ receive antennas. There are three possibilities:

- assign one beam to user (feed back capacity)

$$C = M \log\log n + O(\log\log\log n)$$

- assign M/N beams to user (feed back capacity)

$$C = M \log\log n/N + O(\log\log\log n)$$

- assign one beam to user (treat each antenna independently)

$$C = M \log\log nN + O(\log\log\log n)$$

The last is the best. In any case, multiple receive antennas do not substantially increase the network capacity.

- However, if some fraction of the users have N receive antennas they will receive N-times the rate.

Fairness of the System

- So far we have assumed that the network is homogenous
- Most networks are heterogenous, in the sense that the SNRs for the different users are different
- Thus, if we transmit to the most favorable users, the system may be dominated by the users with the highest SNR
- This is certainly true for single-transmit-antenna systems

Single Antenna Multi-User Fairness

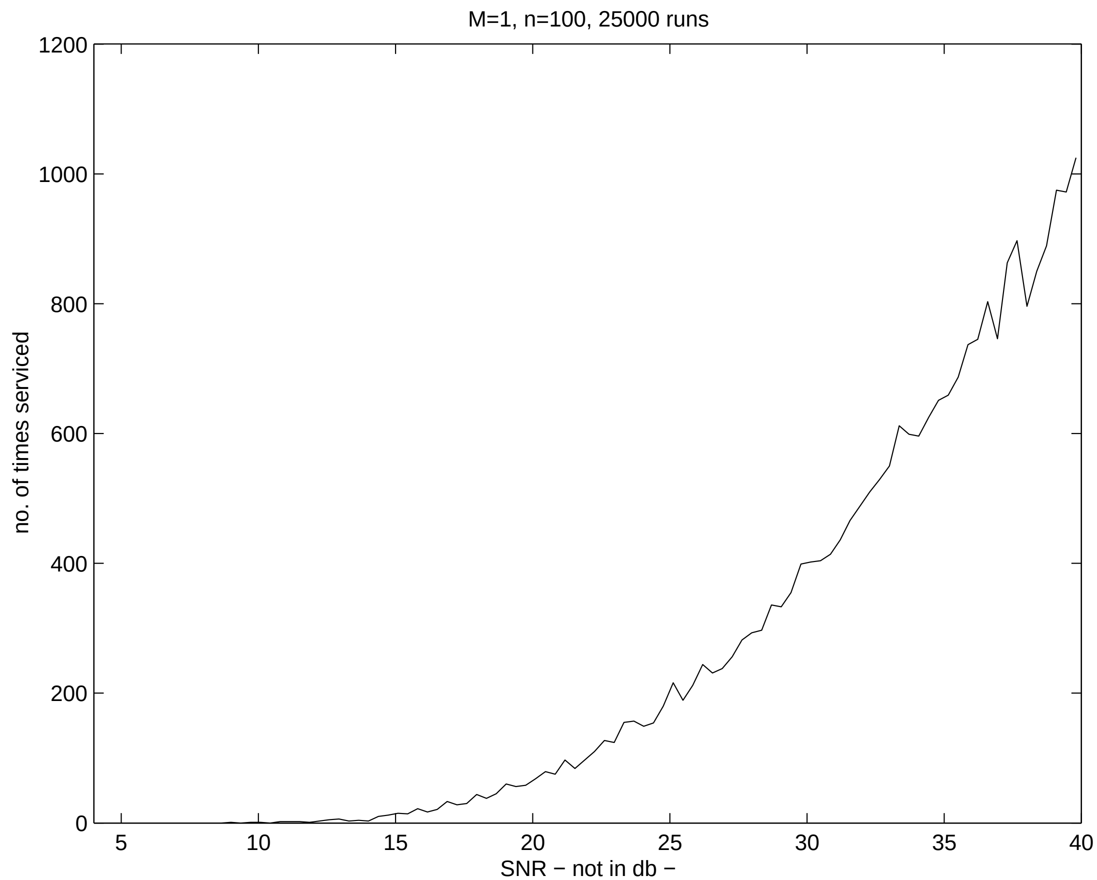


Figure 2: $M = 1$, $n = 100$, SNR 4-40 (6-16db), 25000 runs

Fairness - Continued

In a heterogenous network, the SINRs now become

$$\text{SINR}_{im} = \frac{|h_i \phi_m|^2}{1/\text{SNR}_i + \sum_{n \neq m} |h_i \phi_n|^2}$$

Note that, if the SNR_is are high enough, or if the number of transmit antennas is large enough, then the system is interference-dominated which implies *built-in fairness*.

Result:

$$P(\text{choosing user with SNR}_{min}) \geq \frac{1}{n} e^{-\left(\frac{1}{SNR_{min}} - \frac{1}{SNR_{max}}\right)\left(e^{\frac{\log n}{M}} - 1\right)},$$

This further illustrates the benefits of having $M = \alpha \log n$.

Multi-Antenna Multi-User Fairness

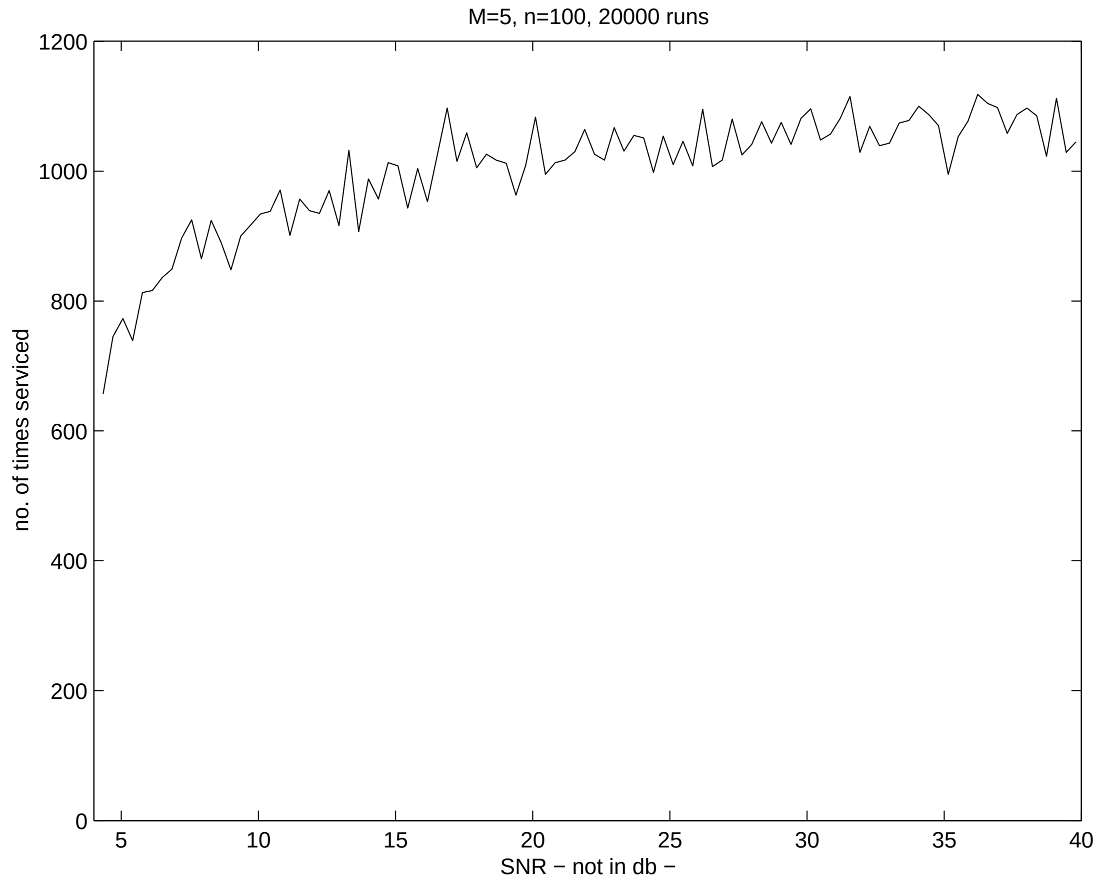


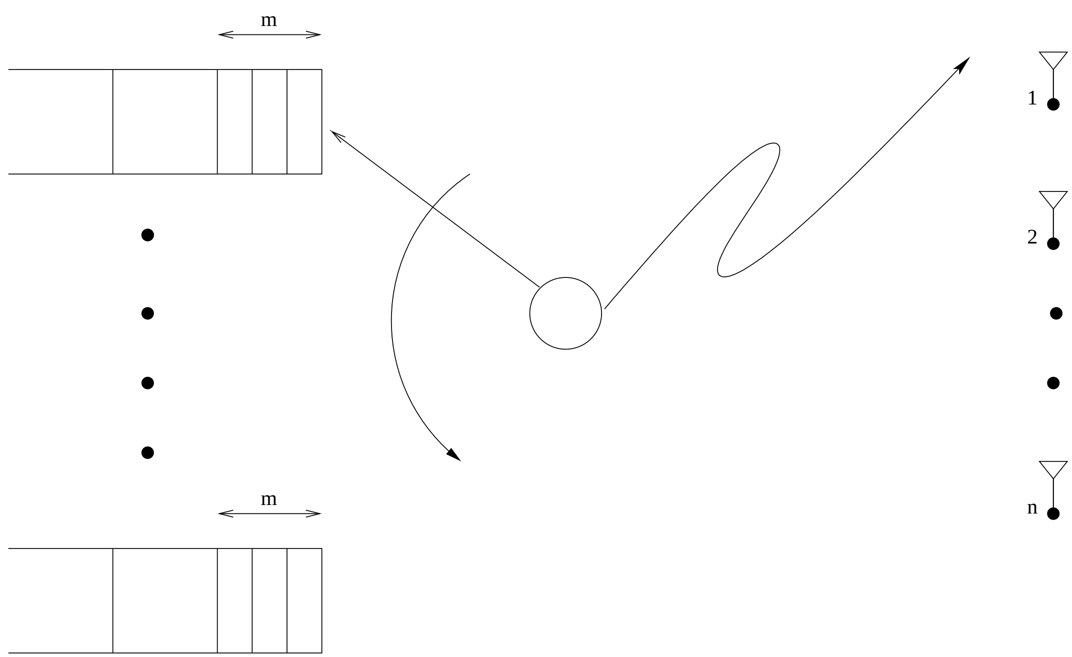
Figure 3: $M = 5$, $n = 100$, SNR 4-40 (6-16db), 20,000 runs

Delay Guarantee in Broadcast Channels

- So far we have considered capacity and fairness.
- The fairness we considered is really "long-term" fairness.
- We now want to focus on "short-term" fairness, which really means the study of delay.

Delay in SISO Broadcast Channels

- For simplicity, we start with SISO BC systems.
- We consider a packet-based transmission scheme.
- Delay, $D_{m,n}$, is defined as the minimum number of channel uses that guarantees that all n users successfully receive m packets.



Assumptions

- Channel is block fading, and is independent from packet to packet.
- Assume all the packets carry C_0 information bits.
- A packet is dropped if outage occurs, i.e., if the instantaneous capacity C goes below C_0 at the time of the transmission.
- The throughput therefore is $R = C_0 \Pr\{C \geq C_0\}$.
- The throughput-optimal strategy is the one that transmits to the user with the best channel condition.

Here are the main questions:

1. How bad is the delay for throughput-optimal strategies?
2. Can we reduce the delay at the expense of a small throughput hit?
3. What is the effect of temporal channel correlation?
4. What happens if we have multiple transmit antennas?

Delay: Round Robin vs Throughput-Optimal

- The minimum delay is mn, and is achieved by a round-robin scheme.
- Given the error probability $P_e = \Pr(C \leq C_0)$, the round-robin scheme achieves a rate of $R = C_0(1 - P_e)$.
- Similarly for the throughput-optimal, the rate would be

$$R = C_0' \Pr(\max_{i=1,\ldots,n} C_i \geq C_0').$$

- We are interested in characterizing the delay of this scheme.

Asymptotic Analysis of the Delay

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

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

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

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

$$E(D_{m,n}) = n \log n + O(n \log\log n) \quad , \quad \sigma^2(D_{m,n}) = O(n^2)$$

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

$$E(D_{m,n}) = 3.126 \underbrace{n \log n}_{mn} + O(n \log\log n).$$

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

$$E(D_{m,n}) = \underbrace{n(\log n)^r}_{mn} + O(n \log n),$$

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

Role of Multiple Antennas

Assuming that the channels fade independently from packet to packet, with M transmit antennas for fixed m and $n \to \infty$, we have

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

which is not much of an improvement over a system that is M times faster.

However, multiple antennas improve the delay in two important ways:

- When the channels have high correlation from packet to packet, the use of independent random beams can decorrelate the SINRs
- When $M = \log n$, using random beams makes all users equally likely to be chosen.

Summary

- scaling laws for sum rate of MIMO BC
  - full CSI in TX, capacity scales like $M \log\log n$
  - no CSI in TX, capacity scales like $\log M$
- proposed a simple scheme requiring very little feedback
  - throughput scales like $M \log\log n$
  - linear growth for $M = O(\log n)$
  - exploits multi-user diversity, but also has built-in fairness
  - optimal number of transmit antennas appears to be $\alpha \log n$
- analyzed the delay in broadcast channels
  - delay for the optimal-throughput scheduling
  - role of multiple antennas