The cardinal role of Scheduling in Downlink Multiuser MIMO Systems

Marios Kountouris

Eurecom Institute / France Telecom R&D

kountour@eurecom.fr

University of Southern California

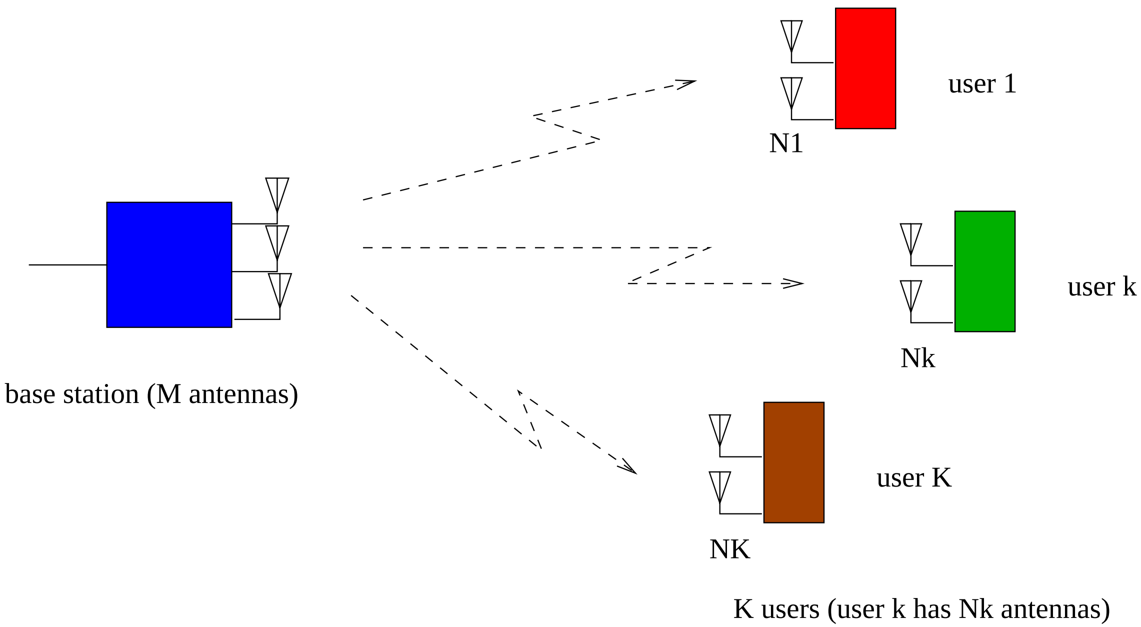
May 1^{st}, 2007

Outline

- Part I: *Scheduling in MIMO BC: living with partial CSIT*
  - Lessons learned from MU Information Theory.
  - The significant role of CSIT in MIMO BC.
  - Living with sparse networks.
  - Exploiting correlation in time domain.
  - Exploiting spatial statistical feedback in MIMO correlated channels.
  - Efficient metrics for MIMO Broadcast Scheduling.
- Part II: *Resource allocation in multicell wireless networks*
  - Network model and Objectives.
  - Distributed resource allocation in multicell networks.
  - Network capacity scaling laws for large number of users.
  - Performance evaluation.
  - Conclusions.

Promises and challenges of MU MIMO networks

Downlink MIMO Channels



- BS or AP equipped with M transmit antennas, and K active terminals. Each user k has N_k antennas.
- Active users: set of users simultaneously downloading data during one given scheduling window (subset of connected users).
- Users might have different QoS constraints/demands (data rate, delay).

What does MU Information Theory tell us?

- In MIMO broadcast channels (BC), the capacity can be boosted by means of SDMA $\longrightarrow$ serving multiple users simultaneously.
- Dirty Paper Coding (DPC) is capacity-achieving, but
  - *difficult* to implement in practice.
  - requires perfect Channel State Information at the Tx/Rx (CSIT/R).

  $\longrightarrow$ *practical* & *low-complexity* downlink transmission techniques are of interest!
- Downlink linear precoding, although suboptimal, is shown to achieve a large portion of DPC capacity, if combined with efficient *user selection* $\longrightarrow$ reduced complexity PHY layer.

Capacity scaling laws in MIMO BC

With full CSIT:

Assuming a block fading channel, an homogeneous network (all users have the same average SNR), $N_k = N$, fixed M and P, and large K:

$$\lim_{K\rightarrow\infty} \frac{\mathbb{E}(\mathcal{R}^{DPC})}{M \log\log(KN)} = 1 \quad (1)$$

- multiplexing gain of M (by sending data to M carefully selected users out of K).
- multiuser diversity offers an extra gain $\log\log(KN)$.

With no CSIT:

the sum-rate is reduced to (in the high SNR regime)

$$\mathbb{E}(\mathcal{R}^{NoCSIT}) \approx \min(M, N) \log SNR \quad (2)$$

Lessons learned from MU Information Theory

- The multiplexing gain is limited by the number of Tx antennas, and the number of simultaneously served users is upper bounded by M^2.
- Unlike in the SU setting, no need for large number of Rx antennas (low cost user terminals).
- MU MIMO schemes are immune to most of propagation limitations (e.g., rank loss). Rich scattering (NLOS) is not required as in SU MIMO.
- Resource allocation techniques help to fulfill the gains of multiuser MIMO systems.

However ...

the advantages of MU MIMO unfortunately come at a price!

The crucial role of CSIT

- The capacity gain of multiuser MIMO systems seems to remain highly dependent on the amount of feedback (CSIT).
  - if a BS communicating with K single-antenna mobiles has perfect CSIT, the multiplexing gain = $\min(M, K)$.
  - if there is a complete lack of CSIT, the multiplexing gain collapses to one!!
- The amount of feedback should be kept *minimal*.
- Full/perfect CSIT is often unrealistic and difficult to obtain at BS/AP.
- The level of CSIT critically affects the slope of the capacity vs. SNR curve (multiplexing gain) of the MIMO downlink channel.
- To achieve *full multiplexing gain* the number of FB bits/user must increase approx. linearly with M and the average SNR (in dB) [Jindal'05].

Challenges in Downlink MIMO Channels

- Can we still benefit from MUDiv and spatial multiplexing gains under limited feedback rate constraint ?
- What kind of partial CSIT is necessary and sufficient ?!
  - to identify "good" users to be scheduled.
  - to design efficient precoding techniques.
  - to achieve near optimal capacity growth.
- How can the scheduler use *efficiently* all possible degrees of freedom?

Part I

Scheduling in downlink MU MIMO systems: living with partial channel knowledge

System Model

- Multiple antenna broadcast channel with M transmit antennas and K single-antenna receivers (with $K \geq M$).
- At the k-th mobile the received signal is given by

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k, \qquad k = 1, \ldots, K \tag{3}$$

Assumptions:

- We consider an i.i.d. block Rayleigh flat fading channel, where each user has perfect knowledge of its own channel $\mathbf{h}_k$ (CSIR).
- The transmitted signal is subject to an average transmit power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} = P$.
- Homogeneous network where all users have the same average signal-to-noise ratio (SNR).

Quantization-based techniques

- Quantization is the first idea that comes to mind when dealing with source compression (random channel matrix/vector or the corresponding precoders are the possible sources).
- The amount of feedback depends on the frequency of feedback (generally a fraction of the coherence time), the number of parameters being quantized, and the resolution of the quantizer.
- A well studied problem in single-user MIMO communication systems (limited feedback precoding, Grassmannian quantization, etc).
- The extension of SU MIMO ideas is not straightforward: need for efficient scheduling metrics.

Dimension-reduction techniques (1/2)

- Projection of the channel matrix/vector onto one or more basis vectors known to the Tx and Rx
$\rightarrow$ CSI matrix of size $N \times M$ is mapped into an p-dimensional vector with $1 \leq p \leq N \times M$ (dimensionality reduction).
- Once the projection is carried out, the receiver feeds back a metric $\xi_k = f(\mathbf{H})$ which is typically related to the square magnitude of the projected signal.
- Antenna selection falls into that category.

Dimension-reduction techniques (2/2)

- Alternatively, the projection can be the result of using a particular precoder at the BS.
  Let $N_k = 1$, the BS designs an arbitrary unitary precoder $\mathbf{Q}_{M \times M}$. Each terminal identifies the projection of its vector channel onto the precoder and reports its best SINR:

$$\xi_k = \max_{1 \leq i \leq M} \frac{|\mathbf{h}_k^H \mathbf{q}_i|^2}{\sigma^2 + \sum_{j \neq i} |\mathbf{h}_k^H \mathbf{q}_j|^2} \tag{4}$$

  where $\mathbf{q}_i$ denotes the i-th column of $\mathbf{Q}$. The scheduling algorithm opportunistically assigns to each beamformer $\mathbf{q}_i$ the user which has selected it and has reported the highest SINR.
- Random Opportunistic BF [Viswanath et al.'02], [Sharif,Hassibi'05] falls into that category (the precoder is randomized to change the system dynamics).

Using power allocation to restore robustness in sparse networks

joint work with Prof. David Gesbert

Living with sparse networks

- Random Beamforming yields optimal capacity growth of $M \log \log K$ for large number of active users.
- The performance degrades severely with decreasing - yet practical - number of users (sparse networks).

How to restore robustness in cases with moderate number of users ?

Idea: modify the random beams for a better matching with the actual users' channels (with little or no extra feedback cost).

- beam power control.
- beam selection (multi-mode scheme).

Beam power control (1/2)

Denote $\mathcal{S}$ the set of selected users and $\mathbf{p}$ the beam power vector.

We assume the BS knows $g_{km} = \left|\mathbf{h}_k^H \mathbf{q}_m\right|^2$ for $k \in \mathcal{S}, m = 1, \ldots, M$.

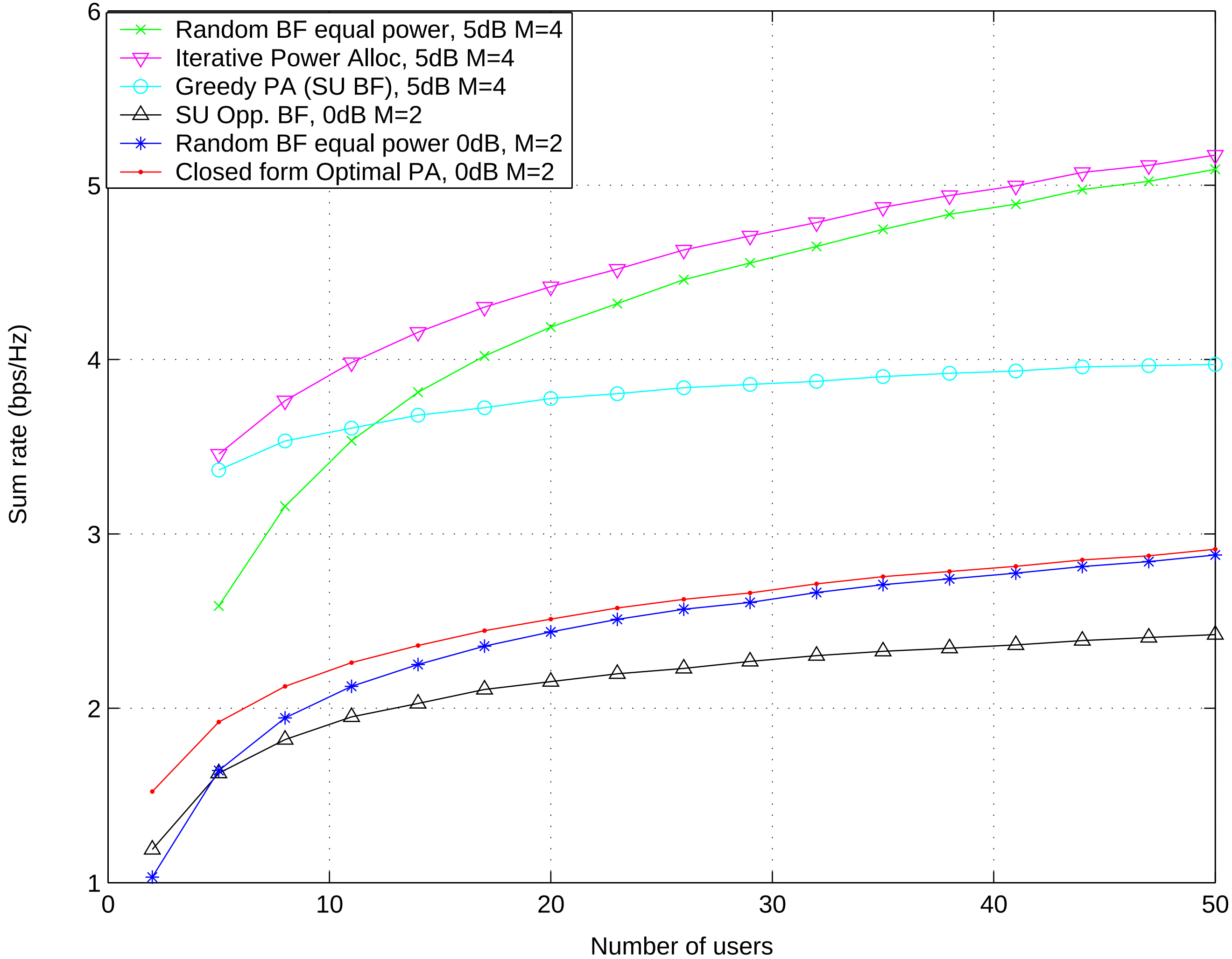
The sum-rate optimal beam power allocation results into the following optimization [SPAWC'05]:

$$\max_{\mathbf{p}} \sum_{k \in \mathcal{S}} \log \left(1 + \frac{P_m g_{km}}{1 + \sum_{j \neq m} P_j g_{kj}} \right)$$

$$\text{subject to} \sum_{i=1}^{M} P_i = P$$

- Closed-form solution for $M = 2$ antennas (optimal).
- Iterative WF-like algorithm for $M > 2$ (optimality is not guaranteed).

Beam power control (2/2)



Performance of Beam Power Allocation vs. the number of users for $M = 2, 4$ Tx antennas.

Beam selection (1/2)

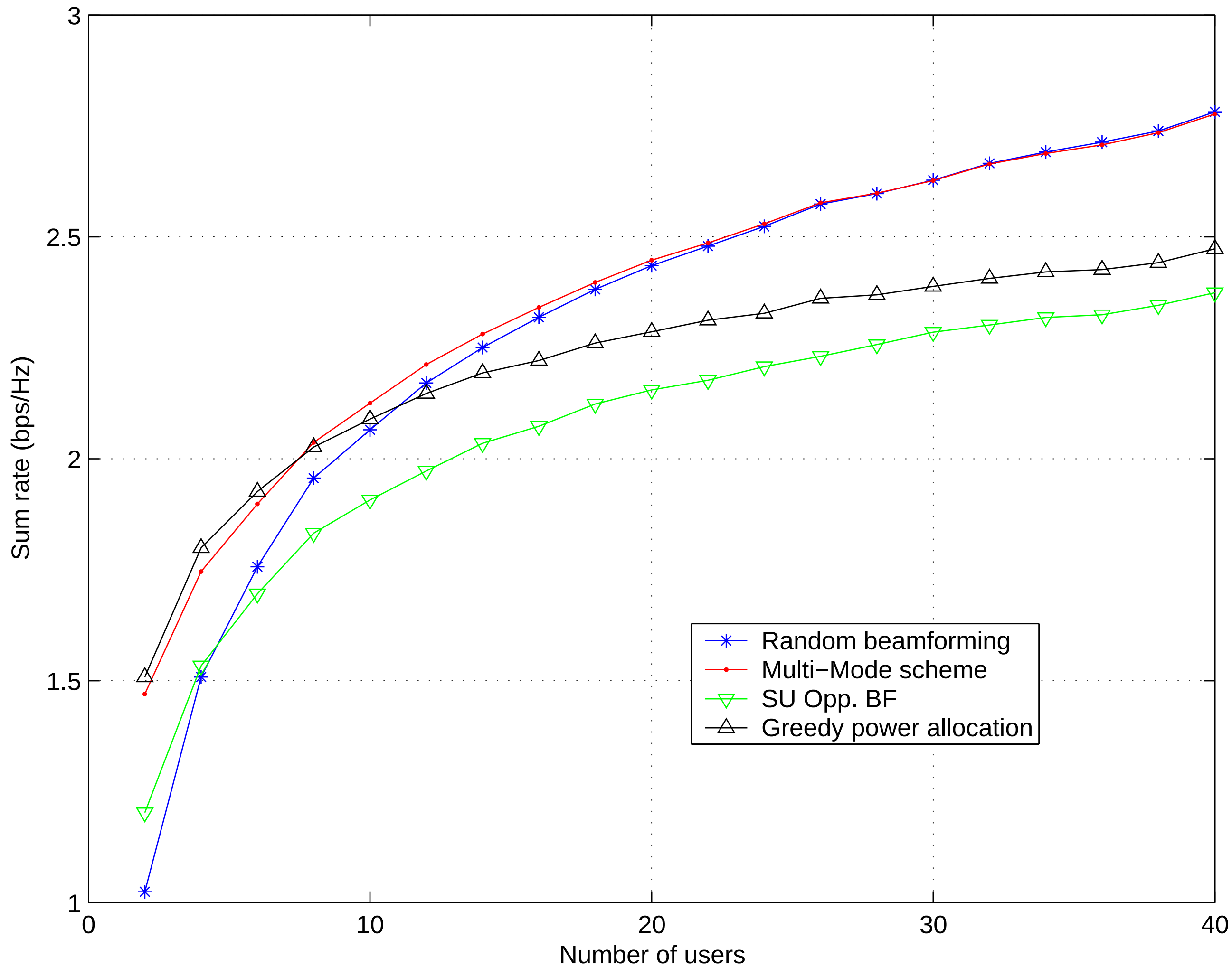
Key idea: When the random beams are not well aligned with users' channels, turn off certain beams $\rightarrow$ reducing inter-user interference.

- Multi-mode scheme switching from TDMA to SDMA transmission mode, depending on the number of active users and average SNR.
- An efficient method of balancing multiuser diversity and multiplexing gain.
- A means to achieve linear capacity growth at high SNR (no capacity ceiling effect in the interference-limited region).

Challenges:

- solutions that require no extra feedback are mostly heuristic.
- mobiles need to know at each scheduling slot the number of active users, K.
- extension to heterogeneous networks, multiple-antenna terminals is not evident.

Beam selection (2/2)



Sum rate as a function of the number of users for $M = 2$ Tx antennas and $SNR = 0$ dB.

Scheduling in temporally correlated channels

joint work with Prof. David Gesbert

User Scheduling exploiting temporal correlation

Assume that the channel is correlated in the time domain ...

"Previously selected 'good' users are highly likely to remain good."

Time correlation can be exploited:

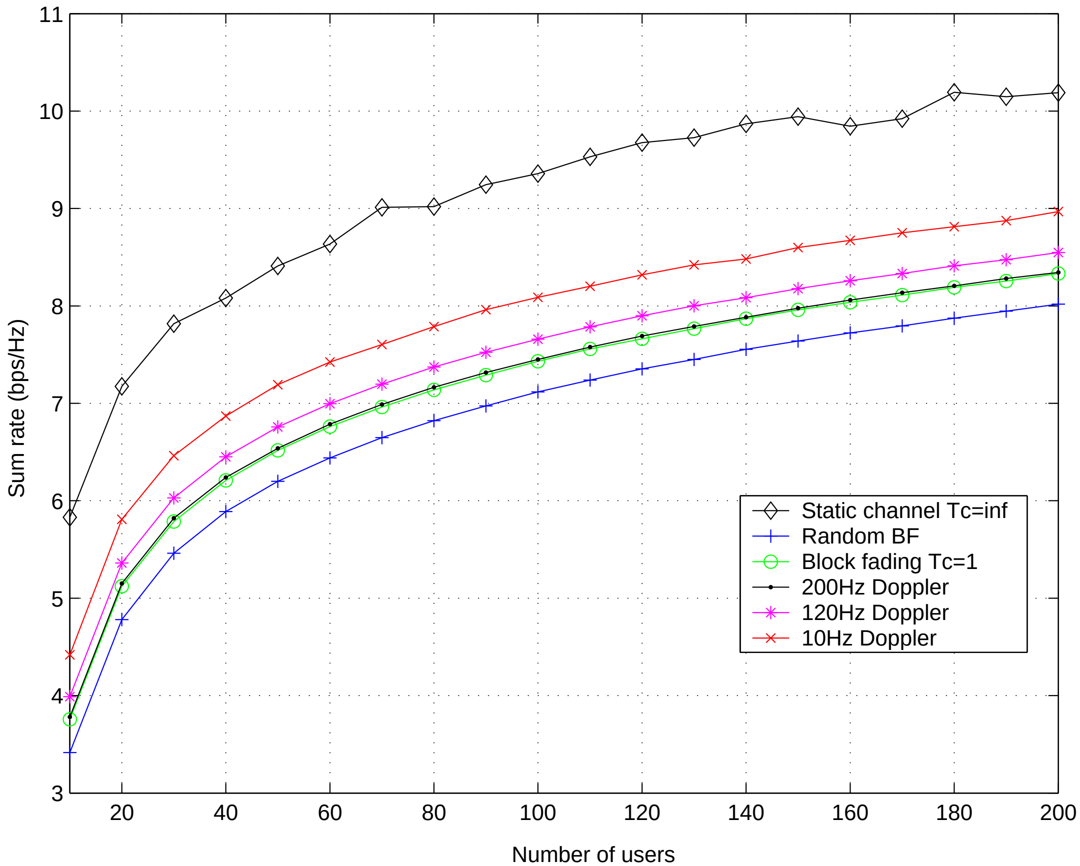
- *Feedback aggregation concept*: information derived from low-rate feedback channel can be cumulated over time to approach the performance of full CSIT scenario.
- *Feedback compression*: the channel can be seen as a Markov source
  - feedback rate reduction (by truncating the transition probabilities).
  - compensate the effect of feedback delay.

Challenges:

- solutions work well for low Doppler users.
- results quite sensitive to channel model (e.g. Jakes, Gauss-Markov, etc.).

An example in Random Opport. BF context

The beams are refined over time by storing and reusing the random precoding matrix that offers the highest sum-rate [ISIT’05].



Sum rate as a function of the number of users for $M = 2$ Tx antennas and $SNR = 10$ dB.

Exploiting statistical feedback in spatially correlated channels

joint work with Prof. David Gesbert

Why use spatial statistical feedback ?

- Spatial channel statistics reveal a great deal of information on the *macroscopic* nature of the channel (multipath's mean AoA, angular spread).
- The advantage of statistical CSI is its long coherence time compared with that of the fading channel.
- Several forms of statistical CSI are reciprocal (second-order correlation matrix, power of Ricean component, etc.) $\rightarrow$ no additional feedback required.
- Second-order statistical information can be used to infer knowledge on users' spatial separability.

Using spatial statistical feedback in MIMO BC

- Consider a correlated Rayleigh MISO channel $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{R}_k)$, where $\mathbf{R}_k \in \mathbb{C}^{M \times M}$ is the transmit covariance matrix (known to BS).
- *Objective*: How to combine long-term CSIT with instantaneous scalar CSIT in order to exploit Multiuser Diversity ?
- A general form of short-term CSIT is

$$\gamma_k = \|\mathbf{h}_k \mathbf{Q}_k\|^2 \tag{5}$$

  where $\mathbf{Q}_k \in \mathbb{C}^{M \times L}$ is a training matrix containing L orthonormal vectors $\{\mathbf{q}_{ki}\}_{i=1}^{L}$.
- Key idea: Conditioned on short-term CSIT γ_k, derive a (coarse) channel estimate.

ML estimation framework

- We estimate the channel as the vector that maximizes the PDF of $\mathbf{h}_k$ under the scalar constraint $\gamma_k = |\mathbf{h}_k \mathbf{q}_k|^2$ ($L = 1$).
- The solution to the optimization problem

$$\begin{aligned} \max_{\mathbf{h}_k} \quad & \mathbf{h}_k \mathbf{R}_k \mathbf{h}_k^H \\ s.t. \quad & |\mathbf{h}_k \mathbf{q}_k|^2 = \gamma_k \end{aligned} \tag{6}$$

is given by [Eusipco'06]

$$\widehat{\mathbf{h}}_k = \arg\max_{\mathbf{h}_k} \frac{\mathbf{h}_k \mathbf{R}_k \mathbf{h}_k^H}{\mathbf{h}_k (\mathbf{q}_k \mathbf{q}_k^H) \mathbf{h}_k^H} \tag{7}$$

which corresponds to the (dominant) generalized eigenvector associated with the largest positive generalized eigenvalue of the Hermitian matrix pair $(\mathbf{R}_k, \mathbf{q}_k \mathbf{q}_k^H)$.

What to do with a coarse $\hat{\text{h}}$ estimate ?

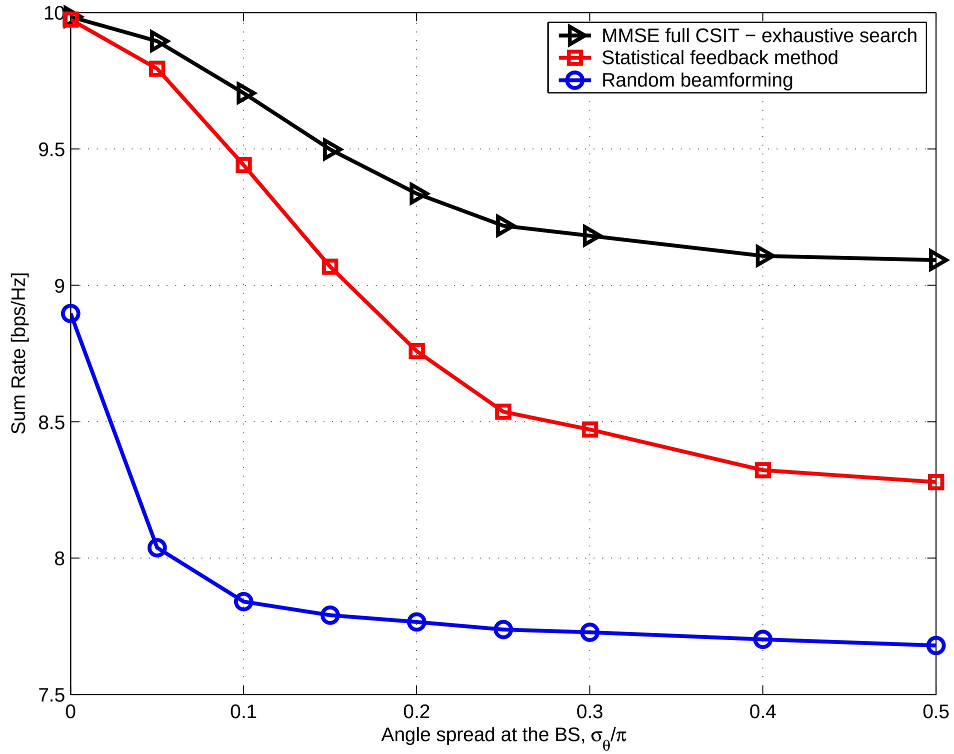
Once the channel estimate is derived for all users ...

- Approach I [ICASSP'06]
  - Find the scheduling set of M users using

$$\mathcal{S}^* = \arg\min_{\mathcal{S}} \mathrm{Tr}\left\{\left(\left(\hat{\mathbf{H}}(\mathcal{S})^H\hat{\mathbf{H}}(\mathcal{S}) + \frac{M}{P}\mathbf{I}\right)^{-1}\right)^2\right\}$$

  - Full CSIT is acquired for the selected M users.
  - Optimal MMSE precoders are used to serve these users.
- Approach II [Eusipco'06]
  User Selection and MMSE beamforming are both performed by using the coarse channel estimate.

ML estimation framework - approach I



Sum rate as a function of the angle spread σ_θ at the base station ($M = 2$, SNR = 10 dB and $K = 50$).

- Performance close to that of full CSIT when the multipath angular spread per user at the BS is small enough (less 30° - ideal for wide area networks).
- Robustness to the case of wide angular spread (worst case performance is that of random beamforming).

Efficient metrics for MIMO Broadcast Scheduling

joint work with D. Gesbert, R. de Francisco, D. Slock, T. Sälzer

CSIT feedback model

- Consider now that the feedback channel is divided into 2 types of information:
  - Channel Direction Information (CDI)
  - Channel Quality Information (CQI)

Motivation:

- CDI can be used to achieve full multiplexing gain (with proper feedback load scaling) when $K \leq M$.
- When $K > M$, CDI cannot exploit multiuser diversity gain $\rightarrow$ *an efficient CQI metric is required*.

However, the feedback rate is finite $\rightarrow$ loss of degrees of freedom (MUDiv and MUX gain).

What type of CQI metric allows us to still benefit from the gains promised by MU MIMO with limited CSIT ?

CDI Finite Rate Feedback Model

- Quantization codebook $\mathcal{V}_k = \{\mathbf{v}_{k1}, \mathbf{v}_{k2}, \ldots, \mathbf{v}_{kN}\}$ containing $N = 2^B$ unit norm vectors (known to both the k-th Rx and Tx).
- At each timeslot, the k-th mobile, based on its current channel realization $\mathbf{h}_k$, determines the vector that maximizes

$$\hat{\mathbf{h}}_k = \mathbf{v}_{kn} = \arg\max_{\mathbf{v}_{ki} \in \mathcal{V}_k} |\bar{\mathbf{h}}_k^H \mathbf{v}_{ki}|^2 = \arg\max_{\mathbf{v}_{ki} \in \mathcal{V}_k} \cos^2(\angle(\bar{\mathbf{h}}_k, \mathbf{v}_{ki})) \tag{8}$$

  where $\bar{\mathbf{h}}_k = \mathbf{h}_k / \|\mathbf{h}_k\|$.
- Each user sends the corresponding quantization index n back to the transmitter through an error-free, and zero-delay feedback channel using $B = \lceil \log_2 N \rceil$ bits.

Zero Forcing (ZF) Beamforming (1/2)

- Let $\mathcal{G} = \{1, \cdots, K\}$ be the set of indices of all K users. Let $\mathcal{S} \in \mathcal{G}$, be one such group of $|\mathcal{S}| = \mathcal{M} \leq M$ users selected for transmission at a given time slot.
- The signal model is given by

$$\mathbf{y}(\mathcal{S}) = \mathbf{H}(\mathcal{S})\mathbf{W}(\mathcal{S})\mathcal{P}\mathbf{s}(\mathcal{S}) + \mathbf{n} \tag{9}$$

where $\mathbf{H}(\mathcal{S})$, $\mathbf{W}(\mathcal{S})$, $\mathbf{s}(\mathcal{S})$, $\mathbf{y}(\mathcal{S})$ are the concatenated channel vectors, beamforming vectors, uncorrelated data symbols and received signals respectively.

- We use ZF beamforming on the quantized channel directions as a multiuser transmission strategy:

$$\mathbf{W}(\mathcal{S}) = \hat{\mathbf{H}}(\mathcal{S})^H(\hat{\mathbf{H}}(\mathcal{S})\hat{\mathbf{H}}(\mathcal{S})^H)^{-1}\mathbf{\Lambda} \tag{10}$$

Zero Forcing (ZF) Beamforming (2/2)

- The SINR at the k-th receiver is

$$SINR_k = \frac{P_k|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum\limits_{j \in \mathcal{S}-\{k\}} P_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + 1} \tag{11}$$

  where $\sum_{i \in \mathcal{S}} P_i = P$ in order to satisfy the power constraint on the transmitted signal.

- We focus on the ergodic data rate which, assuming Gaussian inputs, is equal to

$$\mathcal{R}_k = \mathbb{E}\left\{\sum_{k \in \mathcal{S}} \log\left(1 + SINR_k\right)\right\} \tag{12}$$

CQI feedback achieving near optimal sum-rate by means of multiuser diversity

Metric I: Upper Bound on SINR

Bounding the expected interference

- Let $\phi_k = \angle(\hat{\mathbf{h}}_k, \bar{\mathbf{h}}_k)$ be the angle between the normalized channel vector and the quantized channel direction.
- Consider that each user provides information on its effective channel (SINR) by feeding back the following scalar metric

$$\xi_k^{UB} = \frac{P\left\|\mathbf{h}_k\right\|^2 \cos^2\phi_k}{P\left\|\mathbf{h}_k\right\|^2 \sin^2\phi_k + M} \tag{13}$$

- This metric encapsulates information on the channel gain as well as the CDI quantization error ($\sin^2\phi_k$).
- It can be interpreted as an upper bound (UB) on the received SINR$_k$ (equal power allocation) $\longrightarrow$ an estimate of the multiuser interference *at the mobile side* (without cooperation).

Metric II: Lower Bound on SINR (1)

Bounding the actual multiuser interference

- Let $\cos\theta_k = \left|\bar{\mathbf{h}}_k^H \mathbf{w}_k\right|$, ϵ is the orthogonality between the quantized channels and $\epsilon_{ZF} = \max_{i,j\in\mathcal{S}} \left|\mathbf{w}_i^H \mathbf{w}_j\right|$ (worst-case orthogonality under ZFBF).
- **Theorem** [ICASSP'07]: *Given an ϵ-orthogonal set $\mathcal{S}$, with $|\mathcal{S}| = M$, a system that performs ZFBF can guarantee the following SINR for the k^{th} user*

$$SINR_k^{ZF} \geq \frac{P\left\|\mathbf{h}_k\right\|^2 \cos^2\theta_k}{P\left\|\mathbf{h}_k\right\|^2 \overline{I}_{UB_k} + M} \tag{14}$$

where $\overline{I}_{UB_k} = \cos^2\theta_k\ (M-1)\epsilon_{ZF}^2 + \sin^2\theta_k\ [1+(M-2)\epsilon_{ZF}]$
$+2\sin\theta_k\cos\theta_k\,(M-1)\epsilon_{ZF}$

and

$$\epsilon_{ZF} = \vartheta \text{ and } \cos\theta_k = \frac{\left|\cos\phi_k - \sqrt{\vartheta}\right|}{1+\vartheta} \text{ with } \vartheta = \frac{\epsilon}{1-(M-1)\epsilon}$$

Metric II: Lower Bound on SINR (2)

- Motivated by the above results, let each user provide information on its SINR lower bound (LB) and report the following scalar metric

$$\xi_k^{LB} = \frac{\frac{P}{(1+\vartheta)^2} \left\| \mathbf{h}_k \right\|^2 (\cos \phi_k - \sqrt{\vartheta})^2}{P \left\| \mathbf{h}_k \right\|^2 \overline{I}_{UB_k} + M} \tag{15}$$

Metric III: Decomposing the CQI into two scalars (1/2)

- Main drawback of the above metrics: they estimate the SINR by assuming $|\mathcal{S}| = \mathcal{M} = M$.
- However, in extreme regimes (high SNR, low number of users), it is often better to transmit to $\mathcal{M} < M$ users.
- We decompose the CQI feedback by letting each user feed back the following two scalar values:
  - the alignment $\cos^2 \phi_k$
  - the channel norm, $\|\mathbf{h}_k\|^2$

Metric III: Decomposing the CQI into two scalars (2/2)

Now, everything can be done at base station ...

- The transmitter selects the user based on the following lower bound on the received SINR

$$\xi_k^{LBd} = \frac{P \left\|\mathbf{h}_k\right\|^2 \rho_k^2}{P \left\|\mathbf{h}_k\right\|^2 \bar{I}_{UBd_k} + \mathcal{M}} \tag{16}$$

  where $\rho_k^2 = \cos^2(\phi_k + \angle(\hat{\mathbf{h}}_k, \mathbf{w}_k))$ and $1 \leq \mathcal{M} \leq M$.

- Assuming $\epsilon \to 0$, then $\bar{\mathrm{I}}_{UBd_k} \to \sin^2 \phi_k \to$ a more refined metric can be used

$$\xi_k^{UBd} = \frac{P \left\|\mathbf{h}_k\right\|^2 \rho_k^2}{P \left\|\mathbf{h}_k\right\|^2 \sin^2 \phi_k + \mathcal{M}} \tag{17}$$

User Selection Schemes (1/2)

Greedy - Semi-orthogonal US [Yoo et al.'06]

- Define ϵ as the maximum allowed non-orthogonality (maximum correlation) between $\hat{\mathbf{h}}_k$ (set *a priori*).

Step 0 Set $\mathcal{S} = \emptyset$, $\mathcal{Q}^0 = \{1, \ldots, K\}$

Select the first user: $k_1 = \arg\max_{k \in \mathcal{Q}^0} \xi_k^{UB}$

For $i = 1, 2, \ldots, M-1$ repeat

Step 1 Define user set $\mathcal{Q}^i = \left\{1 \leq k \leq K : |\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_{k_j}| \leq \epsilon, 1 \leq j \leq i\right\}$

Step 2 $k_{i+1} = \arg\max_{k \in \mathcal{Q}^i} \xi_k$

Step 3 $\mathcal{S} = \mathcal{S} \cup k_{i+1}$

---

where $\mathcal{R}(\mathcal{S}_i) = \sum_{k \in \mathcal{S}_i} \log_2(1 + \xi_k)$, with ξ_k being either: ξ_k^{UB}, ξ_k^{LB}, ξ_k^{UBd} or ξ_k^{LBd}.

User Selection Schemes (2/2)

Greedy - User Selection

- We extend the greedy user selection algorithm of [Dimic et al.'05] for the case of limited feedback.

Step 0 Set $\mathcal{S}_0 = \emptyset$, $\mathcal{S} \subseteq \mathcal{G}$ and $\mathcal{R}(\mathcal{S}_0) = \emptyset$

For $i = 1, 2, \ldots, M$ repeat

Step 1 $\quad k_i = \arg \max_{k \notin \mathcal{S}_{i-1}} \mathcal{R}(\mathcal{S}_{i-1} \cup \{k\})$

Step 2 $\quad$ if $\mathcal{R}(\mathcal{S}_{i-1} \cup \{k\}) < \mathcal{R}(\mathcal{S}_{i-1})$

set $\mathcal{S} = \mathcal{S}_{i-i}$ and break

else if $i = M$ then $\mathcal{S} = \mathcal{G}$ and break

else set $\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{k\}$ and go to **Step 1**

---

where $\mathcal{R}(\mathcal{S}_i) = \sum_{k \in \mathcal{S}_i} \log_2(1 + \xi_k)$, with ξ_k being either: ξ_k^{UB}, ξ_k^{LB}, ξ_k^{UBd} or ξ_k^{LBd}.

Performance analysis (1/2)

- The sum rate $\mathcal{R}$ of a system using the Greedy-SUS algorithm in conjunction with metric II is lower bounded by:

$$\mathcal{R} \geq M \left(1 - \Pr\left\{|\mathcal{S}| \neq M\right\}\right) \log_2 \left(1 + SINR_{min}^{LB}\right) \tag{18}$$

where $SINR_{min}^{LB} = \min_{k \in \mathcal{S}} \xi_k^{LB}$.

- If ϵ is set such that $\lim_{K \to \infty} \epsilon = 0$ (e.g. $\epsilon = \frac{1}{\log K}$), then $SINR_{min}^{LB} \stackrel{K \to \infty}{=} \min_{k \in \mathcal{S}} \xi_k^{UB}$ and

$$\lim_{K \to \infty} \left(\mathcal{R}_{ZF}^{opt} - \mathcal{R}\right) = \lim_{K \to \infty} \left[M \log_2 \frac{1 + \frac{P}{M} \log K}{1 + \frac{P}{M} \log\left(\frac{K}{c}\right)} \right] = 0$$

where $c = 2^{-B} (P/M)^{M-1}$.

Performance analysis (2/2)

- In the interference-limited region ($P \to \infty$), the sum rate $\mathcal{R}$ is given by

$$\mathcal{R} \approx \frac{M}{M-1}\left(H_K + B\right) + b \qquad (19)$$

where $H_K = \sum_{k=1}^{K} \frac{1}{k}$ is the harmonic number and b is a constant.

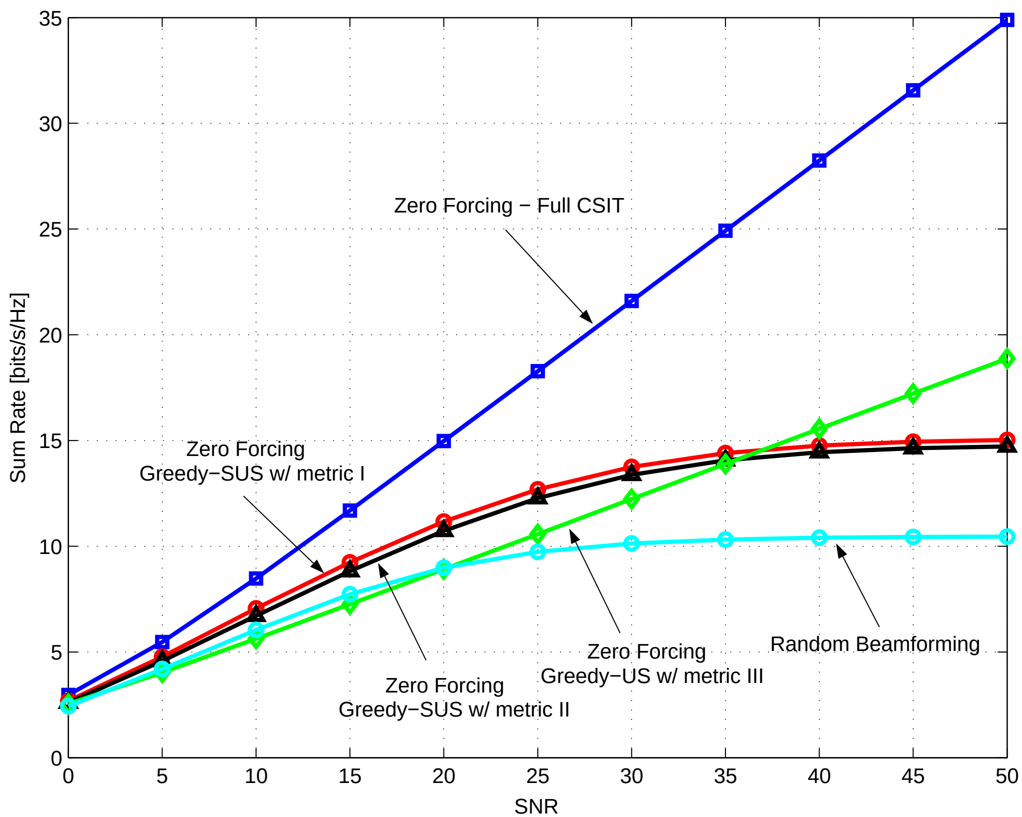
- If the number of user is very large,

$$\lim_{K \to \infty} H_K = \log K + \gamma \qquad (20)$$

where γ is the Euler-Mascheroni constant.

- Multiuser diversity can reduce significantly the feedback load in the high SNR regime.

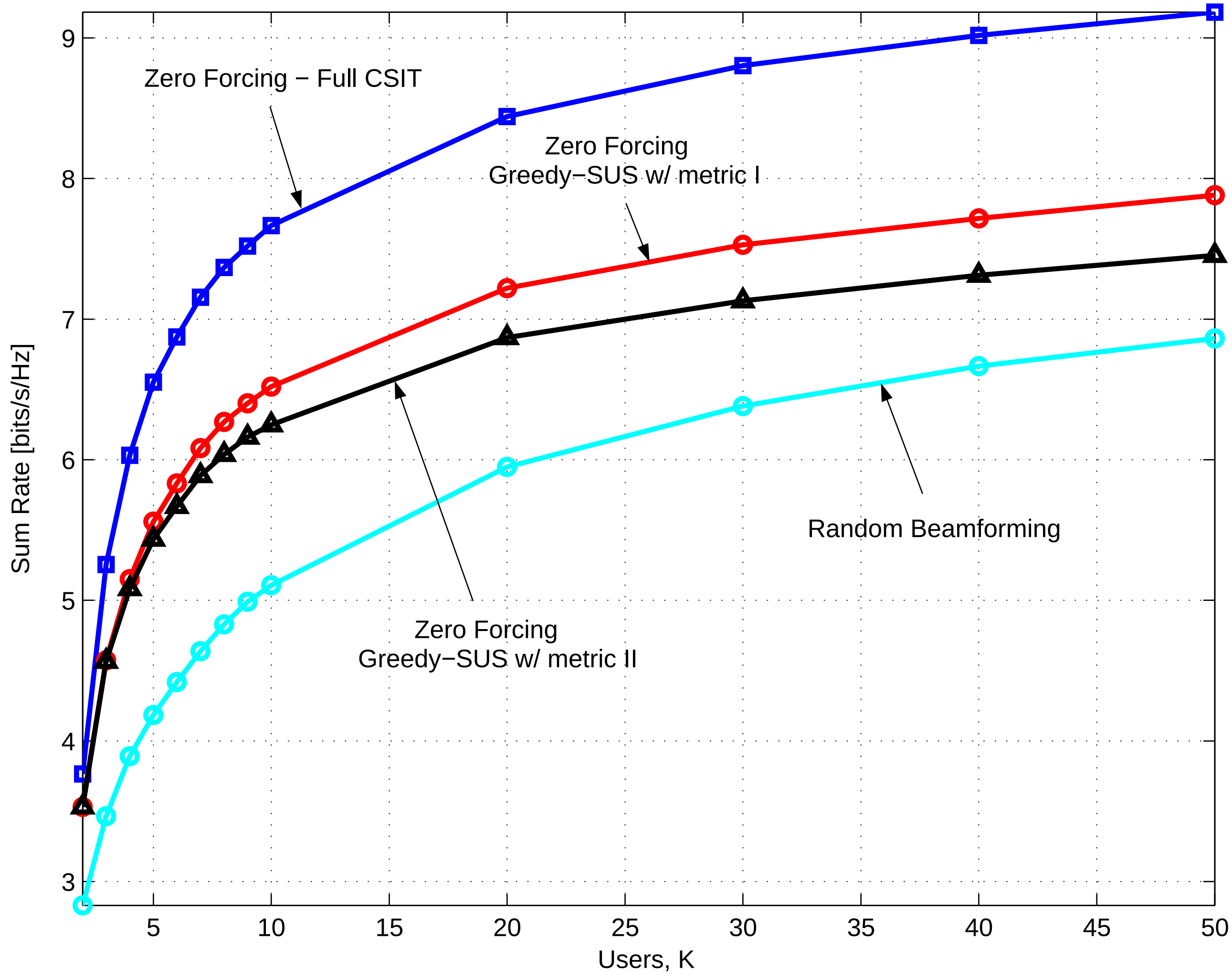
Numerical Results (1)



Sum rate vs. the average SNR for $B = 4$ bits, $M = 2$ Tx antennas and $K = 20$ users.

- Scheme III exhibits a linear sum rate growth in the interference-limited region (high SNR).
- Adaptive multi-mode schemes, switching from SDMA to TDMA, is of particular interest in MISO BC with limited feedback.

Numerical Results (2)



Sum rate as a function of the number of users for $B = 4$ bits, $M = 2$ Tx antennas and $SNR = 10$ dB.

Conclusions

- Combining simple MU MIMO techniques with smart scheduling seems a promising direction to realize the gains of downlink MIMO channels.
- Several approaches that exploit multiuser diversity can be used to restore robustness with respect to the lack of perfect CSIT.
- Scalar feedback metrics, providing an estimate of the multiuser interference at the receiver side, have been proposed.
- These metrics, combined with efficient scheduling and ZF beamforming, can achieve a significant fraction of the capacity of the full CSIT case.

The road ahead ...

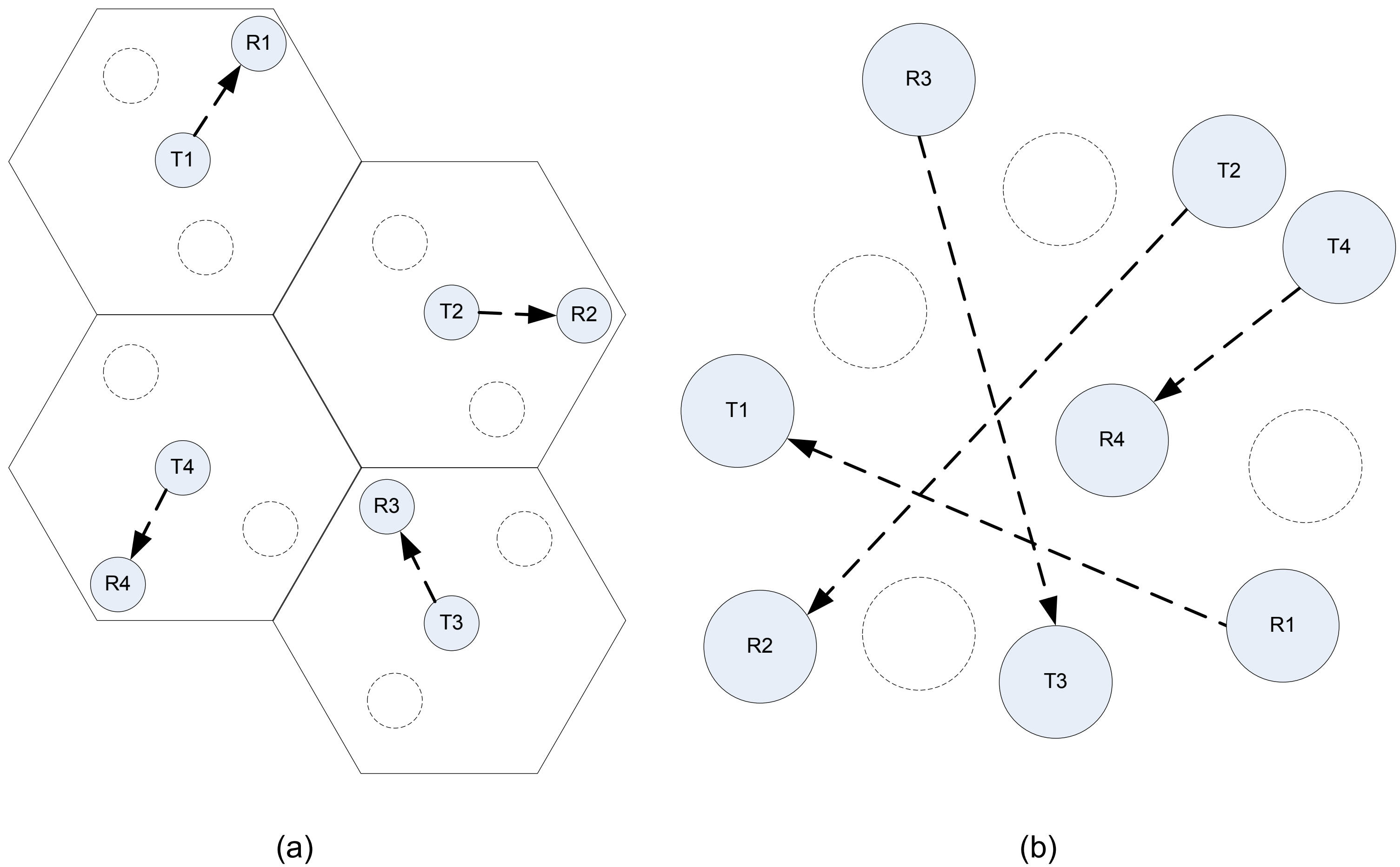
- Extensions to different settings (multi-antenna receivers, wideband, asymmetric users' SNRs ...). Many challenging problems remain open!
- More realistic assumptions need to be considered (feedback delay/errors, pathloss, fairness, training vs. CSIT gain tradeoff, cost of feedback, etc.)

Part II

Resource allocation in multicell wireless networks

joint work with Prof. David Gesbert

Cellular or Ad-hoc wireless networks ?



- cellular network with reuse one (any wireless system where users ultimately connect to an infrastructure point, e.g. BS, AP etc.).
- single-hop ad-hoc (peer-to-peer) network.

Introduction

Objective

- Optimization of the sum-rate (network capacity) in multicell interference-limited wireless networks.
- Joint resource allocation policies are considered: *power control* and *user scheduling*.

Challenges

- computational complexity of joint resource allocation policies.
- requirement for increased Rx-to-Tx feedback across all network cells.

Network model

- We consider a wireless network with $\mathcal{N}$ transmitters and receivers, with $N \leq \mathcal{N}$ transmit-receive active pairs selected for transmission by the scheduling protocol.
- Inter-cell interference due to full reuse of spectral resource.

In cell n, the received signal Y_{k_n} at user k_n is given by

$$Y_{k_n} = \alpha_{k_n,n} X_{k_n} + \sum_{i \neq n}^{N} \alpha_{k_n,i} X_{k_i} + Z_{k_n}$$

where

X_{k_n}: the message-carrying signal from the serving AP, subject to a peak power constraint P_{max}.

$\sum_{i \neq n}^{N} \alpha_{k_n,i} X_{k_i}$: the sum of interfering signals from other cells.

Z_{k_n}: additive noise, modeled (for convenience) as white Gaussian with power $\mathbb{E}|Z_{k_n}|^2 = \sigma^2$.

Multicell resource allocation problem

Motivations

- When network grows large, distributed MIMO is too complex, or multicell routing of data not desirable.
- Multicell R.A. addresses the allocation of spectral resources (time, frequency, codes, power) to optimize the system capacity.
- Here we focus on user scheduling and power control.
- Fundamentally exploits the variability (diversity) of channel in all dimensions.

Multicell R.A. can be

- Centralized
- Distributed

Definitions for scheduling and power control

Definition 1. *A* **scheduling vector** $\boldsymbol{K}$ *contains the set of users simultaneously scheduled across all* N *cells in the same slot:*

$$\boldsymbol{K} = [k_1\, k_2\, \cdots\, k_n\, \cdots\, k_N]$$

where $[\boldsymbol{K}]_n = k_n$.
Noting that $1 \leq k_n \leq K_n$, *the feasible set of scheduling vectors is given by* $\Upsilon = \{\boldsymbol{K} \mid 1 \leq k_n \leq K_n \,\forall\, n = 1, \ldots, N\}$.

Definition 2. *A* **transmit power vector** $\boldsymbol{P}$ *contains the transmit power values used by each AP to communicate with its respective user:*

$$\boldsymbol{P} = [P_{k_1}\, P_{k_2}\, \cdots\, P_{k_n}\, \cdots\, P_{k_N}]$$

where $[\boldsymbol{P}]_n = P_{k_n} = \mathbb{E}|X_{k_n}|^2$.
Due to the peak power constraint $0 \leq P_{k_n} \leq P_{\max}$, *the feasible set of transmit power vectors is given by* $\Omega = \{\boldsymbol{P} \mid 0 \leq P_{k_n} \leq P_{\max} \,\forall\, n = 1, \ldots, N\}$.

Capacity optimal resource allocation

Optimal power control and scheduling (centralized approach)

The SINR of the selected user in cell n is given by

$$\Gamma([\boldsymbol{K}]_n, \boldsymbol{P}) = \frac{G_{k_n,n} P_{k_n}}{\sigma^2 + \sum_{i \neq n}^{N} G_{k_n,i} P_{k_i}} \qquad (21)$$

where $G_{k_n,i} = |\alpha_{k_n,i}|^2$ is the channel power gain from cell i to receiver k_n.

The overall network capacity is given by:

$$\mathcal{C}(\boldsymbol{K}, \boldsymbol{P}) \triangleq \frac{1}{N} \sum_{n=1}^{N} \log\Big(1 + \Gamma([\boldsymbol{K}]_n, \boldsymbol{P})\Big) \qquad (22)$$

The capacity optimal resource allocation problem is:

$$(\boldsymbol{K}^*, \boldsymbol{P}^*) = \arg \max_{\substack{\boldsymbol{K} \in \Upsilon \\ \boldsymbol{P} \in \Omega}} \mathcal{C}(\boldsymbol{K}, \boldsymbol{P}) \qquad (23)$$

Towards distributed resource allocation

- The optimal R.A. problem is non-convex and exhaustive search is prohibitive.
- We want to maximize system capacity, by having each cell making an independent decision (power, user).

However ...

Key intuition: max sum-rate scheduling tends to increase the perceived SINR in each cell.

- If there are many users to choose from, one may pick users less sensitive to interference.
- How large is the gap in capacity due to interference, when the number of users is large ?

A bounding approach

We study two bounds on the capacity [RAWNET'07] :

- Upper bound obtained with no interference.
- Lower bound obtained with full-powered interference.

In two different network scenarios

- *symmetric*: all users have the same average received power (located on a circle around the BS).
- *non-symmetric*: users are uniformly located in the cell.

Upper bound on the capacity

- If we simply ignore the inter-cell interference:

$$\mathcal{C}(\boldsymbol{K}^*, \boldsymbol{P}^*) \leq \mathcal{C}^{ub} = \frac{1}{N}\sum_{n=1}^{N} \log\left(1 + \Gamma_n^{ub}\right). \qquad (24)$$

where the upper bound on SINR is given by:

$$\Gamma_n^{ub} = \max_{k_n=1..K}\{G_{k_n,n}\}P_{max}/\sigma^2 \qquad (25)$$

- In the absence of interference, the optimal capacity is reached by transmitting at a level equal to the power constraint $\rightarrow$ $\mathbf{P}_{max} = [P_{max}, .., P_{max}]$ + select the user with largest channel gain in each cell (*max SNR* scheduler) (*fully distributed*).

Lower bound on the capacity

- Assume now that all transmitters (interferers) transmit at P_{max}:

$$\mathcal{C}(\boldsymbol{K}^*, \boldsymbol{P}^*) \geq \mathcal{C}^{lb} = \mathcal{C}(\boldsymbol{K}^*_{FP}, \mathbf{P}_{max}) \tag{26}$$

where $\boldsymbol{K}^*_{FP}$ denotes the optimal scheduling vector assuming full power everywhere, defined by

$$[\boldsymbol{K}^*_{FP}]_n = \arg\max_{\boldsymbol{K}\in\Upsilon} \Gamma_n^{lb} \tag{27}$$

and Γ_n^{lb} is a lower bound on the best SINR given by:

$$\Gamma_n^{lb} = \max_{k_n=1..K} \frac{\{G_{k_n,n}\}P_{max}}{\sigma^2 + \sum_{i\neq n}^{N} G_{k_n,i}P_{max}} \tag{28}$$

- The corresponding scheduler is the *max SINR* scheduler (also *fully distributed*).

Capacity scaling with many users ($K \to \infty$)

Symmetric network - Interference-free case

Lemma: Let $G_{k_n,n} = \gamma_{k_n,n}|h_{k_n,i}|^2$, $k_n = 1..K, n = 1..N$, where $\gamma_{k_n,n} = \gamma_n$ (pathloss coefficient). Assume $|h_{k_n,n}|^2 \sim \chi^2(2)$ and i.i.d across users. Then for fixed N and K asymptotically large, the upper bound on the SINR in cell n scales as

$$\Gamma_n^{ub} \approx \frac{P_{max}\gamma_n}{\sigma^2} \log K \tag{29}$$

Theorem: For fixed N and K asymptotically large, the average of the upper bound on the network capacity scales as

$$\mathbb{E}(\mathcal{C}^{ub}) \approx \log \log K \tag{30}$$

where the expectation is taken over the complex fading gains.

Capacity scaling with many users ($K \to \infty$)

Symmetric network - Full-powered interference case

Lemma: Let $G_{k_n,i} = \gamma_{k_n,i}|h_{k_n,i}|^2$, $k_n = 1..K, n = 1..N$, where $\gamma_{k_n,n} = \gamma_n$, $\gamma_{k_n,i} = \beta d_{k_n,i}^{-\epsilon}$ (generic pathloss model, $\epsilon > 2$) for $i \neq n$. Assume $|h_{k_n,i}|^2 \sim \chi^2(2)$ and are i.i.d across users, cells. Then for fixed N and K asymptotically large, the lower bound on the SINR in cell n scales as

$$\Gamma_n^{lb} \approx \frac{P_{max}\gamma_n}{\sigma^2} \log K \tag{31}$$

Theorem: For fixed N and K asymptotically large, the average of the upper bound on the network capacity scales as

$$\mathbb{E}(\mathcal{C}^{ub}) \approx \log \log K \tag{32}$$

where the expectation is taken over the complex fading gains.

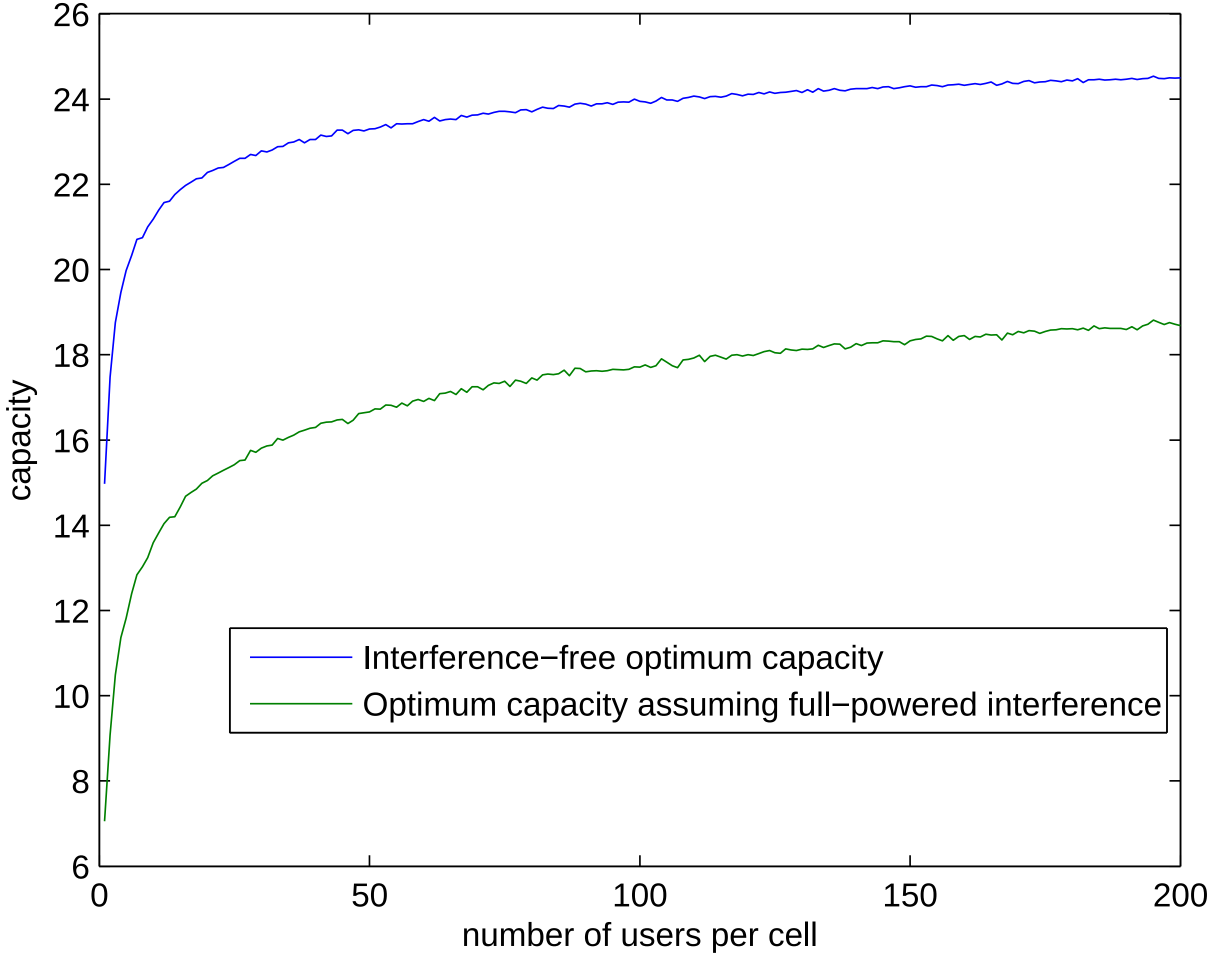
Capacity scaling with many users ($K \to \infty$)

- Upper bounds and lower bounds have same growth rates!
- Interference creates vanishing loss for large number of users.
- Physically, the scheduler looks for users shielded from interference and large SNR.
- When number of users is large, interference becomes small compared with noise.

Average capacity with optimum power control and scheduling scales like:

$$\mathbb{E}(\mathcal{C}(\boldsymbol{K}^*, \boldsymbol{P}^*)) \approx \log \log K \tag{33}$$

Capacity scaling for symmetric network



Scaling of upper and lower bounds of capacity, versus K for a symmetric network

$(N = 3,\ \beta = 1/16,\ \epsilon = 4,\ P_{max} = 1,\ \sigma^2 = 0.02).$

Capacity scaling for non-symmetric network

- We assume the pathloss is determined by the user's distance to the emitting AP (serving and interfering)
  $\rightarrow d_{k_n,n}$ is a random variable with non-uniform distribution
  $f_D(d) = 2d/R^2$, in $[0, R]$ (i.i.d. across users and cells, if users in each cell are dropped randomly in the network).
- Users close to BS have better SNR.

Assuming $R = 1$, the distribution of $\gamma_{k_n,n} = \beta d_{k_n,i}^{-\epsilon}$ is given by

$$f_\gamma(g) = \frac{2}{\epsilon}(\frac{g}{\beta})^{-\frac{2}{\epsilon}}\frac{1}{g} \text{ with } g \in [\beta, \infty) \quad (34)$$

and zero elsewhere.

Capacity scaling for non-symmetric network

Theorem [Breiman'65] Let X and Y two independent R.V. such that X is regularly varying with exponent $-a$, i.e. the cdf of X, $F_X(x)$, is such that $\frac{1-F_X(x)}{1-F_X(tx)} \rightarrow t^a$ when $t \rightarrow \infty$. Assuming Y has finite moment $E(Y^a)$, then the tail behavior of the product $Z = XY$ is governed by:

$$1 - F_Z(z) \rightarrow E(Y^a)(1 - F_X(z)) \text{ when } z \rightarrow \infty \tag{35}$$

Lemma Let $X = \gamma_{k_n,n}$ be a R.V. with distribution given by (34). Let Y be an independent R.V. such that $E(Y^{\frac{2}{\epsilon}}) < \infty$. Then the tail of $Z = XY$ is governed by:

$$1 - F_Z(z) \rightarrow E(Y^{\frac{2}{\epsilon}}) \left(\frac{\beta}{z}\right)^{\frac{2}{\epsilon}} \text{ when } z \rightarrow \infty \tag{36}$$

Capacity scaling for non-symmetric network

As the limiting distribution of Z is of Frechet type, we can show that:
Both the upper and lower bound on capacity behave like:

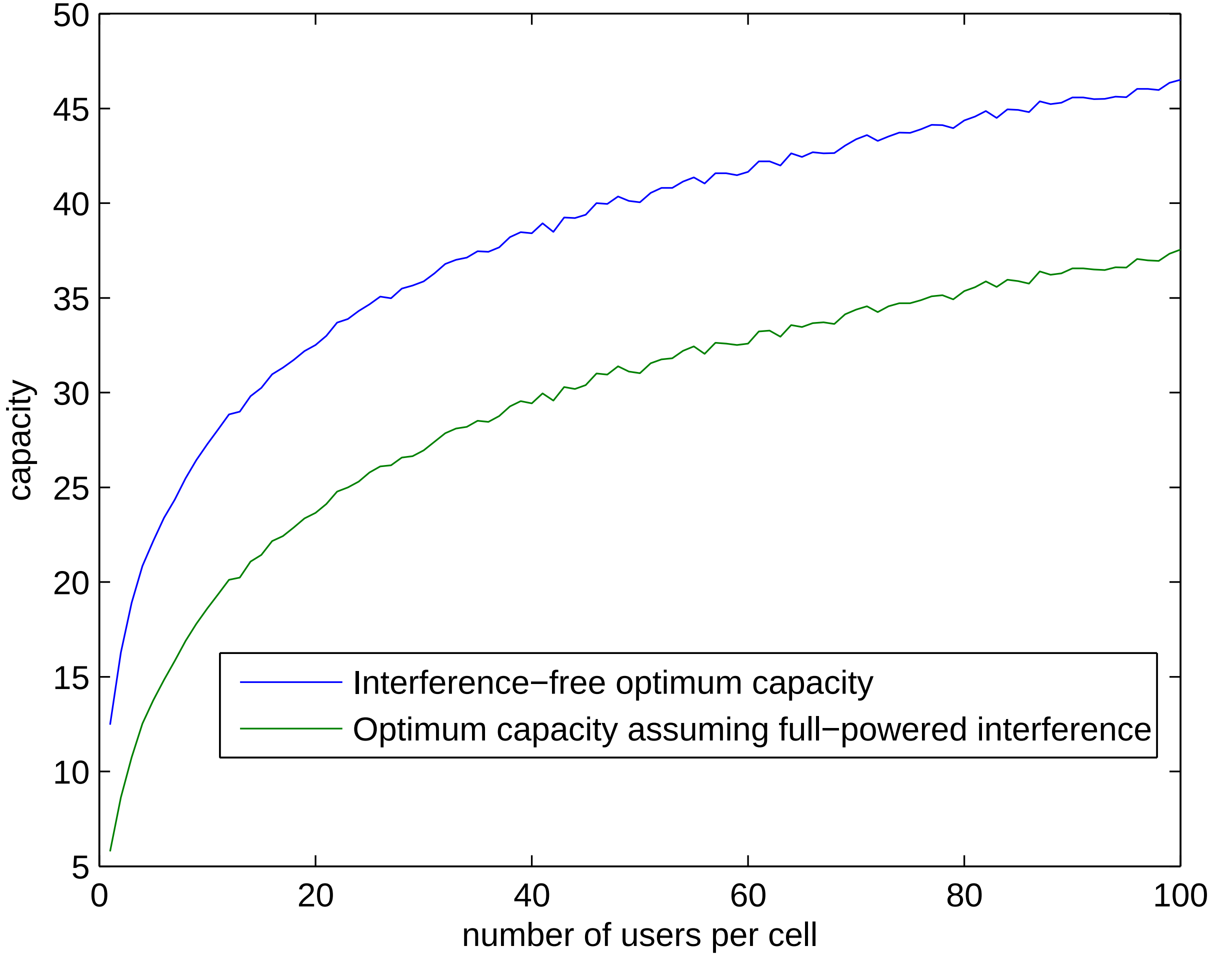
$$E(\mathcal{C}^{ub}) \approx \frac{\epsilon}{2} \log K \ \text{ for large } K \tag{37}$$

- multicell interference, no matter how strong, does not affect the scaling of the network capacity, if enough users exist *and* capacity optimal scheduling is applied.
- for a network with path loss-based average SNR, the maximum capacity behaves like

$$\mathcal{C}(\boldsymbol{K}^*, \boldsymbol{P}^*) \approx \frac{\epsilon}{2} \log K \tag{38}$$

- amplified multiuser diversity gain due to the presence of unequal path loss.

Capacity scaling for non-symmetric network



Scaling of upper and lower bounds of capacity versus K for a non-symmetric network

($N = 3$, $\beta = 1/16$, $\epsilon = 4$, $P_{max} = 1$, $\sigma^2 = 0.02$).

Conclusions

- Asymptotic analysis (in number of users) reveals simple structure of the resource allocation problem.
- Scheduling makes price paid in interference almost negligible.
- Dense networks can operate with reuse one, if the right scheduler is used!
- Scaling laws for upper and lower bounds for the network capacity (with $K \to \infty$) are derived, corresponding to two forms of distributed resource allocation schemes. These bounds are identical asymptotically.
- Two main problems:
  - fairness issues.
  - large number of users is required.

References

[Asilomar'06] M. Kountouris, R. de Francisco, D. Gesbert, D.T.M. Slock, and T. Sälzer, 'Multiuser diversity - multiplexing tradeoff in MIMO broadcast channels with limited feedback,' in *Proc. Asilomar'06*, Oct. 2006.

[ICASSP'06] D. Gesbert, L. Pittman, and M. Kountouris, 'Transmit Correlation-aided Scheduling in Multiuser MIMO Networks,' in *Proc. ICASSP'06*, May 2006.

[ICASSP'07] M. Kountouris, R. de Francisco, D. Gesbert, D.T.M. Slock, and T. Sälzer, 'Efficient metrics for scheduling in MIMO broadcast channels with limited feedback,' in *Proc. ICASSP'07*, April 2007. (see also journal version)

[ISIT'05] M. Kountouris, and D. Gesbert, 'Memory-based opportunistic multi-user beamforming,' in *Proc. ISIT'05*, Sept. 2005.

[EUSIPCO'06] M. Kountouris, D. Gesbert, and L. Pittman, 'Transmit correlation-aided opportunistic beamforming and scheduling,' in *Proc. EUSIPCO'06*, Sept. 2006.

[RAWNET'07] D. Gesbert, and M. Kountouris, 'Resource allocation in multicell wireless networks: Some capacity scaling laws,' in *Proc. IEEE WiOpt - RAWNET'07*, April 2007.

[SPAWC'05] M. Kountouris, and D. Gesbert, 'Robust multi-user opportunistic beamforming for sparse networks,' in *Proc. SPAWC'05*, June 2005. (see also journal version)

References

[Breiman'65] L. Breiman, 'On some limit theorems similar to the arc-sin law,' in *Theory of Probability and its applications*, vol. 10, 1965.

[Dimic et al.'05] G. Dimic and N.D. Sidiropoulos, 'On Downlink Beamforming with Greedy User Selection: Performance Analysis and a Simple New Algorithm,' in *IEEE Trans. Signal Processing*, vol. 53(10), Oct. 2005.

[Jindal'05] N. Jindal, 'MIMO Broadcast Channels with Finite-Rate Feedback,' in *IEEE Trans. Inform. Theory*, vol. 52(11), Nov. 2006.

[Sharif,Hassibi'05] M. Sharif and B. Hassibi, 'On the capacity of MIMO broadcast channel with partial side information,' in *IEEE Trans. Inform. Theory*, vol. 51(2), Febr. 2005.

[Viswanath et al.'02] P. Viswanath, D. Tse, and R. Laroia, 'Opportunistic beamforming using dumb antennas,' in *IEEE Trans. Inform. Theory*, vol. 48(6), June 2002.

[Yoo et al.'06] T. Yoo and A. Goldsmith, 'On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,' in *IEEE JSAC*, vol. 24(3), March 2006.