Exploiting Multiuser Diversity in MIMO Broadcast Channels with Limited Feedback

Marios Kountouris

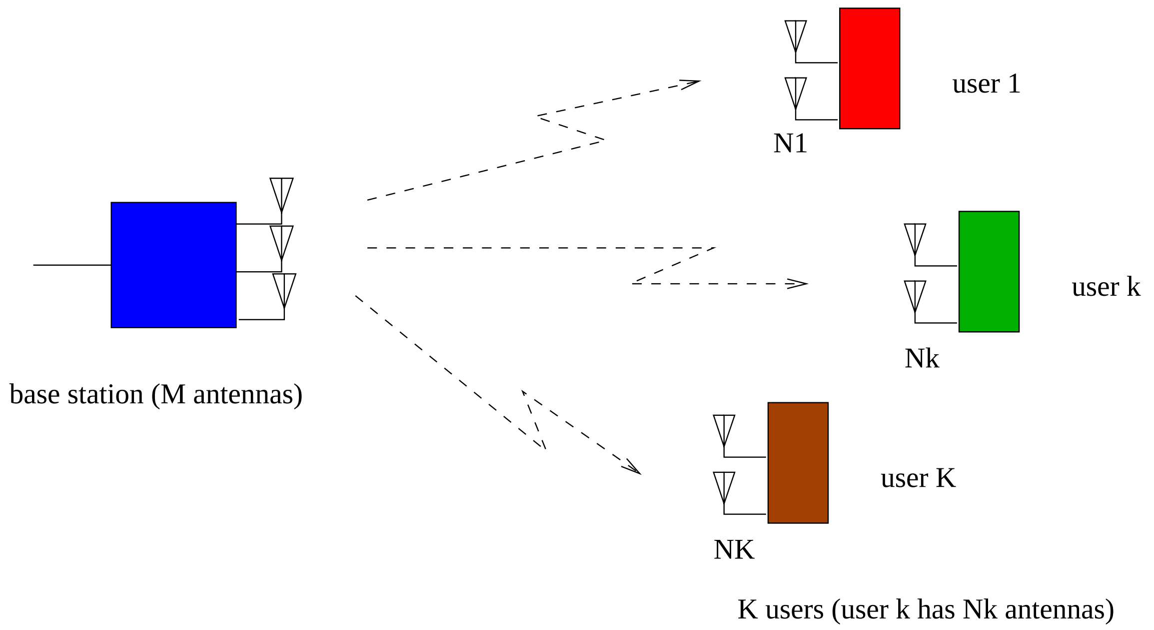
joint work with R. de Francisco, D. Gesbert,
D. T. M. Slock and T. Sälzer

Mobile Communications Dept. Seminar

Eurecom Institute

May 24^{th}, 2007

Context: downlink MU 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 (exploiting the spatial multiplexing capability of Tx antennas and transmit to 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.

The cardinal role of CSIT

- The capacity gain of multiuser MIMO systems seems to remain highly dependent on the channel state information available at the transmitter (CSIT).
  - if a BS with M transmit antennas 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!!
- However, full/perfect CSIT is unrealistic ...

Can we benefit from MUDiv and spatial multiplexing gains under limited feedback rate constraint ?

Effect of Limited Feedback in MIMO BC

In point-to-point systems:

- even a few feedback bits are sufficient for near-optimal performance.
- the level of CSIT only affects the SNR-offset; it does not affect the slope of the capacity vs. SNR curve (multiplexing gain).

However, in MIMO Broadcast Channels ...

- The level of CSIT critically affects the multiplexing gain of the MIMO downlink channel.
- It was recently shown that *in order to achieve full multiplexing gain* the number of FB bits/user must increase approximately linearly with the number of Tx antennas and the average SNR (in dB).

Feedback link design challenges

- The need for channel information is a limiting factor in multiuser MIMO systems.
- Such information must be fed back at the Tx, but the amount of feedback should be kept *minimal*.
- What kind and portion of partial CSIT is necessary and sufficient ?!
  - to identify "good" users to be scheduled.
  - to design efficient beamforming techniques (amplify the received signal and reduce the multiuser interference).
  - to achieve near optimal capacity growth.

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

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

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).

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

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

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

$$\mathbf{y}(\mathcal{S}) = \mathbf{H}(\mathcal{S})\mathbf{W}(\mathcal{S})\mathbf{P}\mathbf{s}(\mathcal{S}) + \mathbf{n} \tag{3}$$

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. $\mathbf{P}$ is a diagonal power allocation matrix with equal entries.

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

$$\mathbf{W}(\mathcal{S}) = \hat{\mathbf{H}}(\mathcal{S})^H(\hat{\mathbf{H}}(\mathcal{S})\hat{\mathbf{H}}(\mathcal{S})^H)^{-1}\mathbf{\Lambda} \tag{4}$$

where $\hat{\mathbf{H}}(\mathcal{S}) = \left[\hat{\mathbf{h}}_{k_1}\ \hat{\mathbf{h}}_{k_2} \ldots \hat{\mathbf{h}}_{k_\mathcal{M}}\right]^H$ with $\{k_i\}_{i=1}^{\mathcal{M}} \in \mathcal{S}$.

Zero Forcing (ZF) Beamforming (2/2)

- The SINR at the k-th receiver is

$$SINR_k = \frac{P_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum\limits_{j \in \mathcal{S}-\{k\}} P_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + 1} \tag{5}$$

  where $\sum_{i \in \mathcal{S}} P_i = P$ in order to satisfy the power constraint on the transmitted signal.
  For simplicity, equal power allocation is considered, i.e. $P_i = \frac{P}{\mathcal{M}}, \forall i \in \mathcal{S}$.

- The expected sum rate, assuming Gaussian inputs, is equal to

$$\mathcal{R} = \mathbb{E}\left\{ \sum_{k \in \mathcal{S}} \log\left(1 + SINR_k\right) \right\} \tag{6}$$

CQI Finite Rate Model

Efficient Feedback Metrics exploiting 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.
- We consider that each user provides information on its effective channel (SINR) by feeding back the following scalar metric

$$\xi_k^{UB} = \frac{P\left\|\mathbf{h}_k\right\|^2 \cos^2\phi_k}{P\left\|\mathbf{h}_k\right\|^2 \sin^2\phi_k + M} \tag{7}$$

- 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 <u>*at the mobile side*</u> (without cooperation).

Metric II: Lower Bound on SINR (1/3)

Bounding the actual multiuser interference

- Let $\cos\theta_k = \left|\bar{\mathbf{h}}_k^H \mathbf{w}_k\right|$, $\mathbf{\Psi}_k(\mathcal{S}) = \sum_{j\in S, j\neq k} \mathbf{w}_j \mathbf{w}_j^H$, the operator $\lambda_{max}\{\cdot\}$, which returns the largest eigenvalue, and $\mathbf{U}_k \in \mathbb{C}^{M\times(M-1)}$ an orthonormal basis spanning the null space of $\mathbf{w}_k$.
- *Theorem 1: Given a set $\mathcal{S}$ of unit-norm BF vectors, an upper bound on the interference over the normalized channel $\bar{I}_k(\mathcal{S}) = \sum_{j\in\mathcal{S}, j\neq k} \left|\bar{\mathbf{h}}_k^H \mathbf{w}_j\right|^2$ is given by*

$$\bar{I}_k(\mathcal{S}) \leq \cos^2\theta_k \alpha_k(\mathcal{S}) + \sin^2\theta_k \beta_k(\mathcal{S}) + 2\sin\theta_k\cos\theta_k\gamma_k(\mathcal{S}) \quad (8)$$

where

$$\begin{cases} \alpha_k(\mathcal{S}) = \mathbf{w}_k^H \mathbf{\Psi}_k(\mathcal{S})\mathbf{w}_k \\ \beta_k(\mathcal{S}) = \lambda_{max}\left\{\mathbf{U}_k^H \mathbf{\Psi}_k(\mathcal{S})\mathbf{U}_k\right\} \\ \gamma_k(\mathcal{S}) = \left\|\mathbf{U}_k^H \mathbf{\Psi}_k(\mathcal{S})\mathbf{w}_k\right\| \end{cases} \quad (9)$$

Metric II: Lower Bound on SINR (2/3)

- We now impose an ϵ-orthogonality constraint between $\hat{\mathbf{h}}_k$ and $\epsilon_{ZF} = \max_{i,j\in\mathcal{S}} \left|\mathbf{w}_i^H \mathbf{w}_j\right|$ (worst-case orthogonality under ZFBF).
- *Lemma 1: The orthogonality of the set of M normalized ZFBF vectors (ϵ_{ZF}) and alignment with the normalized channel ($\cos\theta_k$) are bounded as:*

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

- *Theorem 2: 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*

$$SINR_k^{ZF} \geq \frac{P\left\|\mathbf{h}_k\right\|^2 \cos^2\theta_k}{P\left\|\mathbf{h}_k\right\|^2 \overline{I}_{UB_k} + M} \tag{10}$$

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

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

Metric II: Lower Bound on SINR (3/3)

- Motivated by the above results, we propose that each user provides information on its SINR lower bound (LB) and reports the following scalar metric

$$\xi_k^{LB} = \frac{\frac{P}{(1+\vartheta)^2} \left\| \mathbf{h}_k \right\|^2 (\cos\phi_k - \sqrt{\vartheta})^2}{P \left\| \mathbf{h}_k \right\|^2 \overline{I}_{UB_k} + M} \tag{11}$$

Metric III: Decoupling 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: Decoupling 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

$$\xi_k^{LBd} = \frac{P \left\| \mathbf{h}_k \right\|^2 \rho_k^2}{P \left\| \mathbf{h}_k \right\|^2 \bar{I}_{UBd_k} + \mathcal{M}} \tag{12}$$

  where $\bar{I}_{UBd_k} = \rho_k^2 \alpha_k(\mathcal{S}) + \left(1 - \rho_k^2\right) \beta_k(\mathcal{S}) + 2\rho_k \sqrt{1-\rho_k^2}\gamma_k(\mathcal{S})$ can be explicitly calculated at the transmitter using (9), and $\rho_k^2 = \cos^2(\phi_k + \angle(\hat{\mathbf{h}}_k, \mathbf{w}_k))$.

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

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

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: $k_1 = \arg\max_{k \in \mathcal{Q}^0} \xi_k$

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 (1/4)

Asymptotic sum rate in the large K regime

- We analyze the sum-rate performance $\mathcal{R}$ of a system using the Greedy-SUS algorithm in conjunction with metric II at large K and M fixed.
- The sum rate is lower bounded by:

$$\mathcal{R} \geq M\left(1 - \Pr\left\{|\mathcal{S}| \neq M\right\}\right)\log_2\left(1 + \min_{k\in\mathcal{S}} \xi_k^{LB}\right) \tag{14}$$

- Choosing $\epsilon = \frac{1}{\log K}$, so that $\lim_{K\to\infty}\left|\mathcal{Q}^i\right| = \infty$ and $\lim_{K\to\infty}\epsilon = 0$, then $\xi_k^{LB} \overset{K\to\infty}{\to} \xi_k^{UB}$. Denoting $\beta = 2^{-B}\left(P/M\right)^{M-1}$, we can show:
- *Theorem 3:* The sum rate $\mathcal{R}$ converges to the optimum capacity of full CSI case $\mathcal{R}_{opt}$, for $K \to \infty$, i.e.,

$$\lim_{K\to\infty}\left(\mathcal{R}_{opt} - \mathcal{R}\right) = \lim_{K\to\infty}\left[M\log_2\frac{1+\frac{P}{M}\log K}{1+\frac{P}{M}\log\left(K/\beta\right)}\right] = 0 \tag{15}$$

with probability one.

Performance analysis (2/4)

Sum rate in the interference-limited region

- We characterize the sum-rate performance of our scheme in the high-power regime ($P \to \infty$).
- *Theorem 4:* The sum rate of the proposed scheme at high SNR with finite B and K is upper bounded by

$$\mathcal{R} \leq \frac{M}{M-1}\left(B + \frac{log_2 e}{\kappa_{max}} H_K\right) \quad (16)$$

where $H_K = \sum_{k=1}^{K} \frac{1}{k}$ is the harmonic number and $\kappa_{max} = \max_{i=1,\ldots,M} \kappa_{i-1}$.

$\kappa_i = \left|\mathcal{Q}^i\right| / K \approx I_{\epsilon^2}(i, M-i)$ (regularized incomplete beta function) captures the multiuser diversity gain reduction due to greedy scheduling.

Performance analysis (3/4)

Sum rate in the interference-limited region

- The system becomes interference-limited $\rightarrow$ bounded sum rate at high SNR (even for arbitrary large but finite B and K).
- As $\partial\mathcal{R}/\partial M < 0 \rightarrow \mathcal{R}$ is monotonically decreasing with M $\rightarrow M = 1$ at high SNR (switch from SDMA to TDMA).
- For large K, we have $\lim_{K\rightarrow\infty} H_K = \log K + \gamma$
  (γ: Euler-Mascheroni constant)
- The sum rate with $P \rightarrow \infty$ and $K \rightarrow \infty$ exhibits logarithmic growth with $K \rightarrow$ multiuser diversity compensates for the loss in degrees of freedom.

Performance analysis (4/4)

Sum rate in the low-power regime

- When $P \to 0$, $\xi_k^{LB} = \frac{P}{M(1+\vartheta)^2} \|\mathbf{h}_k\|^2 (\cos\phi_k - \sqrt{\vartheta})^2 \leq \frac{P}{M} \|\mathbf{h}_k\|^2 \cos^2\phi_k$.
- *Lemma 2*: The distribution of $X = \|\mathbf{h}_k\|^2 \cos^2\phi_k$ is given by

$$f_X(x) = (1-\delta) \sum_{k=0}^{\infty} \frac{\zeta_k x^{k+M-1} e^{-x/(1-\delta)}}{(1-\delta)^{k+M} \Gamma(k+M)} u(x) \tag{17}$$

where $u(\cdot)$ is the unit step function, $\delta = 2^{-\frac{B}{M-1}}$, and ζ_k is obtained recursively by

$$\zeta_0 = 1 \;\text{ and }\; \zeta_{k+1} = \frac{1}{k+1} \sum_{i=1}^{k+1} \left(\delta^i\right) \zeta_{k+1-i}, k = 0, 1, 2, \ldots \tag{18}$$

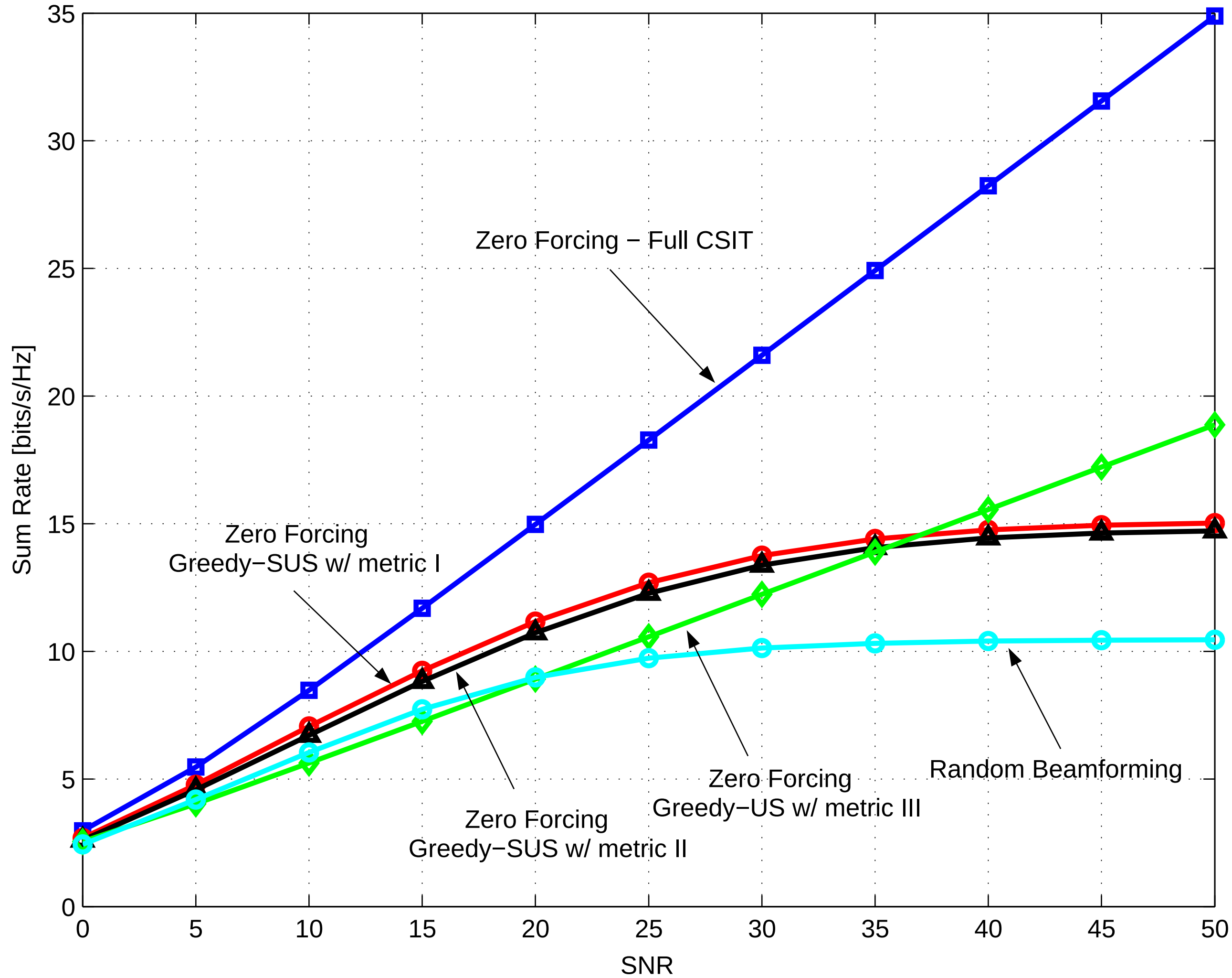
- *Proposition*: The r.v X stochastically dominates $\tilde{X}$ ($F_X(x) \leq F_{\tilde{X}}(x)$), whose CDF is given by

$$F_{\tilde{X}}(x) = \int_0^x (1 - e^{-x/(1-\delta)})^{M-1} dx \tag{19}$$

- For finite B, the sum rate is lower bounded as

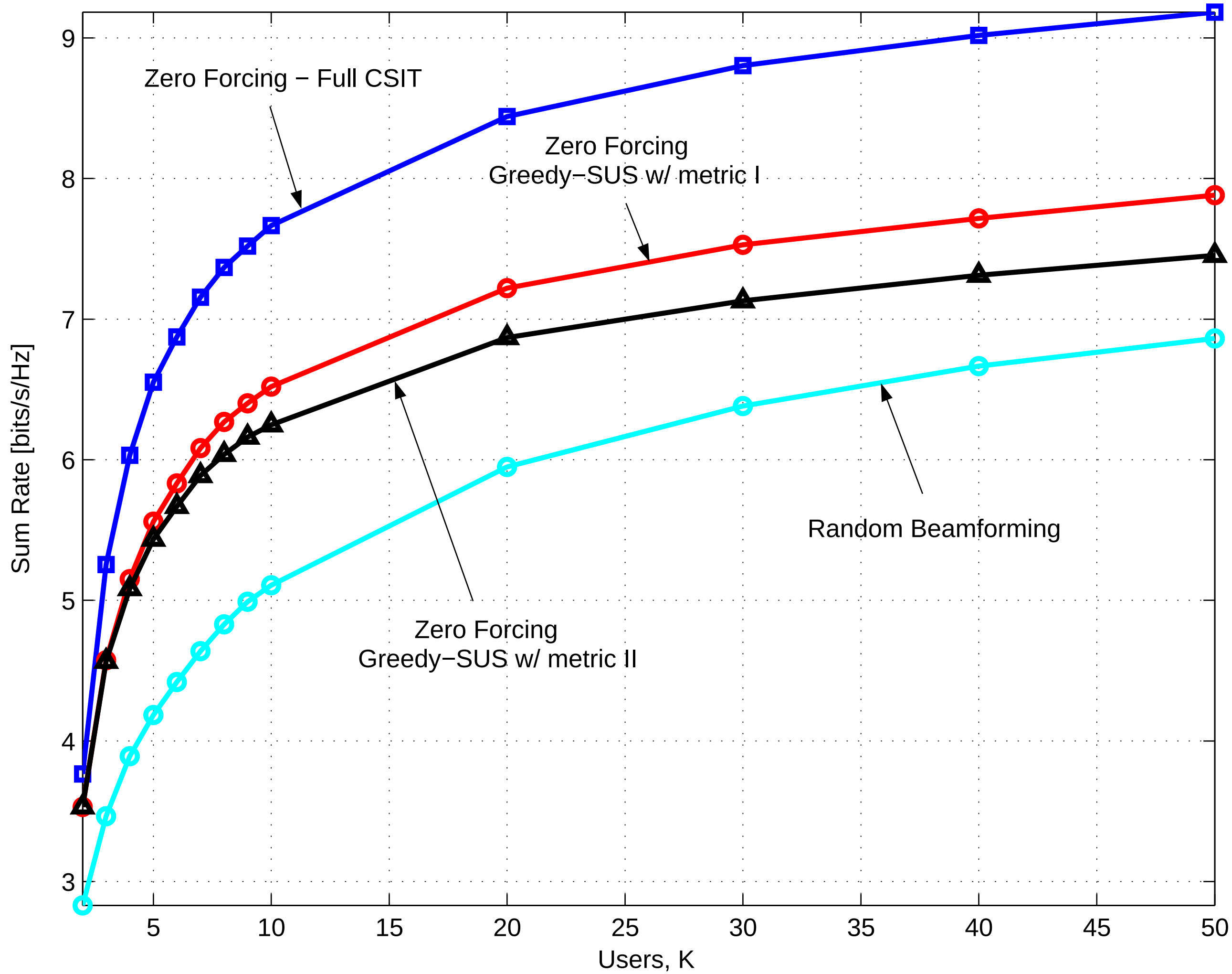
$$\mathcal{R} \geq \sum_{i=1}^{M} \int_0^\infty \log_2\left(1 + \frac{P}{M}x\right) dF_{\tilde{X}}^{\mathcal{K}_i}(x) \approx \log_2 e \frac{P}{M} \sum_{i=1}^{M} \int_0^\infty x dF_{\tilde{X}}^{\mathcal{K}_i}(x) \tag{20}$$

Numerical Results (1)



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

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.

In brief ...

- We studied a multiple antenna broadcast channel in which partial CSIT is conveyed via a limited rate feedback channel.
- Several scalar feedback metrics have been proposed, which provide an efficient estimate of the received SINR at the receiver side.
- These metrics, combined with greedy scheduling and ZFBF on the quantized channel directions, can achieve a significant fraction of full CSI capacity by means of multiuser diversity.
- We derived upper bounds on the instantaneous MU interference that allows us to analytically predict the worst case interference and SINR in a system employing ZF on the quantized channel directions.
- A multi-mode switching scheme from MU to SU transmission is proposed. This scheme exhibits a linear sum-rate growth at high SNR.

References

[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.

[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.