Joint Space-Division and Multiplexing: How to Achieve Massive MIMO Gains in FDD Systems

Giuseppe Caire
University of Southern California, Viterbi School of Engineering, Los Angeles, CA

Phuket, Thailand, June 23-26, 2013
Channel estimation bottleneck on MU-MIMO

- High-SNR capacity of $N_t \times N_r$ single-user MIMO with coherence block-length $T$ [Zheng-Tse, 2003]:

$$C(\text{SNR}) = M^* (1 - M^*/T) \log \text{SNR} + O(1), \quad M^* = \min\{N_t, N_r, T/2\}$$

- Trivial cooperative bound: for large $M = N_t$ and $N = K N_r$, the coherence block $T$ is the limiting factor.

- Disappointing theoretical performance of “CoMP” (base station cooperation), in FDD.

![Diagram of Inter-cell Cooperation]
In FDD, for large macro-cellular base stations, we have to exploit channel dimensionality reduction while still exploiting the large number of antennas at the BS.

Idea: exploit the asymmetric spatial channel correlation at the BS and at the UTs.

Isotropic scattering, \( |\mathbf{u} - \mathbf{u}'| = \lambda D \):

\[
\mathbb{E} [h(\mathbf{u})h^*(\mathbf{u}')] = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j2\pi D \cos(\alpha)} d\alpha = J_0(2\pi D)
\]

Two users separated by a few meters (say 10 \( \lambda \)) are practically uncorrelated.
• In contrast, the base station sees user groups at different AoAs under narrow AS $\Delta \approx \arctan(r/s)$.

• This leads to the Tx antenna correlation model

$$h = U\Lambda^{1/2}w, \quad R = U\Lambda U^H$$

with

$$[R]_{m,p} = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{j\mathbf{k}^T(\alpha+\theta)(\mathbf{u}_m-\mathbf{u}_p)} d\alpha.$$
Joint Space Division and Multiplexing (JSDM)

- $K$ users selected to form $G$ groups, with $\approx$ same channel correlation.
  \[
  \mathbf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_G], \text{ with } \mathbf{H}_g = \mathbf{U}_g \mathbf{\Lambda}_g^{1/2} \mathbf{W}_g.
  \]

- Two-stage precoding: $\mathbf{V} = \mathbf{B}\mathbf{P}$.

- $\mathbf{B} \in \mathbb{C}^{M \times b_g}$ is a pre-beamforming matrix function of $\{\mathbf{U}_g, \mathbf{\Lambda}_g\}$ only.

- $\mathbf{P} \in \mathbb{C}^{b_g \times S_g}$ is a precoding matrix that depends on the effective channel.

- The effective channel matrix is given by
  \[
  \mathbf{H}^H = \begin{bmatrix}
  \mathbf{H}_1^H\mathbf{B}_1 & \mathbf{H}_1^H\mathbf{B}_2 & \cdots & \mathbf{H}_1^H\mathbf{B}_G \\
  \mathbf{H}_2^H\mathbf{B}_1 & \mathbf{H}_2^H\mathbf{B}_2 & \cdots & \mathbf{H}_2^H\mathbf{B}_G \\
  \vdots & \vdots & \ddots & \vdots \\
  \mathbf{H}_G^H\mathbf{B}_1 & \mathbf{H}_G^H\mathbf{B}_2 & \cdots & \mathbf{H}_G^H\mathbf{B}_G 
  \end{bmatrix}.
  \]
• **Per-Group Processing:** If estimation and feedback of the whole $\mathbf{H}$ is still too costly, then each group estimates its own diagonal block $\mathbf{H}_g = \mathbf{B}_g^H \mathbf{H}_g$, and $\mathbf{P} = \text{diag}(\mathbf{P}_1, \cdots, \mathbf{P}_G)$.

• This results in

$$y_g = \mathbf{H}_g^H \mathbf{B}_g \mathbf{P}_g \mathbf{d}_g + \sum_{g' \neq g} \mathbf{H}_g^H \mathbf{B}_{g'} \mathbf{P}_{g'} \mathbf{d}_{g'} + \mathbf{z}_g$$
Achieving capacity with reduced CSIT

- Let \( r = \sum_{g=1}^{G} r_g \) and suppose that the channel covariances of the \( G \) groups are such that \( \mathbf{U} = [\mathbf{U}_1, \ldots, \mathbf{U}_G] \) is \( M \times r \) tall unitary (i.e., \( r \leq M \) and \( \mathbf{U}^H \mathbf{U} = \mathbf{I}_r \)).

- Eigen-beamforming (let \( b_g = r_g \) and \( \mathbf{B}_g = \mathbf{U}_g \)) achieves exact block diagonalization.

- The decoupled MU-MIMO channel takes on the form

\[
\mathbf{y}_g = \mathbf{H}_g^H \mathbf{P}_g \mathbf{d}_g + \mathbf{z}_g = \mathbf{W}_g^H \Lambda_g^{1/2} \mathbf{P}_g \mathbf{d}_g + \mathbf{z}_g, \quad \text{for } g = 1, \ldots, G,
\]

where \( \mathbf{W}_g \) is a \( r_g \times K_g \) i.i.d. matrix with elements \( \sim \mathcal{CN}(0, 1) \).

**Theorem 1.** For \( \mathbf{U} \) tall unitary, JSDM with PGP achieves the same sum capacity of the corresponding MU-MIMO downlink channel with full CSIT.
Block Diagonalization

- For given target numbers of streams per group \( \{S_g\} \) and dimensions \( \{b_g\} \) satisfying \( S_g \leq b_g \leq r_g \), we can find the pre-beamforming matrices \( B_g \) such that:
  \[
  U_{g'}^H B_g = 0 \quad \forall \ g' \neq g, \quad \text{and} \quad \text{rank}(U_g^H B_g) \geq S_g
  \]

- Necessary condition for exact BD
  \[
  \text{Span}(B_g) \subseteq \text{Span}^\perp(\{U_{g'} : g' \neq g\}).
  \]

- When \( \text{Span}^\perp(\{U_{g'} : g' \neq g\}) \) has dimension smaller than \( S_g \), the rank condition on the diagonal blocks cannot be satisfied.

- In this case, \( S_g \) should be reduced (reduce the number of served users per group) or, as an alternative, approximated BD based on selecting \( r_g^* < r_g \) dominant eigenmodes for each group \( g \) can be implemented.
The transformed channel matrix $\mathbf{H}$ has dimension $b \times S$, with blocks $\mathbf{H}_g$ of dimension $b_g \times S_g$.

For simplicity we allocate to all users the same fraction of the total transmit power, $p_{g_k} = \frac{P}{S}$.

For PGP, the regularized zero forcing (RZF) precoding matrix for group $g$ is given by

$$
P_{g,\text{rzf}} = \bar{\zeta}_g \bar{K}_g \mathbf{H}_g,$$

where

$$
\bar{K}_g = \left[ \mathbf{H}_g \mathbf{H}_g^H + b_g \alpha \mathbf{I}_{b_g} \right]^{-1}
$$

and where

$$
\bar{\zeta}_g^2 = \frac{S'}{\text{tr} \left( \mathbf{H}_g^H \mathbf{K}_g^H \mathbf{B}_g \mathbf{B}_g \mathbf{K}_g \mathbf{H}_g \right)}.
$$
• The SINR of user $g_k$ given by

$$\gamma_{g_k,\text{pgp}} = \frac{P \bar{\zeta}_g |h_{g_k}^H B_g \bar{K}_g B_h^H h_{g_k}|^2}{P \bar{\zeta}_g |h_{g_k}^H B_g \bar{K}_g B_h^H h_{g_k}|^2 + \frac{P}{S} \sum_{j \neq k} \bar{\zeta}_g |h_{g_k}^H B_g \bar{K}_g B_h^H h_{g_j}|^2 + \frac{P}{S} \sum_{g' \neq g} \sum_{j} \bar{\zeta}_g |h_{g_k}^H B_{g'} \bar{K}_{g'} B_h^H h_{g_j}|^2 + 1}$$

• Using the “deterministic equivalent” method of [Wagner, Couillet, Debbah, Slock, 2011], we can calculate $\gamma_{g_k,\text{pgp}}^o$ such that

$$\gamma_{g_k,\text{pgp}} - \gamma_{g_k,\text{pgp}}^o \xrightarrow{M \to \infty} 0$$
• $M = 100$, $G = 6$ user groups, $\text{Rank}(\mathbf{R}_g) = 21$, effective rank $r_g^* = 11$.

• We serve $S' = 5$ users per group with $b' = 10$, $r^* = 6$ and $r^* = 12$.

• For $r_g^* = 12$: 150 bit/s/Hz at SNR $= 18$ dB: 5 bit/s/Hz per user, for 30 users served simultaneously on the same time-frequency slot.
Training, Feedback and Computations Requirements

- **Full CSI:** $100 \times 30$ channel matrix $\Rightarrow$ 3000 complex channel coefficients per coherence block (CSI feedback), with $100 \times 100$ unitary “common” pilot matrix for downlink channel estimation.

- **JSDM with PGP:** $6 \times 10 \times 5$ diagonal blocks $\Rightarrow$ 300 complex channel coefficients per coherence block (CSI feedback), with $10 \times 10$ unitary “dedicated” pilot matrices for downlink channel estimation, sent in parallel to each group through the pre-beamforming matrix.

- One order of magnitude saving in both downlink training and CSI feedback.

- **Computation:** 6 matrix inversions of dimension $5 \times 5$, with respect to one matrix inversion of dimension $30 \times 30$. 
Non-ideal CSIT

• Parallel downlink training in all groups: a scaled unitary training matrix $X_{tr}$ of dimension $b' \times b'$ is sent, simultaneously, to all groups in the common downlink training phase.

• Received signal at group $g$ receivers is given by

$$Y_g = H_g^H X_{tr} + \sum_{g' \neq g} H_{g'}^H B_{g'} X_{tr} + Z_g.$$ 

• Multiplying from the right by $X_{tr}^H$ and letting $\rho_{tr}$ denote the power allocated to training, we obtain

$$Y_g X_{tr}^H = \rho_{tr} H_g^H + \rho_{tr} \sum_{g' \neq g} H_{g'}^H B_{g'} + Z_g X_{tr}^H.$$
• The relevant observation for the $g_k$-th user effective channel is:

$$\tilde{h}_{g_k} = \sqrt{\rho_{\text{tr}}} h_{g_k} + \sqrt{\rho_{\text{tr}}} \left( \sum_{g' \neq g} B_{g'}^H \right) h_{g_k} + \tilde{z}_{g_k}. $$

• The corresponding MMSE estimator is given by

$$\hat{h}_{g_k} = E \left[ h_{g_k} \tilde{h}_{g_k}^H \right] E \left[ \tilde{h}_{g_k} \tilde{h}_{g_k}^H \right]^{-1} \tilde{h}_{g_k}$$

$$= \sqrt{\rho_{\text{tr}}} \left[ B_g^H R_g \sum_{g'=1}^G B_{g'} \right] \left[ \rho_{\text{tr}} \sum_{g', g''=1}^G B_{g'}^H R_g B_{g''} + I_{b'} \right]^{-1} \tilde{h}_{g_k}$$

$$= \frac{1}{\sqrt{\rho_{\text{tr}}}} \left( M_g \tilde{R}_g O^T \right) \left[ O \tilde{R}_g O^T + \frac{1}{\rho_{\text{tr}}} I_{b'} \right]^{-1} \tilde{h}_{g_k}$$
where we used the fact that $h_{gk} = B^H_g h_{gk}$, and we introduced the $b' \times b$ block matrices

$$M_g = [0, \ldots, 0, \underbrace{I_{b'}}_{\text{block } g}, 0, \ldots, 0]$$

$$O = [I_{b'}, I_{b'}, \ldots, I_{b'}].$$

- Notice that in the case of perfect BD we have that $R_g B_{g'} = 0$ for $g' \neq g$. Therefore, the MMSE estimator reduces to

$$\hat{h}_{gk} = \frac{1}{\sqrt{\rho_{tr}}} \bar{R}_g \left[ \bar{R}_g + \frac{1}{\rho_{tr}} I_{b'} \right]^{-1} \tilde{h}_{gk}$$

where $\bar{R}_g = B^H_g R_g B_g$. 


• Also in this case, the deterministic equivalent approximations of the SINR terms for RZFBF and ZFBF precoding can be computed.

• Eventually, the achievable rate of user $g_k$ is given by

\[
R_{g_k,pgp,csit} = \max \left\{ 1 - \frac{b'}{T}, 0 \right\} \times \log \left( 1 + \hat{\gamma}^o_{g_k,pgp,csit} \right).
\]
• $b'$ large yields better conditioned matrices, but it “costs” more in terms of training phase dimension.

(a) $S' = 4, \text{SNR} = 10\text{dB}$  
(b) $S' = 8, \text{SNR} = 30\text{dB}$
Impact of non-ideal CSIT

(c) $S' = 4$

(d) $S' = 8$
Discussion: is the tall unitary realistic?

• For a Uniform Linear Array (ULA), $\mathbf{R}$ is Toeplitz, with elements

$$[\mathbf{R}]_{m,p} = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{-j2\pi D(m-p)\sin(\alpha+\theta)} d\alpha, \quad m, p \in \{0, 1, \ldots, M - 1\}$$

• We are interested in calculating the asymptotic rank, eigenvalue CDF and structure of the eigenvectors, for $M$ large, for given geometry parameters $D, \theta, \Delta$.

• Correlation function

$$r_m = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{-j2\pi Dm\sin(\alpha+\theta)} d\alpha.$$
• As $M \to \infty$, the eigenvalues of $\mathbf{R}$ tend to the “power spectral density” (i.e., the DT Fourier transform of $r_m$),

$$S(\xi) = \sum_{m=-\infty}^{\infty} r_m e^{-j2\pi \xi m}$$

sampled at $\xi = k/M$, for $k = 0, \ldots, M - 1$.

• After some algebra, we arrive at

$$S(\xi) = \frac{1}{2\Delta} \sum_{m \in [D \sin(-\Delta+\theta) + \xi, D \sin(\Delta+\theta) + \xi]} \frac{1}{\sqrt{D^2 - (m - \xi)^2}}.$$
Theorem 2. The empirical spectral distribution of the eigenvalues of $\mathbf{R}$,

$$F_{\mathbf{R}}^{(M)}(\lambda) = \frac{1}{M} \sum_{m=1}^{M} 1\{\lambda_m(\mathbf{R}) \leq \lambda\},$$

converges weakly to the limiting spectral distribution

$$\lim_{M \to \infty} F_{\mathbf{R}}^{(M)}(\lambda) = F(\lambda) = \int_{S(\xi) \leq \lambda} d\xi.$$
Example: $M = 400$, $\theta = \pi/6$, $D = 1$, $\Delta = \pi/10$. Exact empirical eigenvalue cdf of $\mathbf{R}$ (red), its approximation the circulant matrix $\mathbf{C}$ (dashed blue) and its approximation from the samples of $S(\xi)$ (dashed green).
A less well-known Szego’s Theorem: eigenvectors

Theorem 3. Let $\lambda_0(R) \leq \ldots \leq \lambda_{M-1}(R)$ and $\lambda_0(C) \leq \ldots \leq \lambda_{M-1}(C)$ denote the set of ordered eigenvalues of $R$ and $C$, and let $U = [u_0, \ldots, u_{M-1}]$ and $F = [f_0, \ldots, f_{M-1}]$ denote the corresponding eigenvectors. For any interval $[a, b] \subseteq [\kappa_1, \kappa_2]$ such that $F(\lambda)$ is continuous on $[a, b]$, consider the eigenvalues index sets $I_{[a,b]} = \{m : \lambda_m(R) \in [a, b]\}$ and $J_{[a,b]} = \{m : \lambda_m(C) \in [a, b]\}$, and define $U_{[a,b]} = (u_m : m \in I_{[a,b]})$ and $F_{[a,b]} = (f_m : m \in J_{[a,b]})$ be the submatrices of $U$ and $F$ formed by the columns whose indices belong to the sets $I_{[a,b]}$ and $J_{[a,b]}$, respectively. Then, the eigenvectors of $C$ approximate the eigenvectors of $R$ in the sense that

$$\lim_{M \to \infty} \frac{1}{M} \left\| U_{[a,b]} U_{[a,b]}^H - F_{[a,b]} F_{[a,b]}^H \right\|_F^2 = 0.$$

Consequence 1: $U_g$ is well approximated by a “slice” of the DFT matrix.

Consequence 2: DFT pre-beamforming is near optimal for large $M$. 
Theorem 4. The asymptotic normalized rank of the channel covariance matrix $R$, with antenna separation $\lambda D$, AoA $\theta$ and AS $\Delta$, is given by

$$\rho = \min\{1, B(D, \theta, \Delta)\},$$

with $B(D, \theta, \Delta) = |D \sin(-\Delta + \theta) - D \sin(\Delta + \theta)|$.

Theorem 5. Groups $g$ and $g'$ with angle of arrival $\theta_g$ and $\theta_{g'}$ and common angular spread $\Delta$ have spectra with disjoint support if their AoA intervals $[\theta_g - \Delta, \theta_g + \Delta]$ and $[\theta_{g'} - \Delta, \theta_{g'} + \Delta]$ are disjoint.
• ULA with $M = 400$, $G = 3$, $\theta_1 = -\frac{\pi}{4}$, $\theta_2 = 0$, $\theta_3 = \frac{\pi}{4}$, $D = 1/2$ and $\Delta = 15$ deg.
Super-Massive MIMO
• Idea: produce a 3D pre-beamforming by Kronecker product of a “vertical” beamforming, separating the sector into $L$ concentric regions, and a “horizontal” beamforming, separating each $\ell$-th region into $G_\ell$ groups.

• Horizontal beam forming is as before.

• For vertical beam forming we just need to find one dominating eigenmode per region, and use the BD approach.

• A set of simultaneously served groups forms a “pattern”.

• Patterns need not cover the whole sector.

• Different intertwined patterns can be multiplexed in the time-frequency domain in order to guarantee a fair coverage.
An example

- Cell radius 600m, group ring radius 30m, array height 50m, $M = 200$ columns, $N = 300$ rows.
- Pathloss $g(x) = \frac{1}{1 + \left(\frac{x}{d_0}\right)^\delta}$ with $\delta = 3.8$ and $d_0 = 30$ m.
- Same color regions are served simultaneously. Each ring is given equal power.
### Sum throughput (bit/s/Hz) under PFS and Max-min Fairness

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Approximate BD</th>
<th>DFT based</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFS, RZFBF</td>
<td>1304.4611</td>
<td>1067.9604</td>
</tr>
<tr>
<td>PFS, ZFBF</td>
<td>1298.7944</td>
<td>1064.2678</td>
</tr>
<tr>
<td>MAXMIN, RZFBF</td>
<td>1273.7203</td>
<td>1042.1833</td>
</tr>
<tr>
<td>MAXMIN, ZFBF</td>
<td>1267.2368</td>
<td>1037.2915</td>
</tr>
</tbody>
</table>

1000 bit/s/Hz × 40 MHz of bandwidth = 40 Gb/s per sector.
Our on-going work

- Compatibility with an in-band Small-Cell tier: eICIC in the spatial domain: turn on and off the “spotbeams”.

- Multi-cell strategies: activate mutually compatible patterns of groups in adjacent sectors.

- User grouping: we developed a very efficient way to cluster users according to their dominant subspaces (quantization according to chordal distance). See [Adhikary, Caire, arXiv:1305.7252].

- Hybrid Beamforming and mm-wave application: the DFT pre-beamforming can be implemented by phase shifters in analog domain.

- Estimation of the long-term channel statistics: revamped interest in super-resolution methods (MUSIC, ESPRIT) especially for the mm-wave case.
Conclusions

- Exploiting transmit antenna correlation reduces the channel to a simpler \( \approx \) block diagonal structure.

- This is **generalized sectorization**! with MU-MIMO independently in each “sector” (group).

- We need only very coarse information on AoA and AS for the users .... DFT pre-beamforming.

- The idea can be easily extended to 3D beamforming (introducing elevation direction, Kronecker product structure).

- Downlink training, CSIT feedback and computation are greatly reduced (suitable for FDD).

- JSDM lends itself naturally to spatial-domain eICIC, simple inter-cell coordination, hybrid beamforming for mm-wave applications.
Thank You