JTG Summer School 2013

Asymptotic Random Matrix Theory with Applications to Wireless Networks

Giuseppe Caire

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

Department of Electrical Engineering, IIT Madras, June 26-29, 2013

Lecture 1:

Models, Performance Measures and Regimes of Interest

The Finite-Dimensional Linear-Gaussian Channel

- Many (almost all?) important scenarios in wireless communication networks yield a PHY layer model in the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{1}$$

where $\mathbf{y}, \mathbf{z} \in \mathbb{C}^n$, $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{H} \in \mathbb{C}^{n \times m}$, and $\mathbf{z} \in \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_z)$.

- (1) is a finite-dimensional Linear-Gaussian channel.

- Different special cases depend on the constraints at the input and output.

- Input constraints: they limit the empirical input distributions that the encoder(s) are allowed to generate.

- Output constraints: they limit the type of processing allowed at the decoder(s).

- Sequence of channel uses over signal-space dimensions: often we think of (1) as one channel use of the channel

$$\mathbf{y}[t] = \mathbf{H}[t]\mathbf{x}[t] + \mathbf{z}[t], \quad t = 1, \ldots, T$$

- Transmission of a block of T channel uses.

- The index $t \in \{1, \ldots, T\}$ denote the dimension over which coding is performed (this may be time, frequency, time-frequency ...).

- $\mathbf{H}[t]$ may change at every t, stay constant for all $t \in \{1, \ldots, T\}$ or change in blocks of some duration $L|T$.

Example 1: Direct-Sequence CDMA

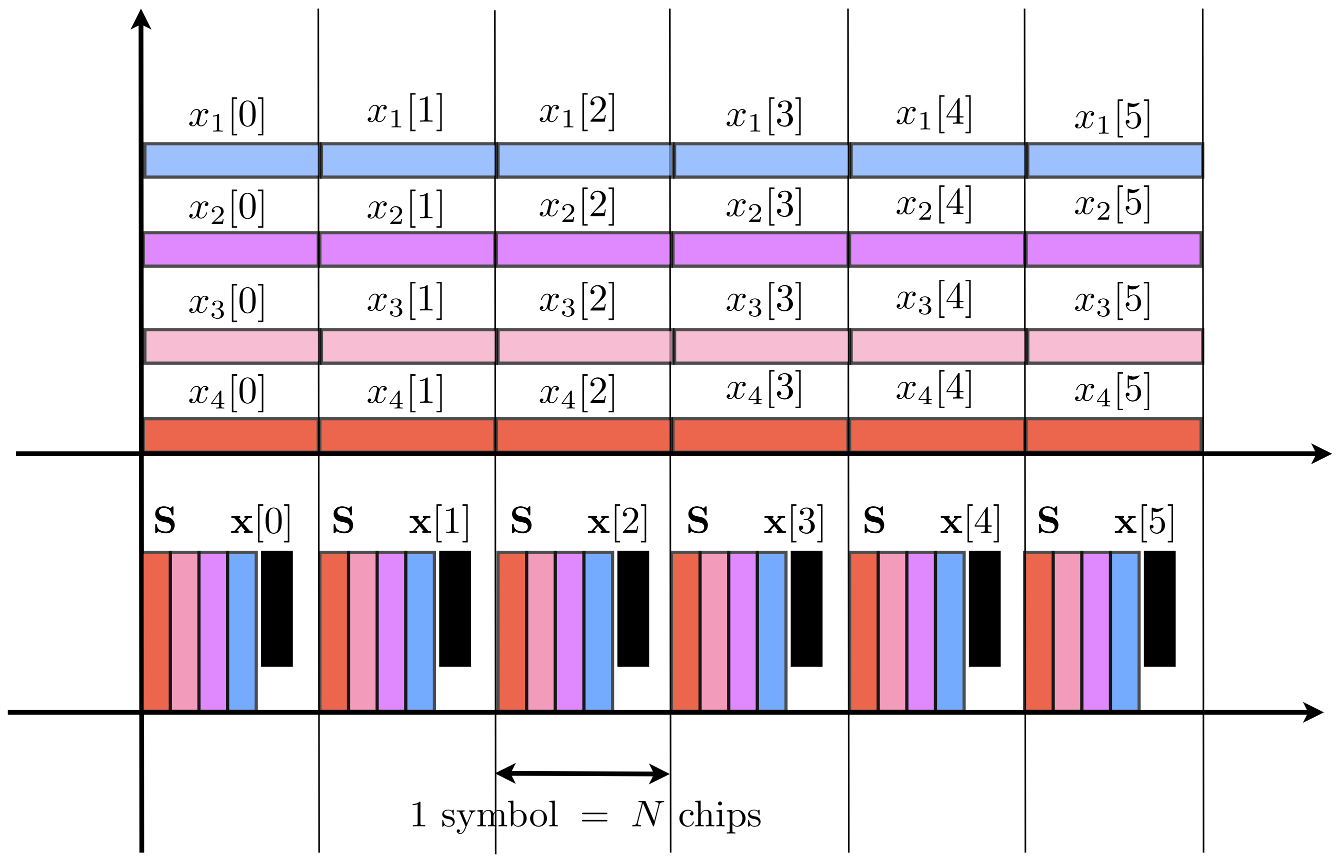
- Each user is given a spreading code (or signature sequence) $\mathbf{s}_k = (s_{1,k}, \ldots, s_{N,k})^{\mathsf{T}} \in \mathbb{C}^N$.
- A chip-synchronous and symbol-synchronous model, sampled at the chip rate, is given by

$$y[tN + i - 1] = \sum_{k=1}^{K} s_{i,k}\, x_k[t] + z[tN + i - 1], \quad i = 1, \ldots, N$$

- $x_k[tN]$ is the information symbol of user k at symbol time t.
- Stacking N consecutive chips into N-dimensional vectors, we obtain

$$\mathbf{y}[t] = \mathbf{S}\mathbf{x}[t] + \mathbf{z}[t]$$

- Chip normalization: $s_{i,k} = \frac{1}{\sqrt{N}} S_{i,k}$ with $|S_{i,k}| = 1$.
- Input power constraint (uplink): $\mathbb{E}[|x_k[t]|^2] \le P_k$.
- Input power constraint (downlink): $\sum_{k=1}^{K} \mathbb{E}[|x_k[t]|^2] \le P$.



- Simple generalization: “long spreading codes” (as in 3G WCDMA and CDMA 2000)

$$\mathbf{y}[t] = \mathbf{S}[t]\mathbf{x}[t] + \mathbf{z}[t]$$

The spreading code of each user changes from symbol to symbol.

- Simple generalization: frequency-flat fading (formally equivalent to shadowing or distance-dependent pathloss):

$$\mathbf{y}[t] = \mathbf{S}[t]\mathbf{A}[t]\mathbf{x}[t] + \mathbf{z}[t]$$

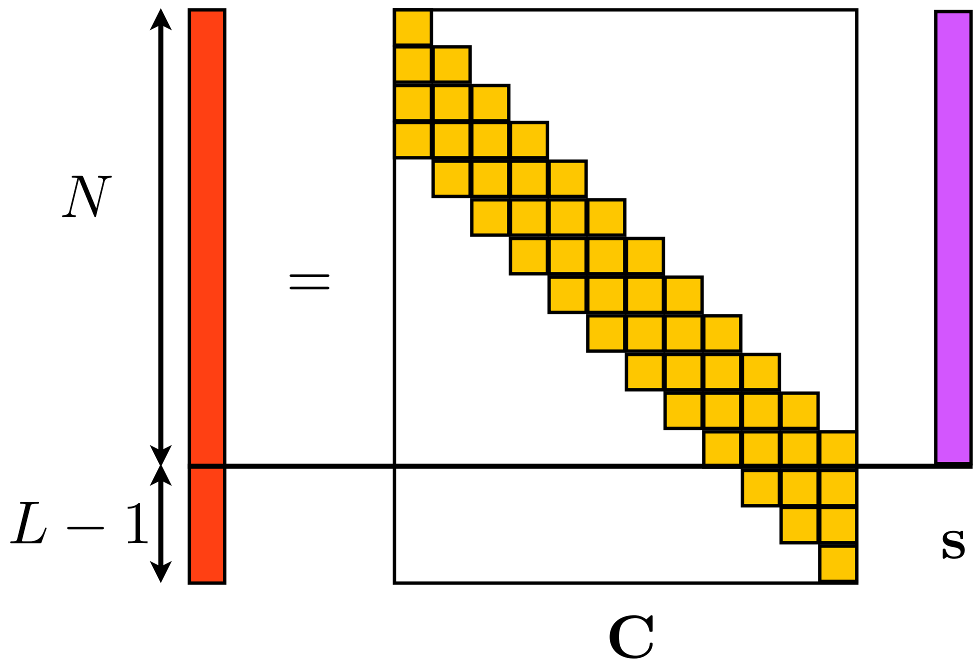
where $\mathbf{A}[t] = \text{diag}\left(A_1[t], \ldots, A_K[t]\right)$.

- Connection to our reference model: $\mathbf{H}[t] = \mathbf{S}[t]\mathbf{A}[t]$, $\mathbf{\Sigma}_z = N_0\mathbf{I}$, $n = N$, $m = K$.

Example 2: Direct-Sequence CDMA with Multipath

- Multipath fading in CDMA is modeled by a "short" channel impulse response with respect to the symbol duration. (We can neglect ISI).

- Effective spreading code is the convolution of $\mathbf{s}_k$ with the channel impulse response $\mathbf{c}_k$



- Uplink: each user is affected by its own frequency-flat pathloss/shadowing and multipath fading channel:

$$\mathbf{H} = [\mathbf{C}_1 \mathbf{s}_1, \ldots, \mathbf{C}_K \mathbf{s}_K] \, \mathbf{A}$$

- Downlink: the signal broadcasted by the base station is received at any given user k through its own pathloss/shadowing and multi path fading channel:

$$\mathbf{H} = \mathbf{C}\mathbf{S}$$

- In both cases, we have the model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$$

(make it time-varying as required).

Example 3: Multi-Carrier CDMA

- Introducing OFDM: the Inter-Symbol-Interference (ISI) channel in general:

$$y[i] = \sum_{\ell=0}^{L-1} c[\ell]x[i-\ell] + z[i]$$

- LTI system, with finite-length impulse response $\mathbf{c} = (c[0], \ldots, c[L-1])$.

- We use Cyclic Prefix (CP) precoding, i.e., fix block length N and send sequences of blocks $\{\mathbf{x}[t]\}$ with the CP precoding defined by

$$(\underbrace{x[tN], \ldots, x[tN+N-L+1], \ldots, x[tN+N-1]}_{\text{block}})$$

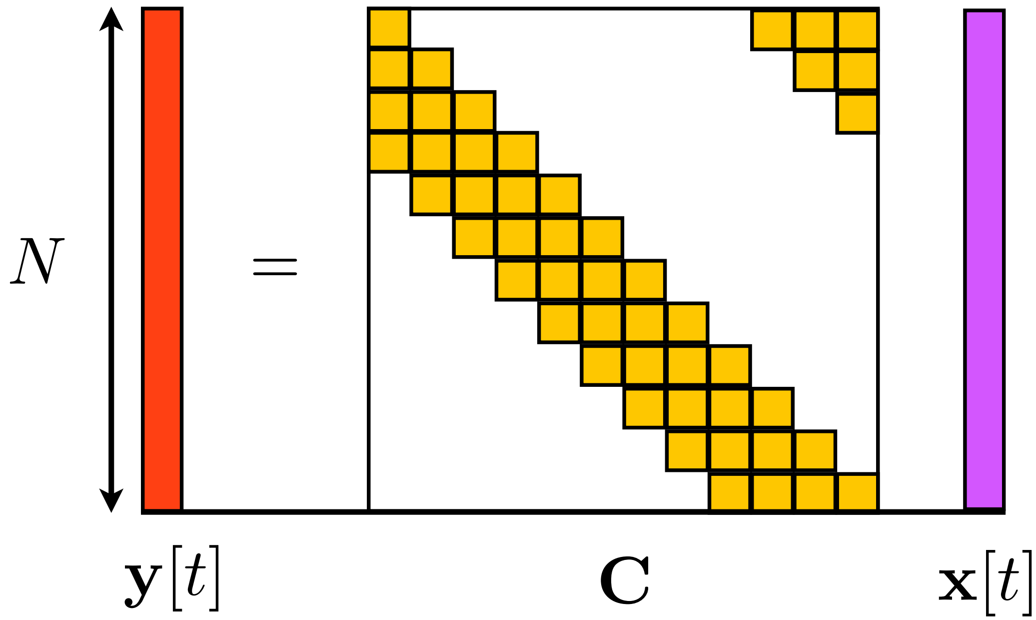
$$\downarrow$$

$$(\underbrace{x[tN+N-L+1], \ldots, x[tN+N-1]}_{\text{cyclic prefix}}, \underbrace{x[tN], \ldots\ldots, x[tN+N-1]}_{\text{block}})$$

- The vectorized channel model becomes

$$\mathbf{y}[t] = \mathbf{C}\mathbf{x}[t] + \mathbf{z}[t]$$

where $\mathbf{y}[t], \mathbf{z}[t], \mathbf{x}[t] \in \mathbb{C}^N$, and $\mathbf{C}$ is a circulant matrix with first column $\begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix}$.



- Result: Any $N \times N$ circulant matrix $\mathbf{C}$ can be written as

$$\mathbf{C} = \mathbf{F}^{\mathsf{H}}\, \mathsf{diag}(G_0, \ldots, G_{N-1})\ \mathbf{F}$$

where $\mathbf{F}$ is the unitary DFT matrix with elements

$$F_{k,\ell} = \frac{e^{-j\frac{2\pi}{N}k\ell}}{\sqrt{N}}, \quad k = 0, \ldots, N-1, \quad \ell = 0, \ldots, N-1$$

and where

$$\begin{bmatrix} G_0 \\ G_1 \\ \vdots \\ G_{N-1} \end{bmatrix} = \sqrt{N}\mathbf{F} \begin{bmatrix} c[0] \\ c[1] \\ \vdots \\ c[L-1] \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

is the vector of DFT coefficients of the impulse response, i.e.,

$$G_k = \sum_{\ell=0}^{L-1} c[\ell] e^{-j\frac{2\pi}{N}k\ell}$$

- IDFT precoding at the transmitter:

$$\mathbf{x}[t] = \mathbf{F}^{\mathsf{H}} \, \check{\mathbf{x}}[t]$$

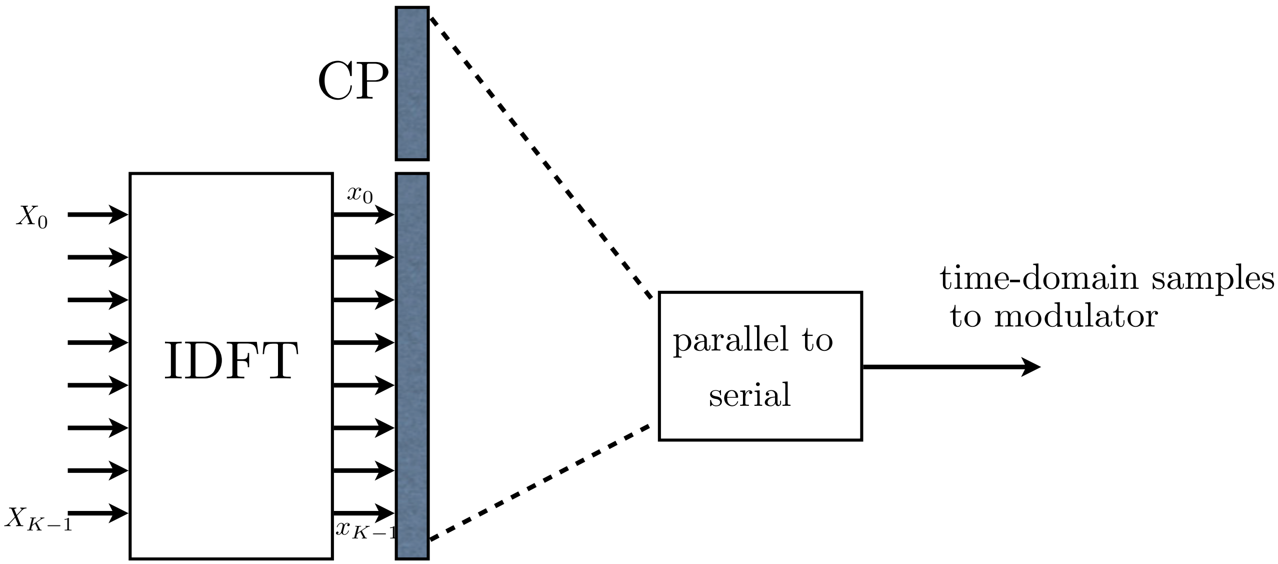
- DFT unitary transformation at the receiver:

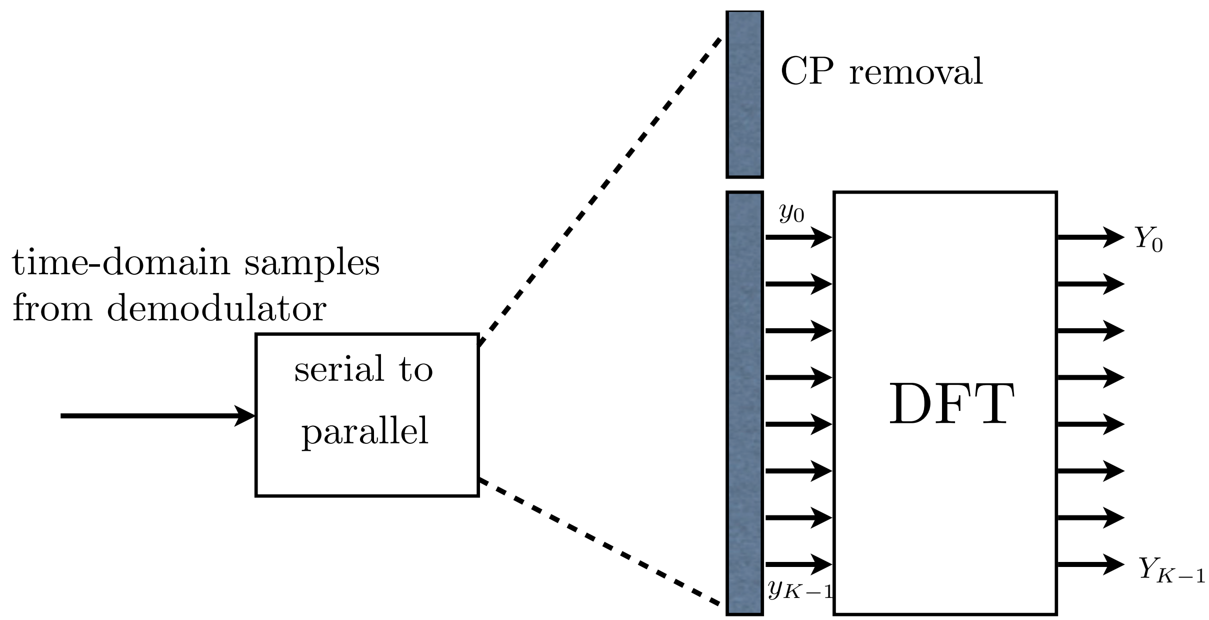
$$\check{\mathbf{y}}[t] = \mathbf{F} \, \mathbf{y}[t]$$

- The resulting frequency-domain OFDM channel is given by

$$\check{\mathbf{y}}[t] = \text{diag}(G_0, \ldots, G_{N-1})\check{\mathbf{x}}[t] + \check{\mathbf{z}}[t]$$

where the frequency-domain noise is $\check{\mathbf{z}}[t] = \mathbf{F}\mathbf{z}[t]$ (if $\mathbf{z}[t] \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ then also $\check{\mathbf{z}}[t] \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$).


CP
X_0
x_0
IDFT
X_{K-1}
x_{K-1}
parallel to serial
time-domain samples to modulator


CP removal
time-domain samples from demodulator
serial to parallel
y_0
DFT
Y_0
y_{K-1}
Y_{K-1}

- In multi-carrier CDMA, the block of frequency domain symbols are obtained by

$$\mathbf{x}_k[t] = \mathbf{s}_k x_k[t]$$

where $\mathbf{s}_k \in \mathbb{C}^N$ is the frequency-domain spreading code.

- Uplink: the resulting channel model is again given by (1) with

$$\mathbf{H} = (\mathbf{G} \odot \mathbf{S})\, \mathbf{A}$$

where $\odot$ is element-wise product,

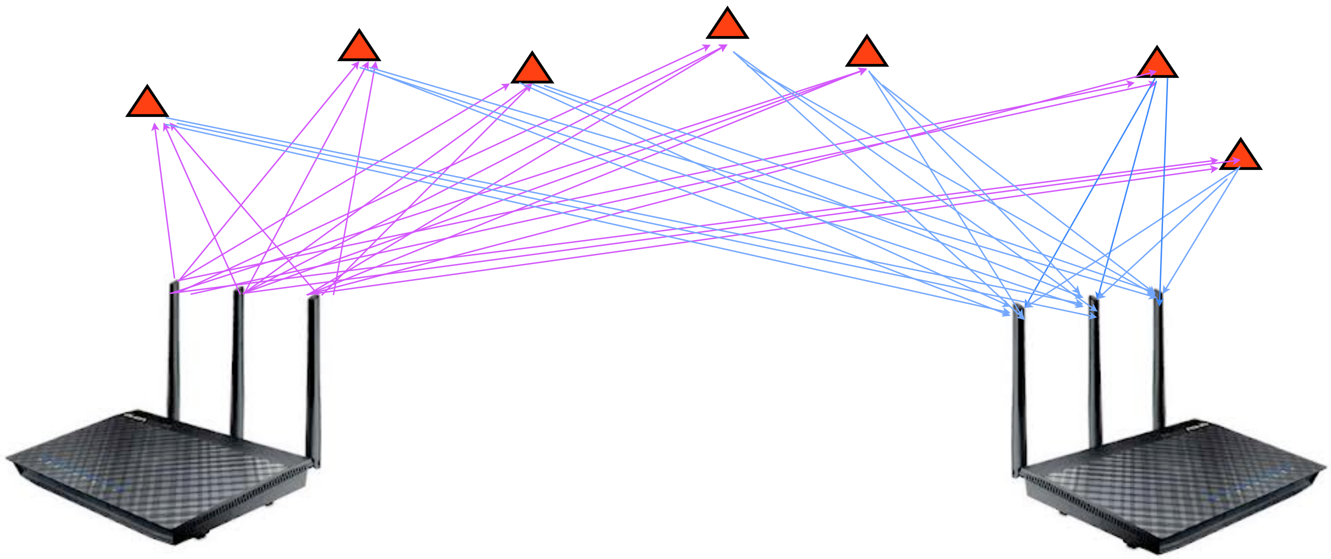
$$\mathbf{G} = \begin{bmatrix} G_{0,1} & \cdots & G_{0,K} \\ \vdots & & \vdots \\ G_{N-1,1} & \cdots & G_{N-1,K} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} s_{0,1} & \cdots & s_{0,K} \\ \vdots & & \vdots \\ s_{N-1,1} & \cdots & s_{N-1,K} \end{bmatrix}$$

and $\mathbf{A} = \text{diag}(A_1, \ldots, A_K)$ represents the frequency-flat pathloss/shadowing.

- Downlink: the resulting channel model is again given by (1) with

$$\mathbf{H} = \text{diag}(G_0, \ldots, G_{N-1})\mathbf{S}$$

Example 4: Frequency-Flat MIMO Point-to-Point



- One channel use of the MIMO point-to-point channel is given by (1) with $\mathbf{H} \in \mathbb{C}^{N \times M}$ and input constraint tr$(\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]) \leq P$.

- Elements $H_{i,j}$ of $\mathbf{H}$ represent the channel coefficients from Tx antenna j to Rx antenna i.

Example 5: MIMO-OFDM Point-to-Point

- Multipath MIMO channel: time-domain

$$\mathbf{y}[t] = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell \, \mathbf{x}[t-\ell] + \mathbf{z}[t]$$

- Using the same OFDM idea explained before, this can be reduced to the set of parallel channels in the frequency domain

$$\check{\mathbf{y}}[f,t] = \check{\mathbf{H}}[f]\check{\mathbf{x}}[f,t] + \check{\mathbf{z}}[f,t]$$

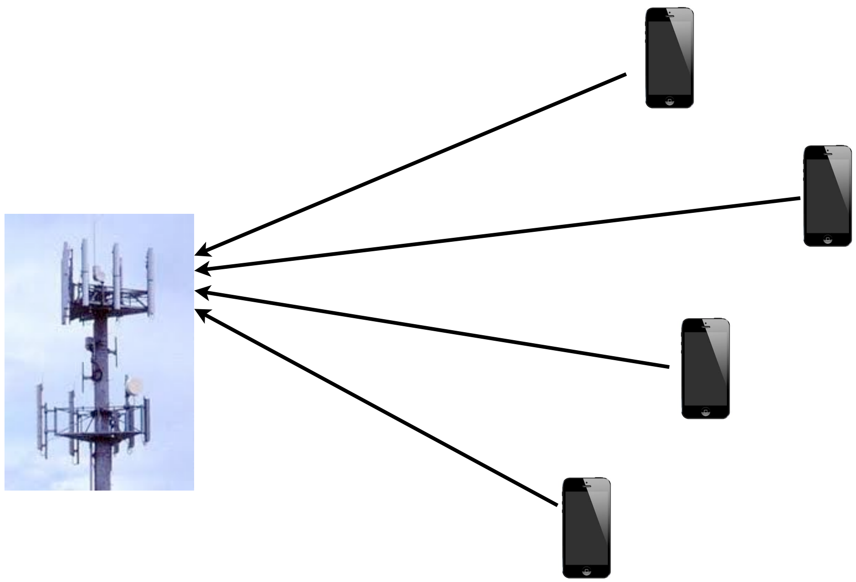
where $\nu \in \{0, \ldots, F-1\}$ is the subcarrier index and

$$\check{\mathbf{H}}[f] = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell e^{-j\frac{2\pi}{F} f \ell}$$

is the DFT of the matrix channel impulse response.

Example 6: MIMO Multiple Access Channel (MIMO-Uplink)

- For the sake of notation simplicity, we shall neglect the time-frequency cannel use index unless necessary.

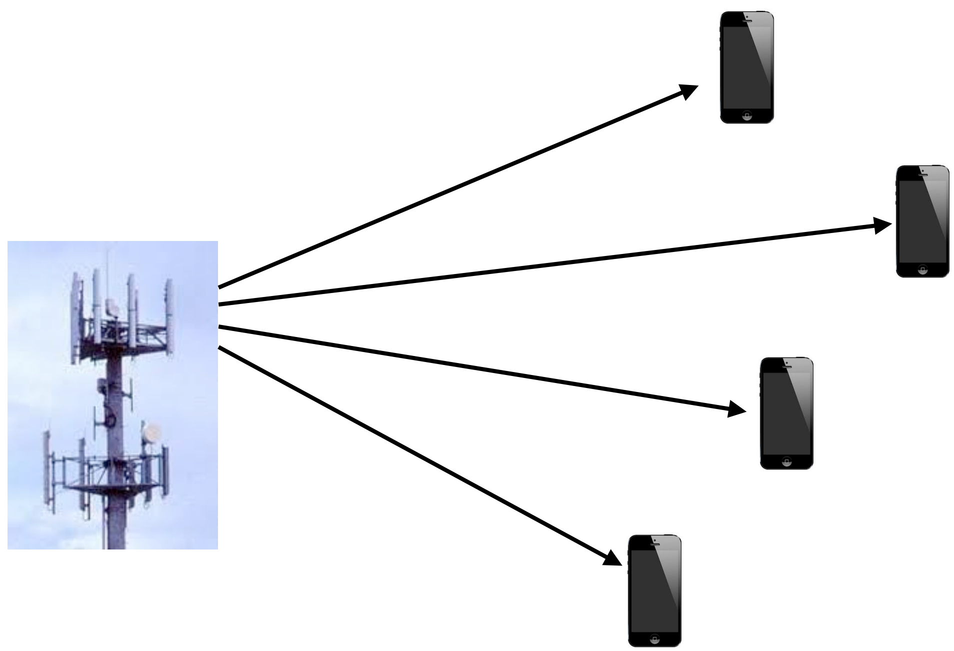


- The channel is still represented by (1) with the constraint that $\mathbf{x}$ is generated by a product distribution, i.e., $\mathbf{x} \sim \prod_{k=1}^{K} P_{X_k}$ (in particular, the input covariance $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]$ is diagonal).

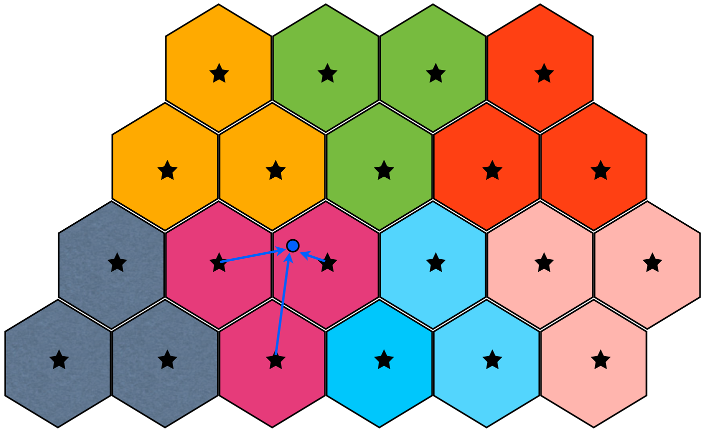
Example 7: MIMO Broadcast Channel (MIMO-Downlink)

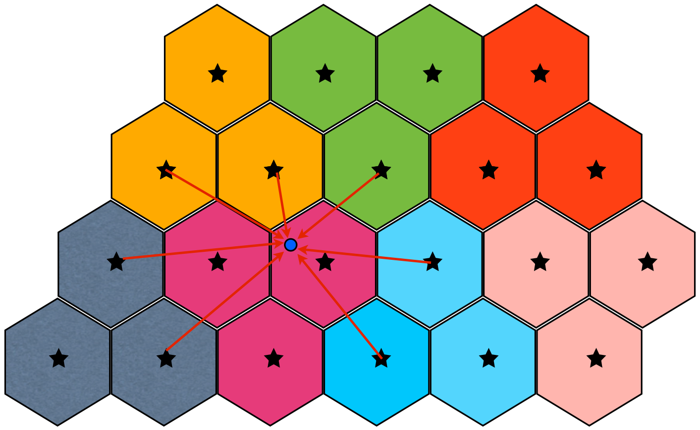
- In this case it is convenient to use the channel model with $\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z}$, and the constraint of decentralized processing at the receivers.

- Collection of channels $y_k = \mathbf{h}_k^{\mathsf{H}}\mathbf{x} + z_k$.

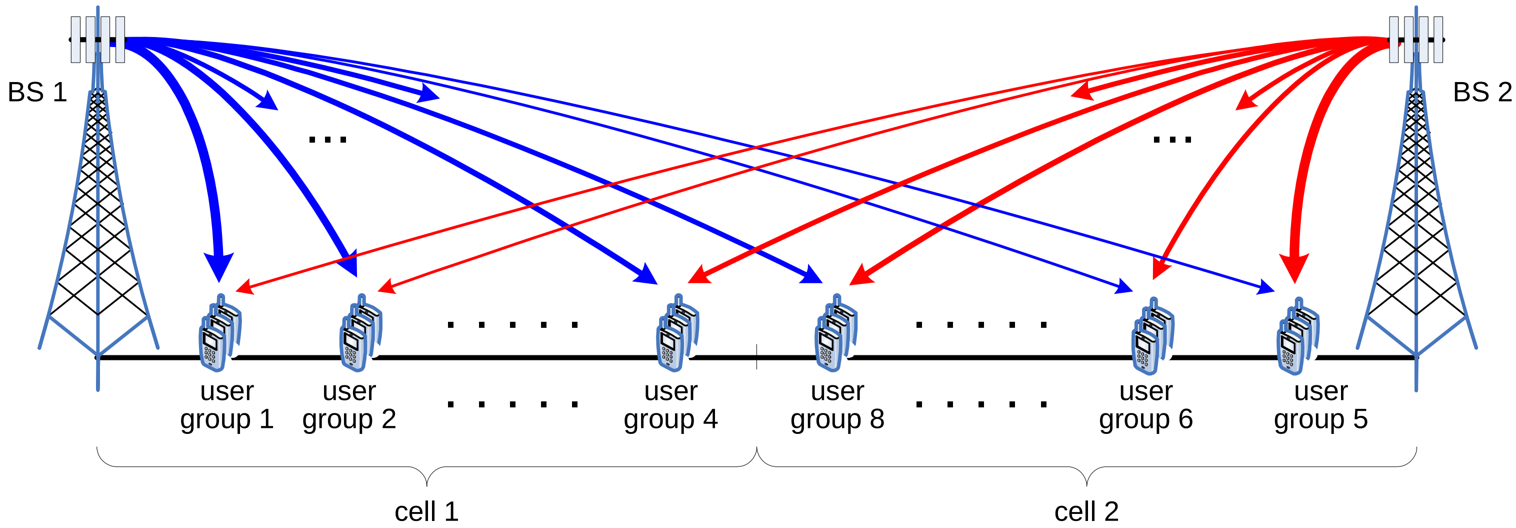


Example 8: Multi-Cell Models





Discretization of the Users Distribution



- We assume that the users are partitioned in co-located groups with N single-antenna terminals each.

- We have A user groups per cluster, and clusters of B cells.

- We have $M = \gamma N$ base station antennas per cell.

Cluster of Cooperating Base Stations

- Modified path coefficients $\beta_{m,k} = \frac{\alpha_{m,k}}{\sigma_k}$ taking into account the ICI power.

- Channel matrix ($B \times A$ blocks of size $\gamma N \times N$):

$$\mathbf{H} = \begin{bmatrix} \beta_{1,1}\mathbf{H}_{1,1} & \cdots & \beta_{1,A}\mathbf{H}_{1,A} \\ \vdots & \ddots & \vdots \\ \beta_{B,1}\mathbf{H}_{B,1} & \cdots & \beta_{B,A}\mathbf{H}_{B,A} \end{bmatrix}.$$

- Reference cluster channel model

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z}$$

where $\mathbf{y} = \mathbb{C}^{AN}$, $\mathbf{x} = \mathbb{C}^{\gamma BN}$, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

Performance measures: K users system

- Assume T large enough such that reliable communication is meaningful.

- Assume for simplicity a single channel matrix state spans T channel uses: i.e., $\mathbf{H}$ is constant for all $t \in \{1, \ldots, T\}$.

- For given $\mathbf{H}$, the Instantaneous achievable rate region is $\mathcal{R}(\mathbf{H}) \subset \mathbb{R}_+^K$.

- This means that for any $\epsilon > 0$ and rate K-tuple $\mathbf{R}(\mathbf{H}) = (R_1(\mathbf{H}), \ldots, R_K(\mathbf{H}))$ such that $\mathbf{R} + \epsilon \mathbf{1} \in \mathcal{R}(\mathbf{H})$ there exists a family of coding schemes for increasing T such that

$$\liminf_{T \to \infty} \frac{1}{T} \log |\mathcal{M}_k| \geq R_k \ \forall\, k, \quad \lim_{T \to \infty} \mathbb{P}\left(\bigcup_{k=1}^{K} \left\{ \widehat{W}_k \neq W_k \right\} \right) = 0$$

- Now we consider a long sequence of blocks of (large) length T, and we are interested in the long-term throughput region of the network, i.e., the region of long-term average rates

$$\overline{R}_k = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} R_k(\mathbf{H}[t]) = \mathbb{E}[R_k(\mathbf{H})]$$

- While $\mathcal{R}(\mathbf{H})$ may not be convex, the long-term throughput region $\overline{\mathcal{R}}$ is always convex, since time-sharing is always possible.

- We are interested in points $\overline{\mathbf{R}}$ on the boundary of $\overline{\mathcal{R}}$.

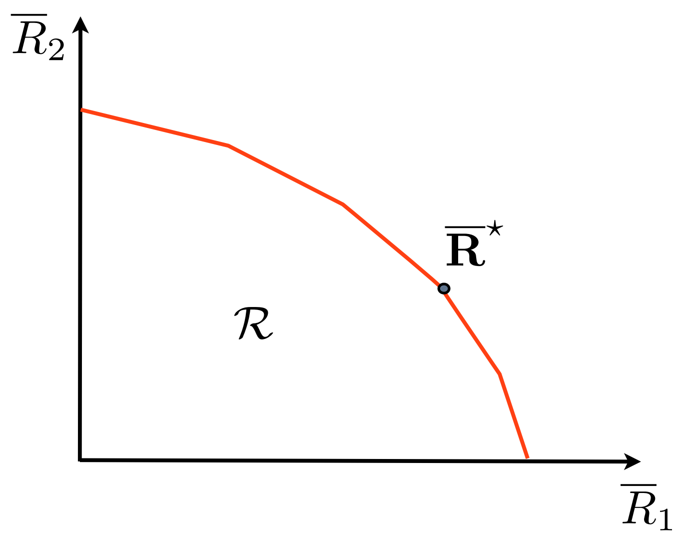
- In particular, we are interested in maximizing some desired concave and componentwise non-decreasing network utility function $U(\overline{\mathbf{R}})$, that reflects some desired notion of fairness.

- Network Utility Maximization (NUM):

$$\begin{array}{ll} \text{maximize} & U(\overline{\mathbf{R}}) \\ \text{subject to} & \overline{\mathbf{R}} \in \overline{\mathcal{R}} \end{array} \tag{2}$$

- The problem is always convex.

- Difficulty: $\overline{\mathcal{R}}$ is typically very hard to express in closed form (curved, uncountable number of supporting hyperplanes).



Scheduling and resource allocation

- We have an *instantaneous coding strategy* achieving points $\mathbf{R}(\mathbf{H}) \in \mathcal{R}(\mathbf{H})$ for any channel state $\mathbf{H}$.

- Over the sequence of successive blocks, we wish to schedule the users and allocate the network resources such that, in the long-term average sense, we achieve the point $\overline{\mathbf{R}}^\star \in \mathcal{R}$, solution of (2).

- A general method: Drift Plus Penalty (DPP).

- Let $R_k[t] = R_k(\mathbf{H}[t])$ denote the instantaneous achievable rates and define the transmission queues

$$Q_k[t+1] = [Q_k[t] - R_k[t]]_+ + A_k[t]$$

for a set of arrival processes $\{A_k[t]\}$.

- System stability region: convex closure of all arrival rates $\boldsymbol{\lambda}$, with $\lambda_k = \mathbb{E}[A_k[t]]$, such that there exists a transmission policy such that all queues are strongly stable: $\limsup_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \mathbb{E}[Q_k[t]] < \infty$ for all k.

Theorem 1. *(Stability Policy) Suppose that the arrival process $\mathbf{A}[t]$ is i.i.d. over the slots, with elements uniformly bounded in $[0, A_{\max}]$, and that the channel state $\mathbf{H}[t]$ also forms an i.i.d. sequence over the slots. Then, the system stability region coincides with $\overline{\mathcal{R}}$. Furthermore, any $\boldsymbol{\lambda}$ in the interior of $\overline{\mathcal{R}}$ is stabilized by the max-weight dynamic policy, solution of:*

$$\begin{aligned} \textit{maximize} \quad & \sum_{k=1}^{K} Q_K[t] R_k[t] \\ \textit{subject to} \quad & (R_1[t], \ldots, R_K[t]) \in \mathcal{R}(\mathbf{H}[t]) \end{aligned} \tag{3}$$

$\square$

For a proof, see for example [L. Georgiadis, M. J. Neely and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, NOW Pub., 2006].

Theorem 2. *(Utility Maximization) Consider a virtual arrival process defined as follows:* $\mathbf{A}[t] = \mathbf{a}^*$ *with*

$$\mathbf{a}^* = \textit{arg} \max_{\mathbf{a} \in [0, A_{\max}]^K} \left\{ VU(\mathbf{a}) - \sum_{k=1}^{K} a_k Q_k[t] \right\}$$

for some $0 < A_{\max} < \infty$ *and* $V > 0$. *Then, by applying the stability policy of Theorem 1 to such virtual queues, the resulting long-term averaged network utility satisfies*

$$\liminf_{\tau \to \infty} U\left(\frac{1}{\tau} \sum_{t=1}^{\tau} \mathbb{E}[\mathbf{R}[t]] \right) \geq U(\overline{\mathbf{R}}^{\star}) - \frac{\kappa}{V}$$

for some system-dependent constant κ *and for sufficiently large* $A_{\max}$. *In addition, all virtual queues are strongly stable, with*

$$\limsup_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \mathbb{E}[Q_k[t]] = O(V), \quad \forall\, k$$

$\square$

For a proof, see for example [Georgiadis, Neely, Tassiulas, FnT 2006].

A Simple Example

- Orthogonal multiple access channel with individual link capacities C_1, C_2. The instantaneous rate region is the non-convex discrete set of points

$$\mathcal{R} = \{(C_1, 0), (0, C_2)\}$$

- In this case $\overline{\mathcal{R}}$ is the set of all non-negative $(\overline{R}_1, \overline{R}_2)$ such that

$$\frac{\overline{R}_1}{C_1} + \frac{\overline{R}_2}{C_2} \leq 1$$

- We wish to maximize the Proportional Fairness network utility function

$$U(\overline{R}_1, \overline{R}_2) = \sum_{k=1}^{2} \log \overline{R}_k$$

- In this case, we obtain an explicit closed-form solution:

$$\mathcal{L}(\overline{R}_1, \overline{R}_2, \mu) = \sum_{k=1}^{2} \log \overline{R}_k - \mu \left(\frac{\overline{R}_1}{C_1} + \frac{\overline{R}_2}{C_2} - 1 \right)$$

- Differentiating and applying the KKT conditions, we obtain

$$\frac{\partial \mathcal{L}}{\partial \overline{R}_k} = \frac{1}{\overline{R}_k} - \frac{\mu}{C_k} \leq 0$$

yielding

$$\overline{R}_k \geq \frac{C_k}{\mu}$$

Since $\overline{R}_k = 0$ yields an objective function value equal to $-\infty$, the solution must be strictly positive. Hence

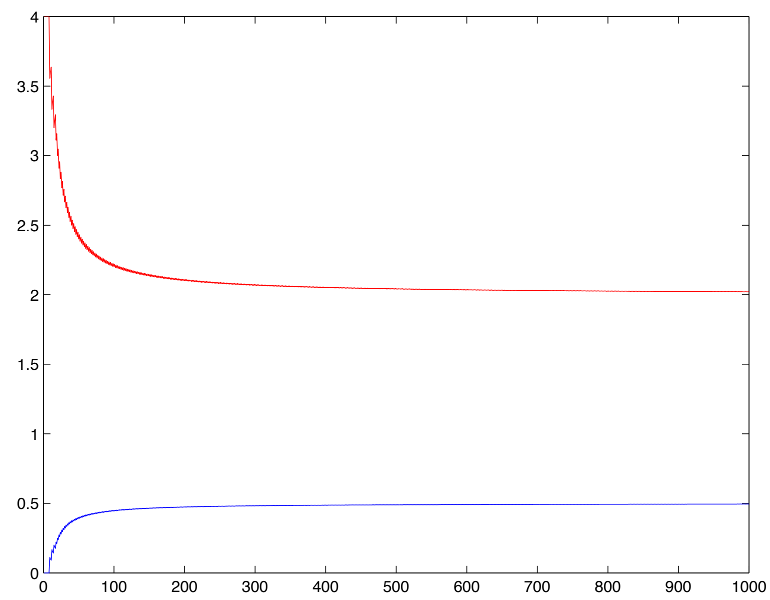
$$\overline{R}_k = \frac{C_k}{\mu}$$

Replacing in the constraint, we obtain $\mu = 2$, such that

$$\overline{R}_k = \frac{C_k}{2}$$

- This corresponds to serving each user for a fraction $1/2$ of the slots (each user is given equal transmission resource).

- Simulation: $C_1 = 1, C_2 = 4$, $V = 10, A_{\max} = 5$. Time-averaged throughput versus slots:



Regime of Interest: Large Systems

- We are interested in the regime of $n, m \to \infty$, with fixed ratio $\frac{m}{n} = \beta$.
- Normalizations: the elements of $\mathbf{H}$ have mean 0, variance $O(\frac{1}{n})$ and higher order moments that vanish sufficiently fast.
- In DS-CDMA systems,

$$s_{i,k} = \frac{1}{\sqrt{N}} S_{i,k}, \quad \text{with} \quad \mathbb{E}[|S_{i,k}|^2] = 1$$

- In downlink or single-user MIMO systems, with total input power constraint P, we have

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z}, \quad \mathrm{tr}\left(\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]\right) = P$$

- We can divide and multiply by M (number of Tx antennas) and have

$$\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z}, \quad \frac{1}{M}\mathrm{tr}\left(\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]\right) = P$$

- In uplink MIMO systems, it is reasonable to assume that the total transmit power is constant, such that

$$\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{H}\mathbf{x} + \mathbf{z}, \quad \frac{\beta}{K}\mathrm{tr}\left(\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]\right) = P$$

  where $\beta = K/M$.

- General idea: under relatively mild and general conditions, in a large number of relevant settings, the instantaneous rates $R_k(\mathbf{H})$ become deterministic constants that depend on the system "geometry", but are independent of the specific realization of $\mathbf{H}$.

- As a consequence: all NUM scheduling problems become as easy as the simple example of before!

End of Lecture 1

Lecture 2:

Basic Results with I.I.D. Matrices

DS-CDMA with i.i.d. Spreading Codes

- Recall the model

$$\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{z} = \sum_{k=1}^{K} \mathbf{s}_k x_k + \mathbf{z}, \tag{4}$$

- $\mathbf{S} \in \mathbb{C}^{N \times K}$, with i.i.d. elements, $s_{i,k} = \frac{1}{\sqrt{N}} S_{i,k}$, with $\mathbb{E}[|S_{i,k}|^2] = 1$ and finite higher order moments (for brevity, we will say "well-behaved").

- $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I})$.

- Uplink symmetric case: $\mathbb{E}[|x_k|^2] \leq P$, where P denotes the energy per symbol (power) for each user.

- We define the SNR per user as

$$\mathsf{snr} = \frac{P}{N_0}$$

Vector Gaussian MAC capacity region

Theorem 3. *The capacity region of the vector Gaussian MAC (4) is given by the set of inequalities*

$$\sum_{k\in\mathcal{K}} R_k \leq \max_{P(\mathbf{x})\in\mathcal{P}} \frac{1}{N} I(\mathbf{x}(\mathcal{K}); \mathbf{y}|\mathbf{x}(\mathcal{K}^c), \mathbf{S})$$

for all subsets $\mathcal{K} \subseteq \{1, \ldots, K\}$, where $\mathbf{x}(\mathcal{K})$ denotes the collection of input variables $\{x_k : k \in \mathcal{K}\}$ and where $\mathcal{P}$ denotes the set of product input distributions satisfying the input power constraint. □

For a proof, see for example [T. Cover and J. Thomas, *Elements of information theory, 2nd Ed.*, Wiley 2012].

- It is not difficult to show that, for any subset $\mathcal{K}$, the corresponding mutual information term is maximized by letting $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, P\mathbf{I})$, such that

$$I(\mathbf{x}(\mathcal{K}); \mathbf{y}|\mathbf{x}(\mathcal{K}^c), \mathbf{S}) = \mathbb{E}\left[\log\left|\mathbf{I} + \mathsf{snr}\mathbf{S}(\mathcal{K})\mathbf{S}^{\mathsf{H}}(\mathcal{K})\right|\right]$$

where $\mathbf{S}(\mathcal{K})$ is the submatrix of $\mathbf{S}$ comprising the columns $\{\mathbf{s}_k : k \in \mathcal{K}\}$.

- The biting constraint for the sum rate is given by

$$R_{\text{sum}} \leq \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I} + \mathsf{snr}\mathbf{S}\mathbf{S}^{\mathsf{H}}\right|\right]$$

- We have

$$
\begin{aligned}
\frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I}+\mathsf{snr}\mathbf{S}\mathbf{S}^{\mathsf{H}}\right|\right] &= \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}\log\left(1+\mathsf{snr}\lambda_i(\mathbf{S}\mathbf{S}^{\mathsf{H}})\right)\right] \\
&= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\log\left(1+\mathsf{snr}\lambda_i(\mathbf{S}\mathbf{S}^{\mathsf{H}})\right)\right] \\
&= \mathbb{E}\left[\log\left(1+\mathsf{snr}\lambda(\mathbf{S}\mathbf{S}^{\mathsf{H}})\right)\right] \\
&= \mathbb{E}\left[\int_0^\infty \log(1+\mathsf{snr}\lambda)dF^{(N)}_{\mathbf{S}\mathbf{S}^{\mathsf{H}}}(\lambda)\right]
\end{aligned}
$$

where $\lambda(\mathbf{M})$ denotes an eigenvalue of a matrix $\mathbf{M}$, and $F^{(N)}_{\mathbf{M}}(\lambda)$ is the Empirical Spectral Distribution (ESD) of the unordered eigenvalues of an $N \times N$ matrix $\mathbf{M}$, defined by the “ladder” function

$$
F^{(N)}_{\mathbf{S}\mathbf{S}^{\mathsf{H}}}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} 1\{\lambda_i(\mathbf{S}\mathbf{S}^{\mathsf{H}}) \le \lambda\}
$$

- In general, for random $\mathbf{S}$ and finite N we have that $F^{(N)}_{\mathbf{S}\mathbf{S}^{\mathsf{H}}}(\lambda)$ is a collection of random variables, for all $\lambda \in \mathbb{R}$.

- Going to the limit: by letting $N \to \infty$ with $K/N = \beta$, under rather mild conditions (always verified in the cases treated here), we have that

$$F^{(N)}_{\mathbf{S}\mathbf{S}^{\mathsf{H}}}(\lambda) \Longrightarrow F_{\mathbf{S}\mathbf{S}^{\mathsf{H}}}(\lambda)$$

  where $F_{\mathbf{M}}(\lambda)$ is the Limit Spectral Distribution (LSD) of the sequence of random matrices $\mathbf{S}\mathbf{S}^{\mathsf{H}}$, for increasing N.

- Technically speaking, this convergence is weak convergence almost everywhere, that is, for each point of continuity λ we have convergence almost surely (with respect to the probability space of the random matrices).

- Under such convergence conditions, the large-system limit of the symmetric sum rate is given by

$$R_{\text{sum}} = \int_0^\infty \log(1 + \mathsf{snr}\lambda) dF_{\mathbf{S}\mathbf{S}^{\mathsf{H}}}(\lambda)$$

One Step Back: Transforms

- The explicit characterization of the LSD of a sequence of random matrices is typically difficult (only a few famous results are known).
- In contrast, we shall follow an implicit characterization, through some appropriate integral transform of the LSD.
- Without trying to be fully exhaustive, we start by introducing here two fundamental transform which have a communication theoretic significance.

Definition 1. *(η-Transform) Let X denote a non-negative RV. The η transform of X is defined by*

$$\eta_X(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right] = \int_0^\infty \frac{1}{1+\gamma x} dF_X(x)$$

for $\gamma \in \mathbb{R}_+$.

We will use the notation $\eta_{\mathbf{M}}(\gamma)$ to indicate the η-transform of $\lambda(\mathbf{M}) \sim F_{\mathbf{M}}(\lambda)$, the LSD of some sequence of random matrices $\mathbf{M}$. $\Diamond$

Definition 2. *(Shannon-Transform) Let X denote a non-negative RV. The Shannon-transform of X is defined by*

$$\mathcal{V}_X(\gamma) = \mathbb{E}\left[\log(1+\gamma X)\right] = \int_0^\infty \log(1+\gamma x) dF_X(x)$$

for $\gamma \in \mathbb{R}_+$.

We will use the notation $\mathcal{V}_{\mathbf{M}}(\gamma)$ to indicate the Shannon-transform of $\lambda(\mathbf{M}) \sim F_{\mathbf{M}}(\lambda)$, the LSD of some sequence of random matrices $\mathbf{M}$. $\Diamond$

- Going back to our DS-CDMA uplink channel, we have

$$R_{\rm sum} = \mathcal{V}_{\mathbf{S}\mathbf{S}^{\sf H}}(\mathsf{snr})$$

Elementary Properties of η and Shannon Transforms

- $\eta_X(\gamma)$ is strictly monotonically decreasing for $\gamma \in \mathbb{R}_+$, with $\eta_X(0) = 1$ and $\lim_{\gamma\to\infty} \eta_X(\gamma) = \mathbb{P}(X = 0)$.

- $\gamma\eta_X(\gamma)$ is strictly monotonically increasing for $\gamma \in \mathbb{R}_+$ from 0 to $\mathbb{E}[1/X]$.

- Asymptotic normalized rank of $\mathbf{M}$ (fraction of non-zero eigenvalues) is $\rho = 1 - \lim_{\gamma\to\infty} \eta_{\mathbf{M}}(\gamma)$, and

$$\lim_{N\to\infty} \frac{1}{N}\text{tr}\left(\mathbf{M}^{-1}\right) = \lim_{\gamma\to\infty} \gamma\eta_{\mathbf{M}}(\gamma)$$

- For any $\mathbf{A}$ of dimension $N \times K$ and $\mathbf{B}$ of dimension $K \times N$, such that $\mathbf{M} = \mathbf{AB}$ is non-negative definite,

$$N\left(1 - \eta_{\mathbf{AB}}(\gamma)\right) = K\left(1 - \eta_{\mathbf{BA}}(\gamma)\right)$$

such that, in the limit of $N \to \infty$ and $K/N = \beta$, we have

$$\eta_{\mathbf{AB}}(\gamma) = 1 - \beta + \beta\eta_{\mathbf{BA}}(\gamma)$$

- Relation between η and Shannon transform:

$$\gamma\frac{d}{d\gamma}\mathcal{V}_X(\gamma) = 1 - \eta_X(\gamma)$$

- Trace Lemma: For a sequence of $N \times N$ matrices with uniformly bounded spectral norm, and a sequence of random vectors $\mathbf{s}$ with i.i.d. components with mean 0 and variance $1/N$, independent of $\mathbf{M}$,

$$\mathbf{s}^{\mathsf{H}}\mathbf{M}\mathbf{s} - \frac{1}{N}\text{tr}(\mathbf{M}) \xrightarrow{a.s.} 0$$

- As a consequence,

$$\mathbf{s}^{\mathsf{H}}\left(\mathbf{I} + \gamma\mathbf{M}\right)^{-1}\mathbf{s} \xrightarrow{a.s.} \eta_{\mathbf{M}}(\gamma)$$

Sum-Capacity of the Symmetric DS-CDMA Uplink

Theorem 4. *Let* $\mathbf{S}$ *be* $N \times K$ *with i.i.d. elements of the type* $s_{i,k} = \frac{1}{\sqrt{N}} S_{i,k}$*, with well-behaved* $S_{i,k}$*. Let* $\mathbf{T}$ *denote a diagonal non-negative definite matrix with well-defined LSD* $F_{\mathbf{T}}(\lambda)$*, as* $K \to \infty$*. Then, as* $N \to \infty$ *with* $K/N = \beta$*, the LSD of* $\mathbf{STS}^{\mathsf{H}}$ *exists and has* η*-transform* $\eta_{\mathbf{STS}^{\mathsf{H}}}(\gamma) = \eta$*, solution of the equation*

$$\beta = \frac{1 - \eta}{1 - \eta_{\mathbf{T}}(\gamma\eta)} \tag{5}$$

The corresponding Shannon transform is given by

$$\mathcal{V}_{\mathbf{STS}^{\mathsf{H}}}(\gamma) = \beta \mathcal{V}_{\mathbf{T}}(\gamma\eta) + \log \frac{1}{\eta} + \eta - 1 \tag{6}$$

□

Sketch of Proof:

- We give an instructive proof with profound communication theoretic significance.

- Consider a receiver that wishes to detect user k (here we assume $\mathbb{E}[|x_k|^2] = 1$ for all k):

$$\mathbf{y} = \mathbf{s}_k \sqrt{T_k} x_k + \sum_{j \neq k} \mathbf{s}_j \sqrt{T_j} x_j + \mathbf{z}$$

- The optimal linear receiver maximizes the Signal-to-Interference plus Noise (SINR) at its output, and is given by by linear MMSE receiver:

$$\widetilde{x}_k = \mathbb{E}[x_k \mathbf{y}^{\mathsf{H}}] \left(\mathbb{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}] \right)^{-1} \mathbf{y}$$

- Explicitly, we have

$$\mathbb{E}[x_k \mathbf{y}^{\mathsf{H}}] = \sqrt{T_k} \mathbf{s}_k^{\mathsf{H}}$$

and

$$\mathbb{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}] = T_k \mathbf{s}_k \mathbf{s}_k^{\mathsf{H}} + \underbrace{\sum_{j \neq k} T_j \mathbf{s}_j \mathbf{s}_j^{\mathsf{H}} + N_0 \mathbf{I}}_{\boldsymbol{\Sigma}_k}$$

- The resulting MMSE is given by

$$\begin{aligned}
\mathsf{MMSE}_k &= \mathbb{E}[|x_k|^2] - \mathbb{E}[x_k \mathbf{y}^{\mathsf{H}}] \left(\mathbb{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}]\right)^{-1} \left(\mathbb{E}[x_k \mathbf{y}^{\mathsf{H}}]\right)^{\mathsf{H}} \\
&= 1 - T_k \mathbf{s}_k^{\mathsf{H}} \left(T_k \mathbf{s}_k \mathbf{s}_k^{\mathsf{H}} + \boldsymbol{\Sigma}_k\right)^{-1} \mathbf{s}_k \\
&= 1 - T_k \mathbf{s}_k^{\mathsf{H}} \left(\boldsymbol{\Sigma}_k^{-1} - \frac{T_k \boldsymbol{\Sigma}_k^{-1} \mathbf{s}_k \mathbf{s}_k^{\mathsf{H}} \boldsymbol{\Sigma}_k^{-1}}{1 + T_k \mathbf{s}_k^{\mathsf{H}} \boldsymbol{\Sigma}_k^{-1} \mathbf{s}_k} \right) \mathbf{s}_k \\
&= \frac{1}{1 + T_k \mu_k}
\end{aligned}$$

with $\mu_k = \mathbf{s}_k^{\mathsf{H}} \boldsymbol{\Sigma}_k^{-1} \mathbf{s}_k$.

- From the well-known relation between MMSE and SINR, we have

$$\mathsf{sinr}_k = \frac{\mathbb{E}[|x_k|^2] - \mathsf{MMSE}_k}{\mathsf{MMSE}_k} = \left(1 - \frac{1}{1 + T_k\mu_k}\right)(1 + T_k\mu_k) = T_k\mu_k$$

- Now, letting λ_i denote the i-th eigenvalue of $\mathbf{STS}^{\mathsf{H}}$, we can write:

$$\begin{aligned}
\sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i + N_0} &= \mathsf{tr}\left(\left(N_0\mathbf{I} + \mathbf{STS}^{\mathsf{H}}\right)^{-1}\mathbf{STS}^{\mathsf{H}}\right) \\
&= \mathsf{tr}\left(\left(N_0\mathbf{I} + \mathbf{STS}^{\mathsf{H}}\right)^{-1}\sum_{k=1}^{K} T_k\mathbf{s}_k\mathbf{s}_k^{\mathsf{H}}\right) \\
&= \sum_{k=1}^{K} T_k\mathbf{s}_k^{\mathsf{H}}\left(N_0\mathbf{I} + \mathbf{STS}^{\mathsf{H}}\right)^{-1}\mathbf{s}_k \\
&= \sum_{k=1}^{K} \frac{\mathsf{sinr}_k}{\mathsf{sinr}_k + 1}
\end{aligned}$$

where the last step follows again by the matrix inversion lemma and by using the SINR expression found before.

- We argue that in the limit for large k, the quantity

$$\mu_k = \mathbf{s}_k^{\mathsf{H}} \boldsymbol{\Sigma}_k^{-1} \mathbf{s}_k = \frac{1}{N_0} \mathbf{s}_k^{\mathsf{H}} \left(\mathbf{I} + \frac{1}{N_0} \mathbf{S}_k \mathbf{T}_k \mathbf{S}_k^{\mathsf{H}} \right)^{-1} \mathbf{s}_k \xrightarrow{a.s.} \frac{1}{N_0} \eta_{\mathbf{STS}^{\mathsf{H}}} \left(\frac{1}{N_0} \right)$$

does not depend on k any longer.

- Hence, in the limit, we have

$$\begin{aligned} \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \frac{T_k \mu_k}{T_k \mu_k + 1} &= 1 - \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{1 + \frac{T_k}{N_0} \eta_{\mathbf{STS}^{\mathsf{H}}} \left(\frac{1}{N_0} \right)} \\ &= 1 - \eta_{\mathbf{T}} \left(\frac{1}{N_0} \eta_{\mathbf{STS}^{\mathsf{H}}} \left(\frac{1}{N_0} \right) \right) \end{aligned}$$

- By the definition of η-transform, we also have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i}{\lambda_i+N_0}=1-\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+\frac{1}{N_0}\lambda_i}=1-\eta_{\mathbf{STS}^{\mathsf{H}}}\left(\frac{1}{N_0}\right)$$

- Putting things together, letting $\gamma = 1/N_0$, and recalling that $K = \beta N$,

$$1-\eta_{\mathbf{STS}^{\mathsf{H}}}(\gamma)=\beta\left(1-\eta_{\mathbf{T}}\left(\gamma\eta_{\mathbf{STS}^{\mathsf{H}}}(\gamma)\right)\right)$$

such that they key equation (5) is proved.

- Next, we need to proof the expression for $\mathcal{V}_{\mathbf{STS}^{\mathsf{H}}}(\gamma)$. To this purpose, we write $\eta = \eta_{\mathbf{STS}^{\mathsf{H}}}(\gamma)$, $\dot{\eta} = \frac{d}{d\gamma}\eta$ and define T $\sim F_{\mathbf{T}}(t)$ to be a RV distributed as the

LSD of $\mathbf{T}$. Then:

$$\begin{aligned}
\frac{d}{\gamma}\beta\mathcal{V}_{\mathbf{T}}(\gamma\eta) &= \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\gamma\eta\mathsf{T}}(\eta+\gamma\dot{\eta})\right] \\
&= \frac{\beta}{\gamma}\mathbb{E}\left[\frac{\gamma\eta\mathsf{T}}{1+\gamma\eta\mathsf{T}}\right]\left(1+\gamma\frac{\dot{\eta}}{\eta}\right) \\
&= \frac{\beta}{\gamma}\left(1-\mathbb{E}\left[\frac{1}{1+\gamma\eta\mathsf{T}}\right]\right)\left(1+\gamma\frac{\dot{\eta}}{\eta}\right) \\
&= \frac{\beta}{\gamma}\left(1-\eta_{\mathbf{T}}(\gamma\eta)\right)\left(1+\gamma\frac{\dot{\eta}}{\eta}\right) \\
&= \frac{1-\eta}{\gamma}\left(1+\gamma\frac{\dot{\eta}}{\eta}\right)
\end{aligned}$$

where we have used the key equation (5).

- Using the differential relation between Shannon transform and η-transform, we have

$$\frac{d}{d\gamma}\mathcal{V}_{\mathbf{STS}^{\mathsf{H}}}(\gamma) = \frac{1-\eta_{\mathbf{STS}^{\mathsf{H}}}(\gamma)}{\gamma} = \frac{1-\eta}{\gamma}$$

- Identifying terms, we obtain

$$\frac{d}{d\gamma}\mathcal{V}_{\mathbf{STS}^{\mathsf{H}}}(\gamma) = \frac{d}{\gamma}\beta\mathcal{V}_{\mathbf{T}}(\gamma\eta) + \dot{\eta} - \frac{\dot{\eta}}{\eta}$$

- Since for $\eta = 0$ both sides are equal to 0 (initial condition), integrating from 0 to γ we obtain (6).

■

- Going back to our original problem, for the symmetric case we have $T_k = P$ for all k, such that

$$\eta_{\mathbf{T}}(\gamma) = \frac{1}{1+\gamma P}$$

- The sought expression for $\eta_{\mathbf{SS}^\mathsf{H}}(\gamma)$ is obtained by solving the quadratic equation

$$\beta = \frac{1-\eta}{1-\frac{1}{1+P\gamma\eta}}$$

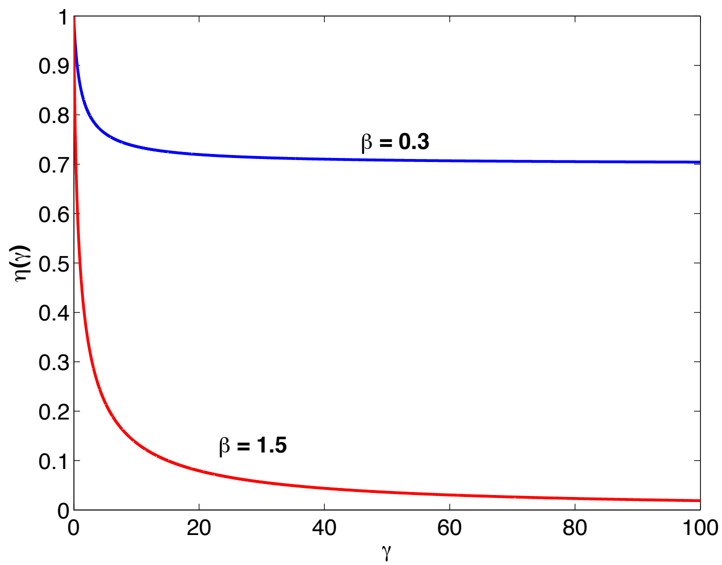
- Recalling that $\gamma = 1/N_0$, and that $\mathsf{snr} = P/N_0$, we can redefine $\gamma = \mathsf{snr}$ and obtain $\eta = \eta_{\mathbf{SS}^\mathsf{H}}(\gamma)$ as the solution of

$$\gamma\eta^2 + (1+\gamma(\beta-1))\eta - 1 = 0$$

- Using the properties of the η-transform, we can choose the root of the above equation corresponding to the sought η-transform:

$$\eta_{\mathbf{SS}^{\mathsf{H}}}(\gamma) = \frac{-(1+\gamma(\beta-1)) + \sqrt{(1+\gamma(\beta-1))^2 + 4\gamma}}{2\gamma}$$

- Finally, using this into the expression of $\mathcal{V}_{\mathbf{SS}^{\mathsf{H}}}(\gamma)$ (see Theorem 4) we obtain the sum rate of our symmetric DS-CDMA system.



Further Results on the i.i.d. Case

- Non-symmetric DS-CDMA: in this case, each user is affected by its own pathloss/shadowing frequency-flat channel gain

$$\mathbf{y} = \mathbf{S}\mathbf{A}\mathbf{x} + \mathbf{z}$$

- Assuming equal transmit power for each user, the sum-rate is given by

$$\mathcal{V}_{\mathbf{S}\mathbf{T}\mathbf{S}^{\mathsf{H}}}(\gamma) = \beta \mathcal{V}_{\mathbf{T}}(\gamma\eta) + \log\frac{1}{\eta} + \eta - 1, \quad \gamma = \mathsf{snr} = \frac{P}{N_0}, \quad \mathbf{T} = \mathbf{A}\mathbf{A}^{\mathsf{H}}$$

and, by Theorem 4 where η is the solution of

$$\beta = \frac{1-\eta}{1-\eta_{\mathbf{T}}(\gamma\eta)}$$

Linear Receivers

- Often practical receivers are constrained to perform linear single-user processing.

- The general structure of a linear receiver for the vector Gaussian MAC channel is

$$\widetilde{x}_k = \mathbf{u}_k^{\mathsf{H}}\mathbf{y} = (\mathbf{u}_h^{\mathsf{H}}\mathbf{s}_k)A_k x_k + \sum_{j \neq k}(\mathbf{u}_k^{\mathsf{H}}\mathbf{s}_j)A_j x_j + \mathbf{u}_k^{\mathsf{H}}\mathbf{z}$$

- The resulting SINR is given by

$$\mathsf{sinr}_k = \frac{|\mathbf{u}_h^{\mathsf{H}}\mathbf{s}_k|^2 T_k}{\mathsf{snr}^{-1}\|\mathbf{u}_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^{\mathsf{H}}\mathbf{s}_j|^2 T_j}$$

with $T_k = |A_k|^2$.

- Among all linear receivers, the one that maximizes the SINR i stye linear MMSE receiver, already discussed before. The resulting SINR is given by

$$\mathsf{sinr}_k = \frac{P_k}{\mathsf{MMSE}_k} - 1 = \mathsf{snr} T_k \mathbf{s}_k^{\mathsf{H}} \left(\mathbf{I} + \mathsf{snr} \sum_{j \neq k} \mathbf{s}_j \mathbf{s}_j^{\mathsf{H}} T_j \right)^{-1} \mathbf{s}_k \xrightarrow{a.s.} \mathsf{snr} T_k \, \eta_{\mathbf{STS}^{\mathsf{H}}}(\mathsf{snr})$$

- Notice that if user k was alone in the system its SINR would be equal to its receiver SNR, $\mathsf{snr}T_k$. Hence, $\eta_{\mathbf{STS}^{\mathsf{H}}}(\mathsf{snr})$ collects the global effect of the multiuser interference on each specific user k.

- In fact, $\eta_{\mathbf{STS}^{\mathsf{H}}}(\mathsf{snr})$ is referred to as the multiuser efficiency of the linear MMSE receiver (this is how the name η-transform was originated in first place).

- The achievable sum rate with linear MMSE is given by

$$R_{\text{sum}}^{\text{mmse}} = \beta \mathbb{E}\left[\log\left(1 + \mathsf{snr}\mathsf{T}\ \eta_{\mathbf{STS}^{\mathsf{H}}}(\mathsf{snr})\right)\right] = \beta \mathcal{V}_{\mathbf{T}}\left(\mathsf{snr}\eta_{\mathbf{STS}^{\mathsf{H}}}(\mathsf{snr})\right)$$

  where $\mathsf{T} \sim F_{\mathbf{T}}(t)$.

- Comparing the optimal sum rate with the sum rate achieved by linear MMSE processing, we arrive at

$$R_{\text{sum}} = R_{\text{sum}}^{\text{mmse}} + \underbrace{\log\frac{1}{\eta} + \eta - 1}_{\text{non-linear gain}}$$

- In passing: this decomposition, observed to hold for a variety of non-Gaussian inputs, AWGN channels, is at the basis of the "MMSE-I" identity [Guo, Verdú, Shamai, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," IT 2005]:

$$\frac{d}{d\gamma} I(X; \sqrt{\gamma}X + Z) = \frac{1}{2}\mathsf{mmse}(X, \sqrt{\gamma}X + Z)$$

MIMO point-to-point

- The capacity of the MIMO channel with perfect CSIR and no CSIT is given by

$$C(\mathsf{snr}) = \max_{P(\mathbf{x}):\mathsf{tr}(\mathbf{\Sigma}_x)\leq P} I(\mathbf{x};\mathbf{Hx}+\mathbf{z}|\mathbf{H})$$

- Maximization of the mutual information:

$$\begin{aligned} I(\mathbf{x};\mathbf{Hx}+\mathbf{z}|\mathbf{H}) &= h\left(\mathbf{Hx}+\mathbf{z}|\,\mathbf{H}\right) - h(\mathbf{z}) \\ &= h\left(\mathbf{Hx}+\mathbf{z}|\,\mathbf{H}\right) - N\log(\pi e N_0) \\ &\leq \mathbb{E}\left[\log\left|\mathbf{I}+\frac{1}{N_0}\mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^{\mathsf{H}}\right|\right] \end{aligned}$$

where the upper bound is achieved by letting $\mathbf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{\Sigma}_x)$.

- In order to obtain the capacity we have to solve a convex optimization problem (maximization with respect to the convex set $\mathcal{S}=\{\mathbf{\Sigma}_x : \mathsf{tr}(\mathbf{\Sigma}_x)\leq P\}$).

- This maximization depends on the statistics of the channel matrix $\mathbf{H}$.

- In the simplest case, $\mathbf{H}$ is formed by i.i.d. elements $\sim \mathcal{CN}(0,1)$ (normalized independent Rayleigh fading).

- This distribution has the unitary invariant property: for any unitary matrix $\mathbf{Q}$ independent of $\mathbf{H}$, $\mathbf{H}$ and $\mathbf{HQ}$ are i.i.d.. It follows that

$$\max_{\mathbf{\Sigma}_x \in \mathcal{S}} \mathbb{E}\left[\log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^{\mathsf{H}}\right|\right] = \max_{\mathbf{\Lambda}_x \in \mathcal{D}} \mathbb{E}\left[\log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{H}\mathbf{\Lambda}_x\mathbf{H}^{\mathsf{H}}\right|\right]$$

where $\mathcal{D}$ is the set of non-negative diagonal matrices with trace not larger than P.

- Letting $\mathbf{\Pi}_\pi$ denote the $M \times M$ permutation matrix corresponding to the permutation π, we have, for any $\Lambda_x \in \mathcal{D}$

$$\begin{aligned}
\mathbb{E}\left[\log\left|\mathbf{I}+\frac{1}{N_0}\mathbf{H}\mathbf{\Lambda}_x\mathbf{H}^{\mathsf{H}}\right|\right] &= \frac{1}{M!}\sum_{\pi}\mathbb{E}\left[\log\left|\mathbf{I}+\frac{1}{N_0}\mathbf{H}\mathbf{\Pi}_\pi\mathbf{\Lambda}_x\mathbf{\Pi}_\pi^{\mathsf{T}}\mathbf{H}^{\mathsf{H}}\right|\right] \\
&\leq \mathbb{E}\left[\log\left|\mathbf{I}+\frac{1}{N_0}\mathbf{H}\left(\frac{1}{M!}\sum_{\pi}\mathbf{\Pi}_\pi\mathbf{\Lambda}_x\mathbf{\Pi}_\pi^{\mathsf{T}}\right)\mathbf{H}^{\mathsf{H}}\right|\right] \\
&\leq \mathbb{E}\left[\log\left|\mathbf{I}+\frac{P}{N_0 M}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right|\right]
\end{aligned}$$

- This upper bound is clearly achievable by letting $\mathbf{\Lambda}_x = (P/M)\mathbf{I}$.

- It follows that the MIMO capacity with perfect CSIR and no CSIT, under the unitary right invariant condition for the channel matrix statistics, is given by

$$C(\mathsf{snr}) = \mathbb{E}\left[\log\left|\mathbf{I}+\frac{\mathsf{snr}}{M}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right|\right]$$

- For the i.i.d. Rayleigh fading case, we can compute this expression exactly for finite M, N (complicated).

- Nevertheless, for large M, N with fixed ratio $M/N = \beta$, we can quickly use the previous developed results to have a very precise expression that yields accurate results also for finite and small M, N.

- Let $\mathbf{S} = \frac{1}{\sqrt{N}}\mathbf{H}$, then

$$C(\mathsf{snr}) = \mathbb{E}\left[\log\left|\mathbf{I} + \frac{\mathsf{snr}}{\beta}\mathbf{S}\mathbf{S}^{\mathsf{H}}\right|\right]$$

- Dividing by N, we obtain $C(\mathsf{snr}) \approx Nc(\mathsf{snr})$ where

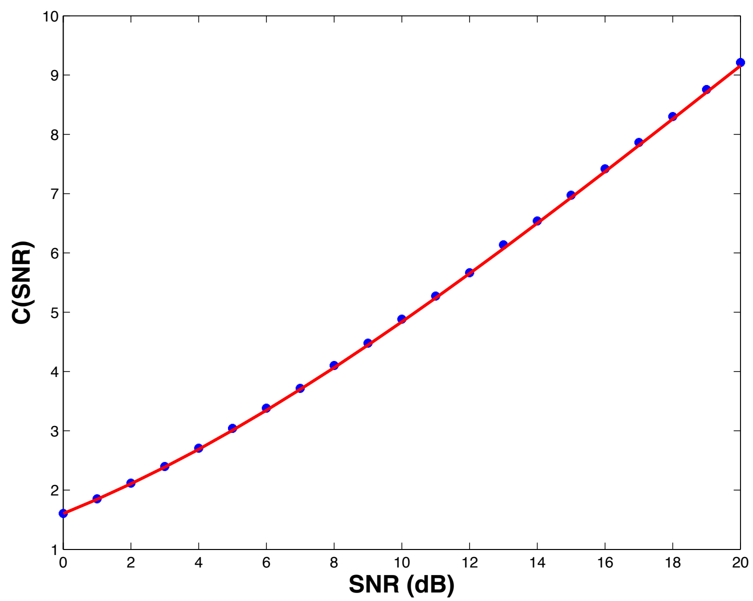
$$\begin{aligned} c(\mathsf{snr}) &= \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I} + \frac{\mathsf{snr}}{\beta}\mathbf{S}\mathbf{S}^{\mathsf{H}}\right|\right] \\ &= \beta\log\left(1 + \frac{\mathsf{snr}}{\beta}\eta\right) + \log\frac{1}{\eta} + \eta - 1 \end{aligned}$$

where η is given by

$$\eta = \frac{-(\beta + \mathsf{snr}(\beta - 1)) + \sqrt{(\beta + \mathsf{snr}(\beta - 1))^2 + 4\beta\mathsf{snr}}}{2\mathsf{snr}}$$

- In order to see this, just notice that this coincides with the previous studied case with $\mathbf{T} = \mathbf{I}$ and $\gamma = \frac{\mathsf{snr}}{\beta}$.

- An example for $N = 3$ and $M = 2$:



End of Lecture 2

Lecture 3:
Matrices with Variance Profile

Structured Matrices $\mathbf{H} = \mathbf{CSA}$

- $\mathbf{H} = \mathbf{CSA}$, with $\mathbf{S}$ i.i.d. $N \times K$ as before (mean zero and variance $1/N$, well-behaved), $\mathbf{C}$ and $\mathbf{A}$ are $N \times N$ and $K \times K$ such that $\mathbf{D} = \mathbf{CC}^{\mathsf{H}}$ and $\mathbf{T} = \mathbf{AA}^{\mathsf{H}}$ have compactly supported LSDs. $\mathbf{C}, \mathbf{S}, \mathbf{A}$ statistically independent.

Theorem 5. *Under the above conditions, as $N \to \infty$ with $K/N = \beta$,*

$$\eta_{\mathbf{HH}^{\mathsf{H}}}(\gamma) = \mathbb{E}\left[\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{D}, \gamma)\right]$$

where $\Gamma_{\mathbf{HH}^{\mathsf{H}}}(d, \gamma)$ is the unique non-negative solution of the following implicit equation:

$$\Gamma_{\mathbf{HH}^{\mathsf{H}}}(d, \gamma) = \frac{1}{1 + \gamma\beta d \mathbb{E}\left[\frac{\mathsf{T}}{1+\gamma \mathsf{T}\mathbb{E}[\mathsf{D}\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{D},\gamma)]}\right]}$$

and where D and T are independent RVs following the LSDs of $\mathbf{D}$ and $\mathbf{T}$, respectively. $\square$

- How do we solve the implicit equation? By discretization.

- Define a suitable discretization $\{d_i : i = 1, \ldots, m\}$ and $\{t_j : j = 1, \ldots, n\}$ (the supports of D andT are bounded).

- For any $\gamma > 0$, we obtain the coupled system of equations given by

$$\begin{aligned}
\Gamma(d_i, \gamma) &= \frac{1}{1 + \gamma\beta d_i \sum_{j=1}^{m} t_j \Upsilon(t_j, \gamma) P_{\mathsf{T}}(j)}, \quad i = 1, \ldots, m \\
\Upsilon(t_j, \gamma) &= \frac{1}{1 + \gamma t_j \sum_{i=1}^{m} d_i \Gamma(d_i, \gamma) P_{\mathsf{D}}(i)}, \quad j = 1, \ldots, n
\end{aligned}$$

- This can be solved recursively, starting from the all-ones initial condition.

- Sanity check: suppose $\mathbf{D} = \mathbf{I}$ and $\mathbf{T} = \mathbf{I}$, then $\mathbf{H}\mathbf{H}^{\mathsf{H}} = \mathbf{S}\mathbf{S}^{\mathsf{H}}$, such that $\eta_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\gamma) = \Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(1, \gamma) = \eta$, and we have

$$\eta = \frac{1}{1 + \gamma\beta \frac{1}{1+\gamma\eta}} \quad \Rightarrow \quad \gamma\eta^2 + (1 + \gamma(\beta - 1))\eta - 1 = 0$$

Theorem 6. *Under the same conditions of Theorem 5 and $\mathbf{S}$ is unitarily invariant, as $N \to \infty$ with $K/N = \beta$, the Shannon transform of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ is given by*

$$\mathcal{V}_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\gamma) = \mathcal{V}_{\mathsf{D}}(\beta\gamma_d) + \beta\mathcal{V}_{\mathsf{T}}(\gamma_t) - \beta\frac{\gamma_d\gamma_t}{\gamma}$$

where γ_d and γ_t are implicitly given by

$$\frac{\gamma_d\gamma_t}{\gamma} = 1 - \eta_{\mathsf{T}}(\gamma_t), \quad \beta\frac{\gamma_d\gamma_t}{\gamma} = 1 - \eta_{\mathsf{D}}(\beta\gamma_d)$$

Furthermore, we have the alternative η-transform expression

$$\eta_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\gamma) = \eta_{\mathsf{D}}(\beta\gamma_d)$$

$\square$

Proof:

- We start with the alternative η-transform expression: using Theorem 5 we have

$$\eta_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\gamma) = \mathbb{E}\left[\frac{1}{1+\beta\gamma_d \mathsf{D}}\right] = \eta_{\mathsf{D}}(\beta\gamma_d)$$

where we define

$$\gamma_t = \gamma\mathbb{E}[\mathsf{D}\Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\mathsf{D},\gamma)], \tag{7}$$

and

$$\gamma_d = \mathbb{E}\left[\frac{\gamma\mathsf{T}}{1+\mathsf{T}\gamma_t}\right]. \tag{8}$$

- Multiplying both sides of (8) by γ_t/γ we find

$$\frac{\gamma_t\gamma_d}{\gamma} = 1 - \eta_{\mathsf{T}}(\gamma_t)$$

consistently with Theorem 6.

- Using the expression of $\Gamma_{\mathbf{HH}^{\mathsf{H}}}(d,\gamma)$ in Theorem 5, rewritten as

$$\Gamma_{\mathbf{HH}^{\mathsf{H}}}(d,\gamma) = \frac{1}{1+\beta d \gamma_d}$$

into (7), we find

$$\gamma_t = \mathbb{E}\left[\frac{\gamma \mathsf{D}}{1+\beta \mathsf{D} \gamma_d}\right],$$

which can be rewritten as

$$\beta \frac{\gamma_t \gamma_d}{\gamma} = 1 - \eta_{\mathsf{D}}(\beta \gamma_d)$$

consistently with Theorem 6.

- In order to prove the Shannon transform expression, we notice that

$$
\begin{aligned}
\mathcal{V}_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\gamma) &= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I}+\gamma\mathbf{C}\mathbf{S}\mathbf{T}\mathbf{S}^{\mathsf{H}}\mathbf{C}^{\mathsf{H}}\right|\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I}+\gamma\mathbf{U}\mathbf{\Lambda}_{\mathsf{D}}^{1/2}\mathbf{V}^{\mathsf{H}}\mathbf{S}\mathbf{\Lambda}_{\mathsf{T}}\mathbf{S}^{\mathsf{H}}\mathbf{V}\mathbf{\Lambda}_{\mathsf{D}}^{1/2}\mathbf{U}^{\mathsf{H}}\right|\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I}+\gamma\mathbf{\Lambda}_{\mathsf{D}}^{1/2}\mathbf{S}\mathbf{\Lambda}_{\mathsf{T}}\mathbf{S}^{\mathsf{H}}\mathbf{\Lambda}_{\mathsf{D}}^{1/2}\right|\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I}+\gamma\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{\mathsf{H}}\right|\right] \\
&= \mathcal{V}_{\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{\mathsf{H}}}(\gamma)
\end{aligned}
$$

where $\widetilde{\mathbf{H}} = \mathbf{\Lambda}_{\mathsf{D}}^{1/2}\mathbf{S}\mathbf{\Lambda}_{\mathsf{T}}^{1/2}$ is an independent but not identically distributed matrix with element variance

$$
\mathbb{E}[|\widetilde{H}_{i,j}|^2] = \frac{d_i t_j}{N}
$$

- This is a special case of a more general class of matrices to be treated next (TO BE CONTINUED).

■

Structured Matrices $\mathbf{H} = \mathbf{A} \odot \mathbf{S}$

- Let $\mathbf{H} = \mathbf{A} \odot \mathbf{S}$, where $\mathbf{S}$ is $N \times K$ as before, and $\mathbf{A}$ is an element weighting matrix with elements $A_{i,j} = \sqrt{P_{i,j}}$, such that $P_{i,j}$ are uniformly bounded and

$$\mathbb{E}[|H_{i,j}|^2] = \frac{P_{i,j}}{N}$$

- We define the variance profile as the function $v^n : [0,1) \times [0,1) \to \mathbb{R}_+$ such that

$$v^N(x,y) = P_{i,j}, \quad \text{for} \ \ (x,y) \in \left[\frac{i-1}{N}, \frac{i}{N}\right) \times \left[\frac{j-1}{K}, \frac{j}{K}\right)$$

- As $N \to \infty$, we assume that $v^N(x,y) \to v(x,y)$ (uniform convergence), where $v(x,y)$ is bounded and measurable.

- The function $v(x,y)$ is referred to as the asymptotic variance profile of $\mathbf{H}$.

Theorem 7. *Under the above conditions, as* $N \to \infty$ *with* $K/N = \beta$*,*

$$\eta_{\mathbf{HH}^{\mathsf{H}}}(\gamma) = \mathbb{E}\left[\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X}, \gamma)\right]$$

with $\Gamma_{\mathbf{HH}^{\mathsf{H}}}(x, \gamma)$ *satisfying the system of coupled fixed-point equations*

$$\Gamma_{\mathbf{HH}^{\mathsf{H}}}(x, \gamma) = \frac{1}{1 + \beta\gamma\mathbb{E}\left[v(x, \mathsf{Y})\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{Y}, \gamma)\right]} \tag{9}$$

$$\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(y, \gamma) = \frac{1}{1 + \gamma\mathbb{E}\left[v(\mathsf{X}, y)\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X}, \gamma)\right]} \tag{10}$$

where X *and* Y *are independent RVs, uniform over* $[0, 1]$. □

Theorem 8. *Under the above conditions, as* $N \to \infty$ *with* $K/N = \beta$*, for any* $a < b$ *with* $a, b \in [0, 1]$ *we have*

$$\frac{1}{N} \sum_{i=\lfloor bN \rfloor}^{\lfloor bN \rfloor} \left[\left(\mathbf{I} + \gamma\mathbf{HH}^{\mathsf{H}}\right)^{-1}\right]_{i,i} \xrightarrow{a.s.} \int_a^b \Gamma_{\mathbf{HH}^{\mathsf{H}}}(x, \gamma) dx$$

where $\Gamma_{\mathbf{HH}^{\mathsf{H}}}(x, \gamma)$ *is defined by (9) - (10).* □

Theorem 9. *Under the above conditions, define the quantity*

$$\mathcal{F}^{(N)}(y,\gamma) = \mathbf{h}_j^{\mathsf{H}}\left(\mathbf{I} + \gamma\sum_{\ell\neq j}\mathbf{h}_\ell\mathbf{h}_\ell^{\mathsf{H}}\right)^{-1}\mathbf{h}_j, \qquad \frac{j-1}{K}\leq y < \frac{j}{K}.$$

As $N\to\infty$ *with* $\frac{K}{N}=\beta$, $\mathcal{F}^{(N)}(y,\gamma)$ *converges almost surely to the limit* $\mathcal{F}(y,\gamma)$, *given by the solution of the fixed-point equation*

$$\mathcal{F}(y,\gamma) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1+\gamma\,\beta\,\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})}{1+\gamma\,\mathcal{F}(\mathsf{Y},\gamma)}\,|\mathsf{X}\right]}\right], \quad y\in[0,1].$$

where X *and* Y *are independent RV, uniform over* $[0,1]$. □

- For a proof of Theorems 7 and 8 see [V. L. Girko, Theory of Random Determinants. Dordrecht: Kluwer Academic Publishers, 1990].

- For a proof of Theorem 9 see [A. M. Tulino, A. Lozano, and S. Verdú, Impact of correlation on the capacity of multi-antenna channels, Bell Labs Technical Memorandum ITD-03-44786F, 2003].

Theorem 10. *Under the above conditions, as $N \to \infty$ with $K/N = \beta$, the Shannon transform of the LSD of $\mathbf{HH}^{\mathsf{H}}$ is given by*

$$\begin{aligned}\mathcal{V}_{\mathbf{HH}^{\mathsf{H}}}(\gamma) &= \beta\mathbb{E}\left[\log\left(1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X},\gamma)|\mathsf{Y}]\right)\right] \\ &+ \mathbb{E}\left[\log\left(1+\gamma\beta\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{Y},\gamma)|\mathsf{X}]\right)\right] \\ &- \gamma\beta\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X},\gamma)\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{Y},\gamma)\right]\end{aligned}$$

where $\Gamma_{\mathbf{HH}^{\mathsf{H}}}(x,\gamma)$ and $\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(y,\gamma)$ are defined by (9) - (10). $\square$

Proof:

- For simplicity of notation we drop the subscript $\mathbf{HH}^{\mathsf{H}}$ everywhere.

- By definition of Shannon transform, we have

$$\mathcal{V}(\gamma) = \int \log(1+\gamma\lambda)dF(\lambda)$$

- Taking the derivative with respect to γ, we have

$$\dot{\mathcal{V}}(\gamma) = \frac{1-\eta(\gamma)}{\gamma} = \frac{1}{\gamma}(1 - \mathbb{E}[\Gamma(\mathsf{X},\gamma)])$$

  where we have used Theorem 7.

- Using (9) - (10) and rearranging terms, we can write

$$\frac{1-\Gamma(x,\gamma)}{\gamma} = \frac{\beta\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]}$$

- Adding and subtracting to the right-hand side the term

$$\frac{\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]}$$

  we obtain

$$\frac{1-\Gamma(x,\gamma)}{\gamma} = \frac{d}{d\gamma}\log\left(1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]\right) - \frac{\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)]}{1+\beta\gamma\mathbb{E}[v(x,\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)]}$$

- Integrating both sides with respect to x, we obtain

$$\begin{aligned}\dot{\mathcal{V}}(\gamma) &= \mathbb{E}\left[\frac{d}{d\gamma}\log\left(1+\beta\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)|\mathsf{X}]\right)\right]\\ &\quad -\beta\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)\Gamma(\mathsf{X},\gamma)\right]\end{aligned}$$

where we used (9).

- We notice that

$$\begin{aligned}-\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\dot{\Upsilon}(\mathsf{Y},\gamma)\Gamma(\mathsf{X},\gamma)\right] &= -\frac{d}{d\gamma}\left(\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)\Gamma(\mathsf{X},\gamma)\right]\right)\\ &\quad +\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)\dot{\Gamma}(\mathsf{X},\gamma)\right]\\ &\quad +\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)\Gamma(\mathsf{X},\gamma)\right]\end{aligned}$$

- Using (10), we can write

$$\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})(\gamma\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma))\Upsilon(\mathsf{Y},\gamma)\right]$$

$$=\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})(\gamma\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma))}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]}\right]$$

$$=\mathbb{E}\left[\frac{\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})(\gamma\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma))|\mathsf{Y}\right]}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]}\right]$$

- Replacing, we arrive at

$$\begin{aligned}\dot{\mathcal{V}}(\gamma) &= \mathbb{E}\left[\frac{d}{d\gamma}\log\left(1+\beta\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)|\mathsf{X}]\right)\right] \\ &\quad -\beta\frac{d}{d\gamma}\left(\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{Y},\gamma)\Gamma(\mathsf{X},\gamma)\right]\right) \\ &\quad +\beta\mathbb{E}\left[\frac{\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})(\gamma\dot{\Gamma}(\mathsf{X},\gamma)+\Gamma(\mathsf{X},\gamma))|\mathsf{Y}\right]}{1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}]}\right]\end{aligned}$$

- Integrating with respect to γ and using $\mathcal{V}(0)=0$, we obtain

$$\begin{aligned}\mathcal{V}_{\mathbf{HH}^{\mathsf{H}}}(\gamma) &= \beta\mathbb{E}\left[\log\left(1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X},\gamma)|\mathsf{Y}]\right)\right] \\ &\quad +\mathbb{E}\left[\log\left(1+\gamma\beta\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{Y},\gamma)|\mathsf{X}]\right)\right] \\ &\quad -\gamma\beta\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X},\gamma)\Upsilon_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{Y},\gamma)\right]\end{aligned}$$

■

Continuation of the Proof of Theorem 6

- Recall that we showed that

$$\mathcal{V}_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\gamma) = \mathcal{V}_{\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{\mathsf{H}}}(\gamma)$$

where $\widetilde{\mathbf{H}} = \mathbf{\Lambda}_{\mathsf{D}}^{1/2}\mathbf{S}\mathbf{\Lambda}_{\mathsf{T}}^{1/2} = (\mathbf{d}\mathbf{t}^{\mathsf{T}}) \odot \mathbf{S}$, such that

$$\mathbb{E}[|\widetilde{H}_{i,j}|^2] = \frac{d_i t_j}{N}$$

- This is a special case of the general variance profile structure, where the limiting variance profile is separable, i.e.,

$$v(x, y) = d(x)t(y)$$

- We introduce the new quantities:

$$\tilde{\Gamma}(\gamma) = \mathbb{E}\left[d(\mathsf{X})\Gamma(\mathsf{X}, \gamma)\right], \quad \text{and} \quad \tilde{\Upsilon}(\gamma) = \mathbb{E}\left[t(\mathsf{Y})\Upsilon(\mathsf{Y}, \gamma)\right]$$

- Rewriting (9) - (10) in this case yields

$$\tilde{\Gamma}(\gamma) = \mathbb{E}\left[\frac{d(\mathsf{X})}{1+\beta\gamma d(\mathsf{X})\tilde{\Upsilon}(\gamma)}\right] \tag{11}$$

$$\tilde{\Upsilon}(\gamma) = \mathbb{E}\left[\frac{t(\mathsf{Y})}{1+\gamma t(\mathsf{Y})\tilde{\Gamma}(\gamma)}\right] \tag{12}$$

- Introducing the RVs $\mathsf{D} = d(\mathsf{X})$ and $\mathsf{T} = t(\mathsf{Y})$, we can rewrite the above system of equations as

$$\tilde{\Gamma}(\gamma) = \frac{1}{\beta\gamma\tilde{\Upsilon}(\gamma)}\left(1-\eta_{\mathsf{D}}\left(\beta\gamma\tilde{\Upsilon}(\gamma)\right)\right) \tag{13}$$

$$\tilde{\Upsilon}(\gamma) = \frac{1}{\gamma\tilde{\Gamma}(\gamma)}\left(1-\eta_{\mathsf{T}}\left(\gamma\tilde{\Gamma}(\gamma)\right)\right) \tag{14}$$

- Also, using the definition of Shannon transform, we have

$$\mathbb{E}\left[\log(1+\gamma\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)|\mathsf{Y}])\right] = \mathbb{E}\left[\log(1+\gamma\mathsf{T}\tilde{\Gamma}(\gamma))\right] = \mathcal{V}_{\mathsf{T}}(\gamma\tilde{\Gamma}(\gamma))$$

and

$$\mathbb{E}\left[\log(1+\gamma\beta\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Upsilon(\mathsf{X},\gamma)|\mathsf{X}])\right] = \mathbb{E}\left[\log(1+\gamma\beta\mathsf{D}\tilde{\Upsilon}(\gamma))\right] = \mathcal{V}_{\mathsf{D}}(\gamma\beta\tilde{\Upsilon}(\gamma))$$

and using separability and the fact that X and Y are independent, we have

$$\gamma\beta\mathbb{E}[v(\mathsf{X},\mathsf{Y})\Gamma(\mathsf{X},\gamma)\Upsilon(\mathsf{Y},\gamma)] = \gamma\beta\tilde{\Gamma}(\gamma)\tilde{\Upsilon}(\gamma)$$

- Eventually, defining

$$\gamma_t = \gamma\tilde{\Gamma}(\gamma), \quad \text{and} \quad \gamma_d = \gamma\tilde{\Upsilon}(\gamma)$$

and putting everything together, we obtain the desired result.

■

A Particularly Simple Case: Doubly-Regular Matrices

- Asymptotic row regularity: if the variance profile $P_{i,j}$ of $\mathbf{H}$ satisfies (for all $\alpha \in \mathbb{R}$)

$$\frac{1}{K}\sum_{j=1}^{K} 1\{P_{i,j} \leq \alpha\} \to G(\alpha), \quad \text{for all rows } \; i$$

- Asymptotic column regularity: $P_{i,j}$ of $\mathbf{H}$ satisfies (for all $\alpha \in \mathbb{R}$)

$$\frac{1}{N}\sum_{i=1}^{N} 1\{P_{i,j} \leq \alpha\} \to G'(\alpha), \quad \text{for all columns } \; j$$

- The matrix is both asymptotically row regular and asymptotically column regular, we say that it is asymptotically doubly regular. In this case

$$\lim_{K\to\infty} \frac{1}{K}\sum_{j=1}^{K} P_{i,j} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} P_{i,j}$$

- If the above limits are equal to 1, we say that the variance profile is standard.

Theorem 11. *Consider a matrix $\mathbf{H}$ in the same conditions of Theorem 7, such that its variance profile is standard asymptotically doubly regular. Then, the LSD of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ coincides with that of $\mathbf{S}\mathbf{S}^{\mathsf{H}}$ (as if the elements of $\mathbf{H}$ were i.i.d.).* $\square$

Proof:

- Combining (9) and (10), we have

$$\eta_{\mathbf{HH}^{\mathsf{H}}}(\gamma) = \mathbb{E}\left[\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X}, \gamma)\right]$$

with

$$\Gamma_{\mathbf{HH}^{\mathsf{H}}}(x, \gamma) = \frac{1}{1 + \beta\gamma\mathbb{E}\left[\frac{v(x,\mathsf{Y})}{1+\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X},\gamma)|\mathsf{Y}\right]}\right]}$$

- We neglect again the subscript $\mathbf{HH}^{\mathsf{H}}$ for simplicity of notation, and notice that because of the column regular condition we have that

$$\mathbb{E}[v(\mathsf{X}, y)\Gamma(\mathsf{X}, \gamma)] = \mu(\gamma), \qquad \forall\ y$$

- Also, because of the row regular condition we have that

$$\mathbb{E}\left[\frac{v(x,\mathsf{Y})}{1+\gamma\mathbb{E}\left[v(\mathsf{X},\mathsf{Y})\Gamma_{\mathbf{HH}^{\mathsf{H}}}(\mathsf{X},\gamma)|\mathsf{Y}\right]}\right]=\mathbb{E}\left[\frac{v(x,\mathsf{Y})}{1+\gamma\mu(\gamma)}\right]=\frac{\mathbb{E}[v(x,\mathsf{Y})]}{1+\gamma\mu(\gamma)}$$

is independent of x.

- We conclude that $\Gamma(x,\gamma)=\Gamma(\gamma)$, independent of x.
- Letting $\mu(\gamma)=\mathbb{E}[v(\mathsf{X},y)\Gamma(\mathsf{X},\gamma)]=\Gamma(\gamma)\mathbb{E}[v(\mathsf{X},y)]=\Gamma(\gamma)$, since by the standardization condition we have

$$\mathbb{E}[v(\mathsf{X},y)]=\mathbb{E}[v(x,\mathsf{Y})]=1$$

we arrive at $\eta_{\mathbf{HH}^{\mathsf{H}}}(\gamma)=\Gamma(\gamma)$, where

$$\Gamma(\gamma)=\frac{1}{1+\beta\gamma\frac{1}{1+\gamma\Gamma(\gamma)}}$$

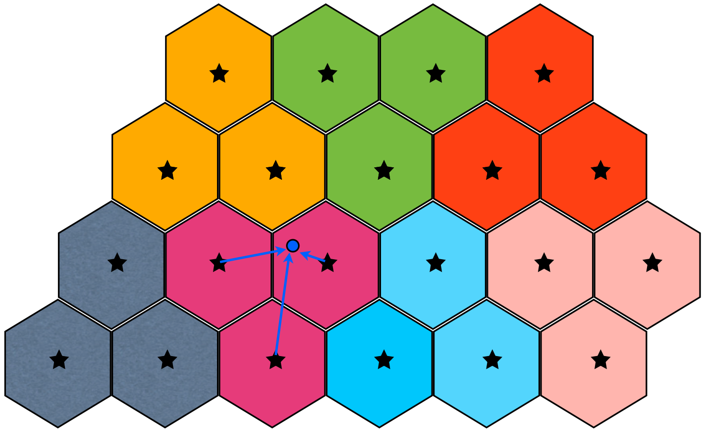
same as the key equation (5) of Theorem 4 for the matrix $\mathbf{SS}^{\mathsf{H}}$.

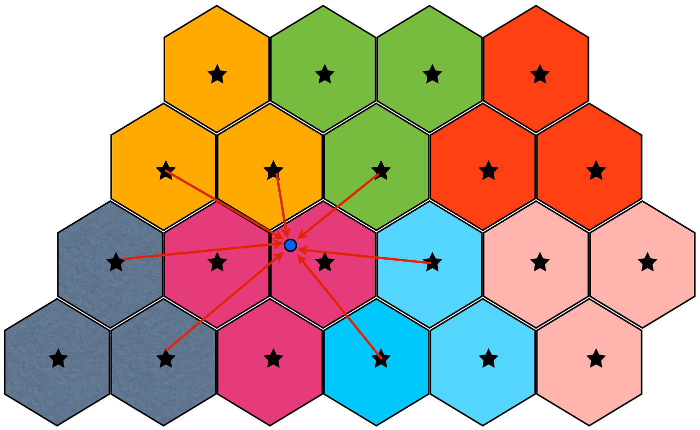
■

End of Lecture 3

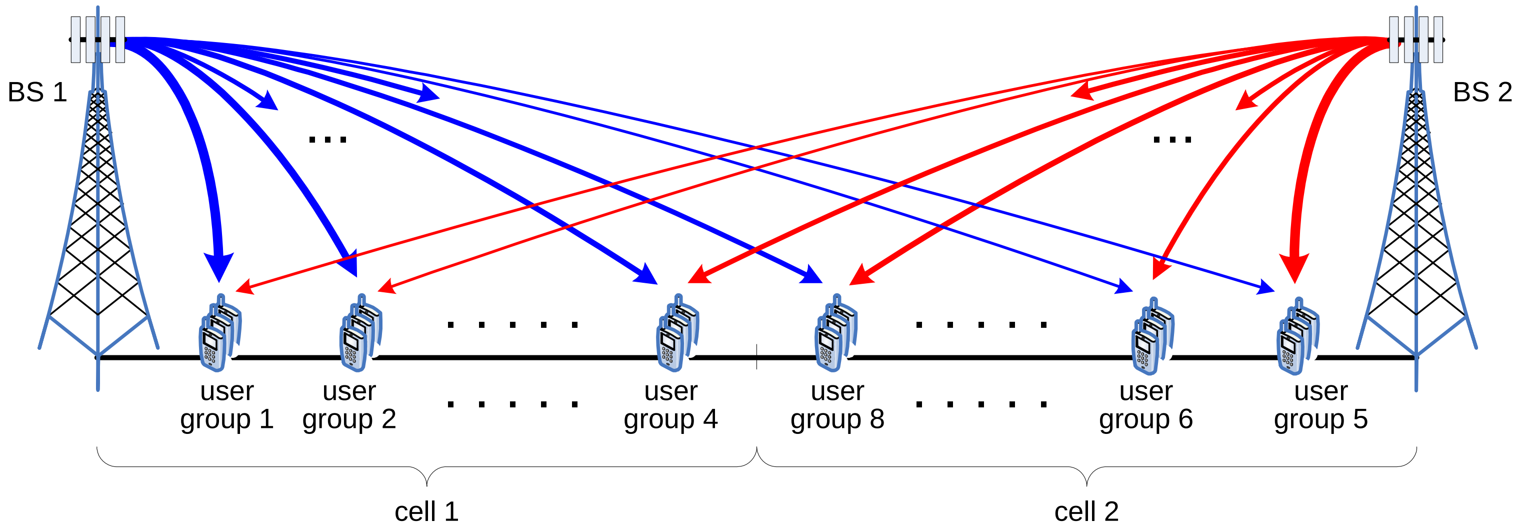
Lecture 4:
Multi-Cell Wireless Networks

Multi-Cell Network Model





Discretization of the Users Distribution



- We assume that the users are partitioned in co-located groups with N single-antenna terminals each.

- We have A user groups per cluster, and clusters of B cells.

- We have $M = \rho N$ base station antennas per cell.

Multiuser MIMO Downlink

- One channel use of the multi-cell MU-MIMO downlink is described by

$$\mathbf{y}_k = \sum_m \alpha_{m,k} \mathbf{H}_{m,k}^{\mathsf{H}} \mathbf{x}_m + \mathbf{n}_k,$$

  for each user location k.

- $\mathbf{H}_{m,k}$ is the $\rho N \times N$ small-scale fading channel matrix from the m-th BS to the k-th user group, with i.i.d. $\sim \mathcal{CN}(0,1)$ elements.

- The per-BS average power constraint is expressed by tr$\left(\mathrm{Cov}(\mathbf{x}_m)\right) \leq P_m$.

- We consider cooperating clusters of BSs (or sectors), such that each cooperating cluster jointly process its signals and serves its user locations using MU-MIMO, and take interference from other clusters.

- More in general: we can consider a multi-band architecture where different intertwined patterns of cooperating clusters are defined for each subband, in order to symmetrize the user performance and avoid "locations in the boundary" on the whole system bandwidth.

- For a given location k in the reference cluster $\mathcal{M}$, the Inter-Cluster Interference (ICI) plus noise variance at any user group k is given by

$$\sigma_k^2 = \mathbb{E}\left[\frac{1}{N}\left\|\sum_{m\notin\mathcal{M}} \alpha_{m,k}\mathbf{H}_{m,k}^{\mathsf{H}}\mathbf{x}_m + \mathbf{n}_k\right\|^2\right] = 1 + \sum_{m\notin\mathcal{M}} \alpha_{m,k}^2 P_m.$$

- Letting A and B denote the number of user locations and BSs in the reference cluster $\mathcal{M}$, and re-normalizing the signal at each user location k by σ_k, such that its ICI plus noise variance is 1, we obtain

$$\mathbf{H} = \left[\begin{array}{ccc} \beta_{1,1}\mathbf{H}_{1,1} & \cdots & \beta_{1,A}\mathbf{H}_{1,A} \\ \vdots & \ddots & \vdots \\ \beta_{B,1}\mathbf{H}_{B,1} & \cdots & \beta_{B,A}\mathbf{H}_{B,A} \end{array} \right],$$

  with $\beta_{m,k} = \frac{\alpha_{m,k}}{\sigma_k}$.

- It follows that from our reference cluster point of view, the relevant downlink channel model is given by

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{v} \tag{15}$$

  with $\mathbf{y} = \mathbb{C}^{AN}$, $\mathbf{x} = \mathbb{C}^{\rho BN}$, and $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

Gaussian Vector Broadcast Channel: A Primer

- Vector Gaussian BC:

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z}$$

- [Caire-Shamai, IT 2003, Viswanath and Tse, IT 2004, Vishwanath, Jindal and Goldsmith, IT 2004, Yu and Cioffi, IT 2004, Weingarten, Steinberg and Shamai, IT 2005].

- Let π denote a precoding order, such that signals are encoded in the order $\pi(K), \pi(K-1), \ldots, \pi(1)$.

- The Dirty-Paper Coding (DPC) achievable region $\mathcal{R}_{\pi}^{\text{dpc}}(\mathbf{H}; \mathbf{S}_{1:K})$ is given by the set of rate points $(R_1, \ldots, R_K)$ such that

$$R_{\pi(k)} \leq \log \frac{1 + \mathbf{h}_{\pi(k)}^{\mathsf{H}} \left(\sum_{j=1}^{k} \mathbf{S}_{\pi(j)} \right) \mathbf{h}_{\pi(k)}}{1 + \mathbf{h}_{\pi(k)}^{\mathsf{H}} \left(\sum_{j=1}^{k-1} \mathbf{S}_{\pi(j)} \right) \mathbf{h}_{\pi(k)}}$$

- The capacity region of the vector Gaussian BC subject to any convex covariance constraint $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}] \in \mathcal{S}$ is given by

$$\mathcal{C}^{\mathrm{BC}}(\mathbf{H}; \mathcal{S}) = \textbf{coh} \left\{ \bigcup_{\pi} \bigcup_{\sum_{k=1}^{K} \mathbf{S}_k \in \mathcal{S}} \mathcal{R}_{\pi}^{\mathrm{dpc}}(\mathbf{H}; \mathbf{S}_{1:K}) \right\}$$

- For the sum power constraint, the set $\mathcal{S}$ is given by $\{\mathbf{\Sigma}_x \preceq 0 : \mathrm{tr}(\mathbf{\Sigma}_x) \leq P\}$.

- Similarly, we can consider per-antenna power constraint, per group of antenna power constraint, or more general linear constraints in the form

$$\mathrm{tr}(\mathbf{\Sigma}_x \mathbf{\Phi}_\ell) \leq P_\ell, \quad \ell = 1, \ldots, c$$

  for some $\mathbf{\Phi}_\ell \preceq 0$ constraint matrices.

- How to calculate points on the boundary of $\mathcal{C}(\mathbf{H}; \mathcal{S})$? Uplink-Downlink Duality.

- Consider the "dual" vector Gaussian MAC:

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{w}$$

- Let π denote a successive decoding order, such that users are decoded in the order $\pi(1), \pi(2), \ldots, \pi(K)$.

- The Successive Interference Cancellation (SIC) achievable region $\mathcal{R}_{\pi}^{\text{sic}}(\mathbf{H}; q_{1:K})$ is the set of rate points $(R_1, \ldots, R_K)$ such that

$$R_{\pi(k)} \leq \log \frac{\left|\mathbf{I} + \sum_{j=k}^{K} \mathbf{h}_{\pi(j)} \mathbf{h}_{\pi(j)}^{\mathsf{H}} q_{\pi(j)}\right|}{\left|\mathbf{I} + \sum_{j=k+1}^{K} \mathbf{h}_{\pi(j)} \mathbf{h}_{\pi(j)}^{\mathsf{H}} q_{\pi(j)}\right|}$$

- It is well-known that the MAC capacity region for user powers $q_1, \ldots, q_K$ is given by

$$\mathcal{C}^{\text{MAC}}(\mathbf{H}; q_{1:K}) = \textbf{coh} \left\{ \bigcup_{\pi} \mathcal{R}_{\pi}^{\text{sic}}(\mathbf{H}; q_{1:K}) \right\}$$

- Duality subject to sum power constraint: for $\mathcal{S} = \{\boldsymbol{\Sigma}_x \preceq 0 : \text{tr}(\boldsymbol{\Sigma}_x) \leq P\}$ we have

$$\mathcal{C}^{\text{BC}}(\mathbf{H}; \mathcal{S}) = \bigcup_{\sum_{k=1}^{K} q_k \leq P} \mathcal{C}^{\text{MAC}}(\mathbf{H}; q_{1:K})$$

- It turns out that it is much easier to compute boundary points on the MAC capacity region than on the BC capacity region.

- This is due to the fact that for fixed powers $q_1, \ldots, q_K$, and channel matrix $\mathbf{H}$ the MAC capacity region is a polymatroid.

- Weighted sum-rate maximization for the MAC:

$$\begin{array}{ll} \text{maximize} & \sum_k W_k R_k \\ \text{subject to} & (R_1, \ldots, R_K) \in \mathcal{C}^{\text{MAC}}(\mathbf{H}; q_{1,\ldots,K}) \end{array}$$

- Result: the solution is the vertex π that orders the weights in increasing order, i.e.,

$$W_{\pi(1)} \leq W_{\pi(2)} \leq \cdots \leq W_{\pi(K)}$$

- In this case, the objective function become

$$\sum_{k=1}^{K} W_{\pi(k)} R_{\pi(k)}^{\text{sic}}(\mathbf{H}; q_{1:K}) = \sum_{k=1}^{K} \left(W_{\pi(k)} - W_{\pi(k-1)}\right) \log \left| \mathbf{I} + \sum_{j=k}^{K} \mathbf{h}_{\pi(j)} \mathbf{h}_{\pi(j)}^{\mathsf{H}} q_{\pi(j)} \right|$$

where, for convenience, we let $\pi(0) = 0$ and $W_0 = 0$.

- In this way, we get rid of the combinatorial problem of choosing the optimal decoding order (otherwise we have to search over all $K!$ orders).

- Since the resulting function is concave, the optimization with respect to the input powers $q_1, \ldots, q_K$ is easily accomplished.

Back to our case ....

- The case of per-antenna or per-group of antennas power constraint is more involved: see [W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," Transactions on Signal Processing, 2007].

- The boundary of the capacity region of the Vector Gaussian BC (15) for fixed channel matrix $\mathbf{H}$ and given per-group-of-antennas power constraints $\{P_1, \ldots, P_B\}$ can be characterized by the solution of a min-max weighted sum-rate problem.

- By symmetry, we restrict ourselves to the case of identical weights for users in the same group.

- Let W_k and $R_k(\mathbf{H}) = \frac{1}{N} \sum_{i=1}^{N} R_{k,i}(\mathbf{H})$ denote the weight for user group k and the corresponding instantaneous per-user rate, respectively.

- Let π denote the permutation that sorts the weights in increasing order $W_{\pi_1} \leq \ldots \leq W_{\pi_A}$.

- Let $\mathbf{H}_k$ denote the k-th $\rho BN \times N$ slice of $\mathbf{H}$.

- Let $\mathbf{Q}_k = \text{diag}(q_{k,1}, \ldots, q_{k,N})$ denote an $N \times N$ non-negative definite diagonal matrix of the dual uplink users' transmit powers.

- Let $\mathbf{Q} = \text{diag}(\mathbf{Q}_1, \ldots, \mathbf{Q}_A)$ and, for given permutation π, let $\mathbf{H}_{k:A} = \left[\mathbf{H}_{\pi_k} \ldots \mathbf{H}_{\pi_A}\right]$ and $\mathbf{Q}_{k:A} = \text{diag}\left(\mathbf{Q}_{\pi_k}, \ldots, \mathbf{Q}_{\pi_A}\right)$.

- The rate point $\{R_k(\mathbf{H}, W_1, \ldots, W_A)\}$ on the boundary of the instantaneous capacity region corresponding to weights $\{W_1, \ldots, W_A\}$ is obtained as solution of the min-max problem

$$\min_{\boldsymbol{\lambda} \geq 0} \max_{\mathbf{Q} \geq 0} \sum_{k=1}^{A} W_{\pi_k} R_{\pi_k}(\mathbf{H}) \tag{16}$$

where the instantaneous per-user rate of each group takes on the expression

$$R_{\pi_k}(\mathbf{H}) = \frac{1}{N} \log \frac{\left|\mathbf{\Sigma}(\boldsymbol{\lambda}) + \mathbf{H}_{k:A}\mathbf{Q}_{k:A}\mathbf{H}_{k:A}^{\mathsf{H}}\right|}{\left|\mathbf{\Sigma}(\boldsymbol{\lambda}) + \mathbf{H}_{k+1:A}\mathbf{Q}_{k+1:A}\mathbf{H}_{k+1:A}^{\mathsf{H}}\right|},$$

where $\mathbf{\Sigma}(\boldsymbol{\lambda})$ is a $\rho BN \times \rho BN$ block-diagonal matrix with $\rho N \times \rho N$ constant diagonal blocks $\lambda_m \mathbf{I}_{\rho N}$, for $m = 1, \ldots, B$ and the maximization with respect to $\mathbf{Q}$ is subject to the trace constraint

$$\mathrm{tr}(\mathbf{Q}) \leq \sum_{m=1}^{B} \lambda_m P_m.$$

- Lagrange Multipliers: The variables $\boldsymbol{\lambda} = \{\lambda_m\}$ are the Lagrange multipliers corresponding to the per-group-of-antennas power constraints.

- Ergodic capacity region (inner bound):

$$\underline{\mathcal{C}}(P_1, \ldots, P_B) = \textbf{coh} \bigcup_{W_1,\ldots,W_A \geq 0} \Big\{\mathbf{R} : 0 \leq R_{k,i} \leq \mathbb{E}\left[R_k(\mathbf{H}, W_1, \ldots, W_A)\right], \\ \forall k = 1, \ldots, A,\ \forall i = 1, \ldots, N\Big\} \qquad (17)$$

- Fairness scheduling problem: let $U(\mathbf{R})$ denote a strictly increasing and concave network utility of the ergodic user rates. Then:

$$\begin{aligned} \text{maximize} \quad & U(\mathbf{R}) \\ \text{subject to} \quad & \mathbf{R} \in \underline{\mathcal{C}}(P_1, \ldots, P_B) \end{aligned} \tag{18}$$

Workplan

- In finite dimension, by applying the already mentioned stochastic optimization framework, the solution $\mathbf{R}^{\star}$ of (18) can be approached by solving a sequence of instantaneous weighted sum-rate maximizations.

- In the large system limit, $N \to \infty$, we can directly compute $\mathbf{R}^{\star}$ by combining asymptotic RMT and convex optimization.

- Preliminary problem: for fixed $\boldsymbol{\lambda}$, solve

$$\begin{aligned} &\text{maximize} \quad \sum_{k=1}^{A} W_{\pi_k} \frac{1}{N} \mathbb{E}\left[\log \frac{\left|\boldsymbol{\Sigma}(\boldsymbol{\lambda}) + \mathbf{H}_{k:A}\mathbf{Q}_{k:A}\mathbf{H}_{k:A}^{\mathsf{H}}\right|}{\left|\boldsymbol{\Sigma}(\boldsymbol{\lambda}) + \mathbf{H}_{k+1:A}\mathbf{Q}_{k+1:A}\mathbf{H}_{k+1:A}^{\mathsf{H}}\right|}\right] \\ &\text{subject to} \quad \mathrm{tr}(\mathbf{Q}) \leq Q \end{aligned} \tag{19}$$

where $Q \triangleq \sum_{m=1}^{B} \lambda_m P_m$.

Solution for finite N

- Letting $\Delta_k \triangleq W_{\pi_k} - W_{\pi_{k-1}}$ with $W_{\pi_0} = 0$, the objective function in (19) can be written as

$$F_{\mathbf{W},\boldsymbol{\lambda}}(\mathbf{Q}) = \sum_{k=1}^{A} \Delta_k \frac{1}{N} \mathbb{E}\left[\log\left|\boldsymbol{\Sigma}(\boldsymbol{\lambda}) + \mathbf{H}_{k:A}\mathbf{Q}_{k:A}\mathbf{H}_{k:A}^{\mathsf{H}}\right|\right] - W_{\pi_A}\frac{1}{N}\log\left|\boldsymbol{\Sigma}(\boldsymbol{\lambda})\right|$$

- The following results follow from the symmetry of the problem:

Lemma 1. *The optimal $\mathbf{Q}$ in (19) allocates equal power to the users in the same group.* $\square$

Lemma 2. *The optimal $\boldsymbol{\lambda}^\star$ for the min-max problem (16) are strictly positive, i.e., $\boldsymbol{\lambda}^\star > \mathbf{0}$.* $\square$

- We can restrict to consider $\mathbf{Q}$ with constant diagonal blocks $\mathbf{Q}_k = \frac{Q_k}{N}\mathbf{I}$.

- We define the modified channel matrix $\overline{\mathbf{H}}_k \triangleq \frac{1}{\sqrt{N}}\mathbf{\Sigma}^{-1/2}(\boldsymbol{\lambda})\mathbf{H}_k$ and rewrite the objective function as

$$F_{\mathbf{W},\boldsymbol{\lambda}}(Q_1,\dots,Q_A) = \sum_{k=1}^{A} \Delta_k \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}\right|\right]$$

subject to the trace constraint $\sum_{k=1}^{A} Q_k \leq Q$.

- The Lagrangian function of our problem becomes

$$\mathcal{L}(Q_1,\dots,Q_A;\xi) = F_{\mathbf{W},\boldsymbol{\lambda}}(Q_1,\dots,Q_A) - \xi\left(\sum_{k=1}^{A} Q_k - Q\right)$$

- Using the differentiation rule $\partial \log |\mathbf{X}| = \mathrm{tr}(\mathbf{X}^{-1}\partial\mathbf{X})$, we write the KKT conditions as

$$\frac{\partial \mathcal{L}}{\partial Q_{\pi_j}} = \sum_{k=1}^{j} \frac{\Delta_k}{N} \mathbb{E}\left[\mathrm{tr}\left(\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}} \left[\mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell} \overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}\right]^{-1} \overline{\mathbf{H}}_{\pi_j}\right)\right] \leq \xi$$

for $j = 1, \ldots, A$, where equality must hold at the optimal point for all j such that $Q_{\pi_j} > 0$.

- After some algebra and the application of the matrix inversion lemma, the trace in the KKT conditions can be rewritten in the more convenient form

$$\begin{aligned}
&\frac{1}{N}\mathrm{tr}\left(\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}} \left[\mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell} \overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}\right]^{-1} \overline{\mathbf{H}}_{\pi_j}\right) \\
&= \frac{1}{N}\mathrm{tr}\left(\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}} \boldsymbol{\Theta}_{k:A\setminus j}^{-1} \overline{\mathbf{H}}_{\pi_j} \left[\mathbf{I} + Q_{\pi_j} \overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}} \boldsymbol{\Theta}_{k:A\setminus j}^{-1} \overline{\mathbf{H}}_{\pi_j}\right]^{-1}\right) \\
&= \frac{1 - \mathsf{MMSE}_{k:A}^{(j)}}{Q_{\pi_j}}
\end{aligned}$$

where we let $\boldsymbol{\Theta}_{k:A\backslash j} = \mathbf{I} + \sum_{\ell=k,\ell\neq j}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}$.

- $\mathsf{MMSE}_{k:A}^{(j)}$ denotes the per-component MMSE for the estimation of $\mathbf{s}_j$ from $\mathbf{r}_{[k:A]}$, for fixed (known) matrices $\overline{\mathbf{H}}_{\pi_k}, \ldots, \overline{\mathbf{H}}_{\pi_A}$, for the observation model

$$\mathbf{r}_{[k:A]} = \sum_{\ell=k}^{A} \sqrt{Q_{\pi_\ell}}\, \overline{\mathbf{H}}_{\pi_\ell}\mathbf{s}_\ell + \mathbf{z}, \tag{20}$$

where $\mathbf{s}_k, \mathbf{s}_{K+1}, \ldots, \mathbf{s}_A$ and $\mathbf{z}$ are Gaussian independent vectors with i.i.d. components $\sim \mathcal{CN}(0,1)$.

- Explicitly, we have

$$\begin{aligned}
\mathsf{MMSE}_{k:A}^{(j)} &= \frac{1}{N}\mathrm{tr}\left(\mathbf{I} - Q_{\pi_j}\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}}\left[\overline{\mathbf{H}}_{\pi_j}\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}} Q_{\pi_j} + \boldsymbol{\Theta}_{k:A\backslash j}\right]^{-1}\overline{\mathbf{H}}_{\pi_j}\right) \\
&= \frac{1}{N}\mathrm{tr}\left(\left[\mathbf{I} + Q_{\pi_j}\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}}\boldsymbol{\Theta}_{k:A\backslash j}^{-1}\overline{\mathbf{H}}_{\pi_j}\right]^{-1}\right)
\end{aligned} \tag{21}$$

- Solving for the Lagrange multiplier, we find

$$\xi = \frac{1}{Q}\sum_{\ell=1}^{A}\sum_{k=1}^{\ell}\Delta_k(1-\mathbb{E}[\mathsf{MMSE}_{k:A}^{(\ell)}])$$

- Finally, we arrive at the KKT conditions

$$Q_{\pi_j} = Q\frac{\sum_{k=1}^{j}\Delta_k(1-\mathbb{E}[\mathsf{MMSE}_{k:A}^{(j)}])}{\sum_{\ell=1}^{A}\sum_{k=1}^{\ell}\Delta_k(1-\mathbb{E}[\mathsf{MMSE}_{k:A}^{(\ell)}])} \tag{22}$$

  for all j such that $Q_{\pi_j} > 0$.

- For all j such that $Q_{\pi_j} = 0$, the following inequality must hold

$$Q\sum_{k=1}^{j}\frac{\Delta_k}{N}\mathbb{E}\left[\mathrm{tr}\left(\overline{\mathbf{H}}_{\pi_j}^{\mathsf{H}}\mathbf{\Theta}_{k:A\setminus j}^{-1}\overline{\mathbf{H}}_{\pi_j}\right)\right] \leq \sum_{\ell=1}^{A}\sum_{k=1}^{\ell}\Delta_k(1-\mathbb{E}[\mathsf{MMSE}_{k:A}^{(\ell)}]) \tag{23}$$

Theorem 12. *The solution $Q_1^\star, \ldots, Q_A^\star$ of problem (19) is given as follows. For all j for which (23) is satisfied, then $Q_{\pi_j}^\star = 0$. Otherwise, the positive $Q_{\pi_j}^\star$ satisfy (22).* $\square$

- In finite dimension, an iterative algorithm that provably converges to the solution can be obtained.

- The amount of calculation is tremendous because the average MMSE terms must be computed by Monte Carlo simulation.

- In the limit for $N \to \infty$, this is greatly simplified because the arguments of the expectations converge to deterministic quantities, that we can compute directly.

Large system limit $N \to \infty$

- Normalized row and column indices $r, t \in [0, 1)$.
- Aspect ratio of the matrix $\nu = \frac{A}{\rho B}$.
- $\mathcal{Q}(t)$: (dual uplink) transmit power profile:

$$\mathcal{Q}(t) = Q_{\pi_k}, \quad \text{for } \frac{k-1}{A} \leq t < \frac{k}{A}.$$

- $\mathcal{G}(r, t)$: channel gain profile:

$$\mathcal{G}(r,t) = \frac{\beta^2_{m,\pi_k}}{\lambda_m} \quad \text{for } \frac{m-1}{B} \leq r < \frac{m}{B}, \quad \frac{k-1}{A} \leq t < \frac{k}{A}.$$

- $\Omega_{k:A}(t)$: average per-component MMSE profile:

$$\Omega_{k:A}(t) = \text{MMSE}_{k:A}^{(j)} \quad \text{for } \frac{k-1}{A} \leq t < 1.$$

- $\mathcal{F}_{k:A}(t)$: SINR profile:

$$\mathcal{F}_{k:A}(t) = \frac{1}{\Omega_{k:A}(t)} - 1.$$

Theorem 13. *As $N \to \infty$, for each $k = 1, \ldots, A$, the SINR functions $\mathcal{F}_{k:A}(t)$ satisfy the fixed-point equation*

$$\mathcal{F}_{k:A}(t) = \int_0^1 \frac{\rho B \mathcal{G}(r,t) \mathcal{Q}(t)\, dr}{1 + \nu \int_{(k-1)/A}^1 \frac{\rho B \mathcal{G}(r,\tau) \mathcal{Q}(\tau)\, d\tau}{1 + \mathcal{F}_{k:A}(\tau)}} \tag{24}$$

Also, the asymptotic $\Omega_{k:A}(t)$ is given in terms of the asymptotic $\mathcal{F}_{k:A}(t)$ as $\Omega_{k:A}(t) = 1/(1 + \mathcal{F}_{k:A}(t))$. □

Proof:

Consider the dual uplink model (20)

$$\mathbf{r}_{[k:A]} = \sum_{\ell=k}^{A} \sqrt{Q_{\pi_\ell}}\, \overline{\mathbf{H}}_{\pi_\ell}\mathbf{s}_\ell + \mathbf{z}$$

The SINR for any user in group $k \leq j \leq A$, is given by

$$\mathsf{sinr}_{\pi_j} \approx Q_{\pi_j}\overline{\mathbf{h}}_{\pi_j}^{\mathsf{H}} \left(\mathbf{I} + \sum_{\ell=k,\ell\neq j}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell} \right)^{-1} \overline{\mathbf{h}}_{\pi_j}$$

We apply Theorem 9, with the caveat that the variance profile of the matrix $\{\overline{\mathbf{H}}_{\pi_\ell} : \ell = k, \ldots, A\}$ is given by

$$v(r,t) = \begin{cases} \rho B \mathcal{G}(r,t)\mathcal{Q}(t) & \text{for } \frac{k-1}{A} \leq t < 1 \\ 0 & \text{for } 0 \leq t < \frac{k-1}{A} \end{cases}$$

■

- Using the fact that all these profile functions are piecewise constant, and defining

$$\mathcal{F}_{k:A}(t) = \mathcal{F}_{k:A}^{(j)}, \quad \text{for } \frac{j-1}{A} \leq t < \frac{j}{A}, \quad j \geq k$$

we obtain

$$\mathcal{F}_{k:A}^{(j)} = \rho \sum_{m=1}^{B} \frac{(\beta_{m,\pi_j}^2/\lambda_m)Q_{\pi_j}}{1 + \sum_{\ell=k}^{A} \frac{(\beta_{m,\pi_\ell}^2/\lambda_m)Q_{\pi_\ell}}{1+\mathcal{F}_{k:A}^{(\ell)}}}. \tag{25}$$

- We also let the MMSE quantities of interest be given by

$$\Omega_{k:A}^{(j)} = \frac{1}{1 + \mathcal{F}_{k:A}^{(j)}}$$

- Finally, we obtain a simple iterative algorithm to compute the optimal power allocation profile for given weights $\{W_k\}$ and Lagrange multipliers $\{\lambda_m\}$.

Power optimization in the large system regime

- For notation simplicity we let $\pi_k = k$ for all $k = 1, \ldots, A$ (arbitrary π is immediately obtained by reordering).
- Initialize $Q_k(0) = Q/A$ for all $k = 1, \ldots, A$.
- For $i = 0, 1, 2, \ldots$, iterate until the following solution settles:

$$Q_j(i+1) = Q \frac{\sum_{k=1}^{j} \Delta_k (1 - \Omega_{k:A}^{(j)}(i))}{\sum_{\ell=1}^{A} \sum_{k=1}^{\ell} \Delta_k (1 - \Omega_{k:A}^{(\ell)}(i))}, \tag{26}$$

for $j = 1, \ldots, A$, where $\Omega_{k:A}^{(j)}(i) = 1/(1 + F_{k:A}^{(j)}(i))$, and $F_{k:A}^{(j)}(i)$ is obtained as the solution of the system of fixed point equations (25), for powers $Q_k = Q_k(i), \ \forall k$.

- Denote by $F^{(j)}_{k:A}(\infty)$, $\Omega^{(j)}_{k:A}(\infty)$ and by $Q_j(\infty)$ the fixed points reached by the iteration at step 2). If the condition

$$Q\sum_{k=1}^{j}\Delta_k F^{(j)}_{k:A}(\infty) \leq \sum_{\ell=1}^{A}\sum_{k=1}^{\ell}\Delta_k\left(1-\Omega^{(\ell)}_{k:A}(\infty)\right) \tag{27}$$

 is satisfied for all j such that $Q_j(\infty)=0$, then stop.

- Otherwise, go back to the initialization step, set $Q_j(0)=0$ for j corresponding to the lowest value of $\sum_{k=1}^{j}\Delta_k F^{(j)}_{k:A}(\infty)$, and repeat the algorithm starting from the new initial condition.

Computation of the asymptotic rates

- The average rate of users in group k is given by

$$R_{\pi_k} = \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}^{\star}\right|\right] - \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I} + \sum_{\ell=k+1}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}^{\star}\right|\right] \tag{28}$$

- In the limit for $N \to \infty$, we can use the asymptotic analytical expression for the mutual information given in Theorem 10, adapted to our case.

- After identifying terms, we obtain

$$
\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left[\log\left|\mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}^{\star}\right|\right] = \\
&\sum_{\ell=k}^{A} \log\left(1 + \rho Q_{\pi_\ell}^{\star} \sum_{m=1}^{B} (\beta_{m,\pi_\ell}^{2}/\lambda_m)\Gamma_m\right) \\
&\quad + \rho \sum_{m=1}^{B} \log\left(1 + \sum_{\ell=k}^{A} (\beta_{m,\pi_\ell}^{2}/\lambda_m) Q_{\pi_\ell}^{\star}\Upsilon_\ell\right) \\
&\quad - \rho \sum_{\ell=k}^{A}\sum_{m=1}^{B} (\beta_{m,\pi_\ell}^{2}/\lambda_m) Q_{\pi_\ell}^{\star}\Gamma_m\Upsilon_\ell
\end{aligned} \tag{29}
$$

- For each $k = 1, \ldots, A$, $\{\Gamma_m : m = 1, \ldots, B\}$ and $\{\Upsilon_\ell : \ell = k, \ldots, A\}$ are the unique solutions to the system of fixed point equations

$$\Gamma_m = \frac{1}{1 + \sum_{\ell=k}^{A} Q^\star_{\pi_\ell}(\beta^2_{m,\pi_\ell}/\lambda_m)\Upsilon_\ell}, \quad m = 1, \ldots, B,$$

$$\Upsilon_\ell = \frac{1}{1 + \rho \sum_{m=1}^{B} Q^\star_{\pi_\ell}(\beta^2_{m,\pi_\ell}/\lambda_m)\Gamma_m}, \quad \ell = k, \ldots, A. \tag{30}$$

- Although (29) is not in a closed form, $\{\Gamma_m\}$ and $\{\Upsilon_\ell\}$ in (30) can be solved by fixed point iterations with $A + B$ variables, that converge very quickly to the solution to any desired degree of numerical accuracy.

Minimization with respect to λ

- In finite dimension and fixed channel matrix, the min-max problem can be solved by an infeasible-start Newton method. See [H. Huh, H. C. Papadopoulos, and G. Caire, Multiuser MISO transmitter optimization for intercell interference mitigation, Transactions on Signal Processing, 2010].

- A direct application to the large system limit requires asymptotic expressions for the *KKT matrix*, involving the second-order derivatives of the Lagrangian function with respect to $\{Q_k\}$ and $\boldsymbol{\lambda}$: not amenable for easily computable asymptotic limits.

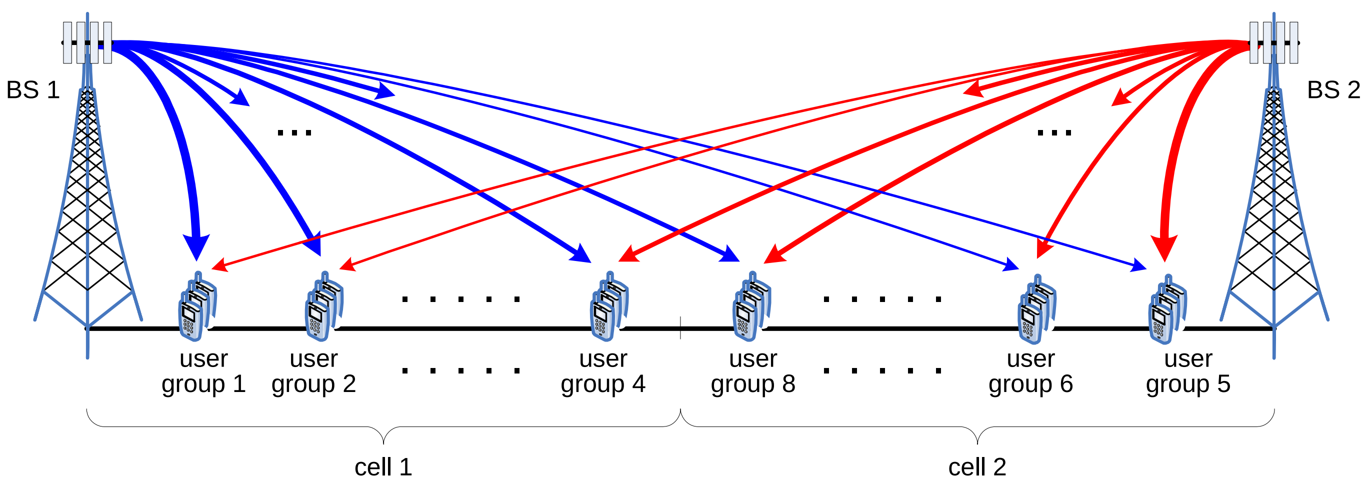
- Idea: let $G_{\mathbf{W}}(\boldsymbol{\lambda})$ denote the solution of (19). This is a convex function of $\boldsymbol{\lambda}$ and the minimizing $\boldsymbol{\lambda}^\star$ must have strictly positive components by Lemma 2.

- Therefore, at the minimizing $\boldsymbol{\lambda}^\star$ we must have $\left.\frac{\partial G_{\mathbf{W}}}{\partial \lambda_m}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^\star} = 0$ for all $m = 1, \cdots, B$ (solution is calculated by descent gradient iteration).

- Obvious upper bound: let $\lambda_m = 1$ for all $m = 1, \ldots, B$. This corresponds to relax the per-BS power constraint to a per-cluster power constraint.

- We can prove that under certain symmetric conditions this bound is tight. In particular, if the channel gain matrix $\boldsymbol{\beta} = \{\beta_{m,k}\}$ can be partitioned into a number of $B \times B$ strongly symmetric blocks, then the minimum is found at $\lambda_m = 1$ for all m.

- Example for $A = 8$, $B = 2$:



- The channel gain matrix takes on the palindrome form

$$\boldsymbol{\beta} = \begin{bmatrix} a & b & b & a & f & e & d & c \\ f & e & d & c & a & b & b & a \end{bmatrix}$$

- This can be decomposed into the 4 strongly symmetric blocks

$$\begin{bmatrix} a & f \\ f & a \end{bmatrix}, \quad \begin{bmatrix} b & e \\ e & b \end{bmatrix}, \quad \begin{bmatrix} b & d \\ d & b \end{bmatrix}, \quad \begin{bmatrix} a & c \\ c & a \end{bmatrix}$$

Computing the fairness rate point by convex optimization

- Recall the general dynamic optimization policy (fairness scheduling).
- Let $W_{k,i}(t)$ denote the virtual queue backlog for user i in group k at time slot t, evolving according to the stochastic difference equation

$$W_{k,i}(t+1) = [W_{k,i}(t) - R_{k,i}(\mathbf{H}(t))]_+ + A_{k,i}(t)$$

- At each time slot t, solve the weighted sum-rate maximization problem

$$\begin{aligned} \text{maximize} \quad & \sum_{k=1}^{A} \sum_{i=1}^{N} W_{k,i}(t) R_{k,i}(\mathbf{H}(t)) \\ \text{subject to} \quad & \mathrm{tr}(\mathbb{E}[\mathbf{x}_m \mathbf{x}_m^{\mathsf{H}}]) \leq P_m \end{aligned} \tag{31}$$

- The virtual arrival processes are given by $A_{k,i}(t) = a^{\star}_{k,i}$, where the vector $\mathbf{a}^{\star}$ is the solution of the maximization problem:

$$\begin{aligned} &\text{maximize} \;\; VU(\mathbf{a}) - \sum_{k=1}^{A} \sum_{i=1}^{N} a_{k,i} W_{k,i}(t) \\ &\text{subject to} \;\; 0 \leq a_{k,i} \leq A_{\max} \end{aligned} \tag{32}$$

for suitable control parameters $V > 0$ and $A_{\max} > 0$.

- The long-time average rates

$$\lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} R_{k,i}(\mathbf{H}(t))$$

are guaranteed to converge to the optimal ergodic rates $R^{\star}_{k,i}$ within a gap factor $O(1/V)$, while the expected backlog of the virtual queues increases as $O(V)$.

- We restrict the network utility function $U(\cdot)$ to be symmetric for users in the same group and Schur-concave.

- Hence, equal average rates for users in the same group is optimal ($R^\star_{k,i} = R^\star_k$).

- The optimum is found on the boundary of the region $\underline{\mathcal{C}}(P_1, \ldots, P_B)$, parameterized by the weights $\{W_1, \ldots, W_A\}$.

- We rewrite (18) using the auxiliary variables $\mathbf{r} = [r_1, \ldots, r_A]$ as:

$$\begin{aligned} &\min_{\boldsymbol{\lambda}} \max_{\mathbf{r},\mathbf{Q},\pi} \quad U(\mathbf{r}) \\ &\text{subject to} \quad r_{\pi_k} \leq \frac{1}{N}\mathbb{E}\left[\log \frac{\left|\mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}\right|}{\left|\mathbf{I} + \sum_{\ell=k+1}^{A} \overline{\mathbf{H}}_{\pi_\ell}\overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell}\right|}\right], \\ &\qquad\qquad\quad \text{tr}(\mathbf{Q}) \leq Q, \quad \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{33}$$

- The Lagrange function for (33) is given by

$$\mathcal{L}(\boldsymbol{\lambda}, \mathbf{r}, \mathbf{Q}, \pi, \mathbf{W}) =$$

$$U(\mathbf{r}) - \sum_{k=1}^{A} W_{\pi_k} \left(r_{\pi_k} - \frac{1}{N}\mathbb{E}\left[\log \frac{\left| \mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell} \overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell} \right|}{\left| \mathbf{I} + \sum_{\ell=k+1}^{A} \overline{\mathbf{H}}_{\pi_\ell} \overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell} \right|} \right] \right)$$

$$= \underbrace{U(\mathbf{r}) - \sum_{k=1}^{A} W_k r_k}_{f_{\mathbf{W}}(\mathbf{r})}$$

$$+ \underbrace{\sum_{k=1}^{A} W_{\pi_k} \frac{1}{N}\mathbb{E}\left[\log \frac{\left| \mathbf{I} + \sum_{\ell=k}^{A} \overline{\mathbf{H}}_{\pi_\ell} \overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell} \right|}{\left| \mathbf{I} + \sum_{\ell=k+1}^{A} \overline{\mathbf{H}}_{\pi_\ell} \overline{\mathbf{H}}_{\pi_\ell}^{\mathsf{H}} Q_{\pi_\ell} \right|} \right]}_{h_{\mathbf{W}}(\boldsymbol{\lambda}, \mathbf{Q}, \pi)} \tag{34}$$

- The Lagrange function can be decomposed into a sum of functions of $\mathbf{r}$ only, denoted by $f_{\mathbf{W}}(\mathbf{r})$, and a function of $\boldsymbol{\lambda}$, $\mathbf{Q}$ and π only, denoted by $h_{\mathbf{W}}(\boldsymbol{\lambda}, \mathbf{Q}, \pi)$.

- The Lagrange dual function for the problem (34) is given by

$$\begin{aligned}\mathcal{G}(\mathbf{W}) &= \min_{\boldsymbol{\lambda}} \max_{\mathbf{r},\mathbf{Q},\pi} \quad \mathcal{L}(\boldsymbol{\lambda}, \mathbf{r}, \mathbf{Q}, \pi, \mathbf{W}) \\ &= \underbrace{\max_{\mathbf{r}} \quad f_{\mathbf{W}}(\mathbf{r})}_{(a)} + \underbrace{\min_{\boldsymbol{\lambda}} \max_{\mathbf{Q},\pi} \quad h_{\mathbf{W}}(\boldsymbol{\lambda}, \mathbf{Q}, \pi)}_{(b)} \end{aligned} \tag{35}$$

  and it is obtained by the decoupled maximization in (a) (with respect to $\mathbf{r}$) and the min-max in (b) (with respect to $\boldsymbol{\lambda}, \mathbf{Q}, \pi$).

- Notice that problems (a) and (b) correspond to the static (deterministic time-invariant) forms of (32) and (31), respectively, after identifying $\mathbf{r}$ with the virtual arrival rates $\mathbf{A}(t)$ and $\mathbf{W}$ with the virtual queue backlogs $\mathbf{W}(t)$.

- Finally, we can solve the dual problem defined as

$$\min_{\mathbf{W} \geq \mathbf{0}} \quad \mathcal{G}(\mathbf{W}) \tag{36}$$

  via inner-outer subgradient iterations:

Inner Problem: For given $\mathbf{W}$, we solve (35) with respect to $\boldsymbol{\lambda}$, $\mathbf{r}$, $\mathbf{Q}$ and π. This can be further decomposed into:

1. Subproblem (a): Since $f_{\mathbf{W}}(\mathbf{r})$ is concave in $\mathbf{r} \geq 0$, the optimal $\mathbf{r}^*$ readily obtained by imposing the KKT conditions.
2. Subproblem (b): Taking the limit of $N \to \infty$, this problem is solved by the iterative algorithm given before for fixed $\boldsymbol{\lambda} > 0$.

Outer Problem: the minimization of $\mathcal{G}(\mathbf{W})$ with respect to $\mathbf{W} \geq \mathbf{0}$ can be obtained by subgradient adaptation.

1. Let $\boldsymbol{\lambda}^*$, π^*, $\mathbf{Q}^*$ and $\mathbf{r}^*(\mathbf{W})$ denote the solution of the inner problem for fixed $\mathbf{W}$.
2. For any $\mathbf{W}'$, we have

$$\begin{aligned}
\mathcal{G}(\mathbf{W}') &= \max_{\mathbf{r}} \; f_{\mathbf{W}'}(\mathbf{r}) + \max_{\mathbf{Q}} \; h_{\mathbf{W}'}(\boldsymbol{\lambda}^*, \mathbf{Q}, \pi^*) \\
&\geq f_{\mathbf{W}'}(\mathbf{r}^*(\mathbf{W})) + h_{\mathbf{W}'}(\boldsymbol{\lambda}^*, \mathbf{Q}^*, \pi^*) \\
&= \mathcal{G}(\mathbf{W}) + \sum_{k=1}^{A} (W_k' - W_k)\,(R_k^*(\mathbf{W}) - r_k^*(\mathbf{W}))
\end{aligned} \tag{37}$$

where $R_k^*(\mathbf{W})$ denotes the k-th group rate resulting from the solution of the inner problem with weights $\mathbf{W}$, which is efficiently calculated by the iterative algorithm in the large-system regime.

3. A subgradient for $\mathcal{G}(\mathbf{W})$ is given by the vector with components $R_k^*(\mathbf{W}) - r_k^*(\mathbf{W})$.
4. The dual variables $\mathbf{W}$ are updated at the n-th outer iteration according to

$$W_k(n+1) = W_k(n) - \mu(n) \left(R_k^*(\mathbf{W}(n)) - r_k^*(\mathbf{W}(n)) \right), \;\; \forall\, k$$

for some step size $\mu(n) > 0$ which can be determined by standard efficient methods.

Proportional fairness scheduling

- The network utility function for PFS is given as

$$U(\mathbf{r}) = \sum_{k=1}^{A} \log(r_k)$$

- In this case, the KKT conditions for the inner subproblem (a) yield the solution

$$r_k^*(\mathbf{W}) = \frac{1}{W_k}, \quad \forall\, k$$

- Observation: the dual variables play the role of the virtual queue backlogs in the dynamic scheduling policy, while the auxiliary variables $\mathbf{r}$ correspond to the virtual arrival rates.

- At the n-th outer iteration these variables are related by

$$W_k(n) = \frac{1}{r_k^*(\mathbf{W}(n))}.$$

- The virtual arrival rates of the dynamic scheduling policy are designed in order to be close to the ergodic rates $\mathbf{R}^\star$ at the optimal fairness point.

- It follows that the usual intuition of PFS, for which the scheduler weights are inversely proportional to the long-term average user rates, is recovered.

Max-min (or "Hard") fairness scheduling

- The network utility function for HFS is given as

$$U(\mathbf{r}) = \min_{k=1,\dots,A} r_k.$$

- This objective function is not strictly concave and differentiable everywhere. Therefore, it is convenient to rewrite subproblem (a) by introducing an auxiliary variable δ, as follows:

$$\begin{aligned} &\max_{\delta,\mathbf{r}\geq 0} \quad \delta - \sum_{k=1}^{A} W_k r_k \\ &\text{subject to} \quad r_k \geq \delta, \;\; \forall\, k \end{aligned} \tag{38}$$

- The solution must satisfy $r_k = \delta$ for all k, leading to

$$\max_{\delta > 0} \quad (1 - \sum_{k=1}^{A} W_k)\delta.$$

- Since the original maximization is bounded while the above may be unbounded, we must have that $\sum_{k=1}^{A} W_k = 1$ and δ must enforce this condition.

- The subgradient iteration for the weights $\mathbf{W}$, using $r_k^*(\mathbf{W}(n)) = \delta^*(\mathbf{W}(n))$, becomes

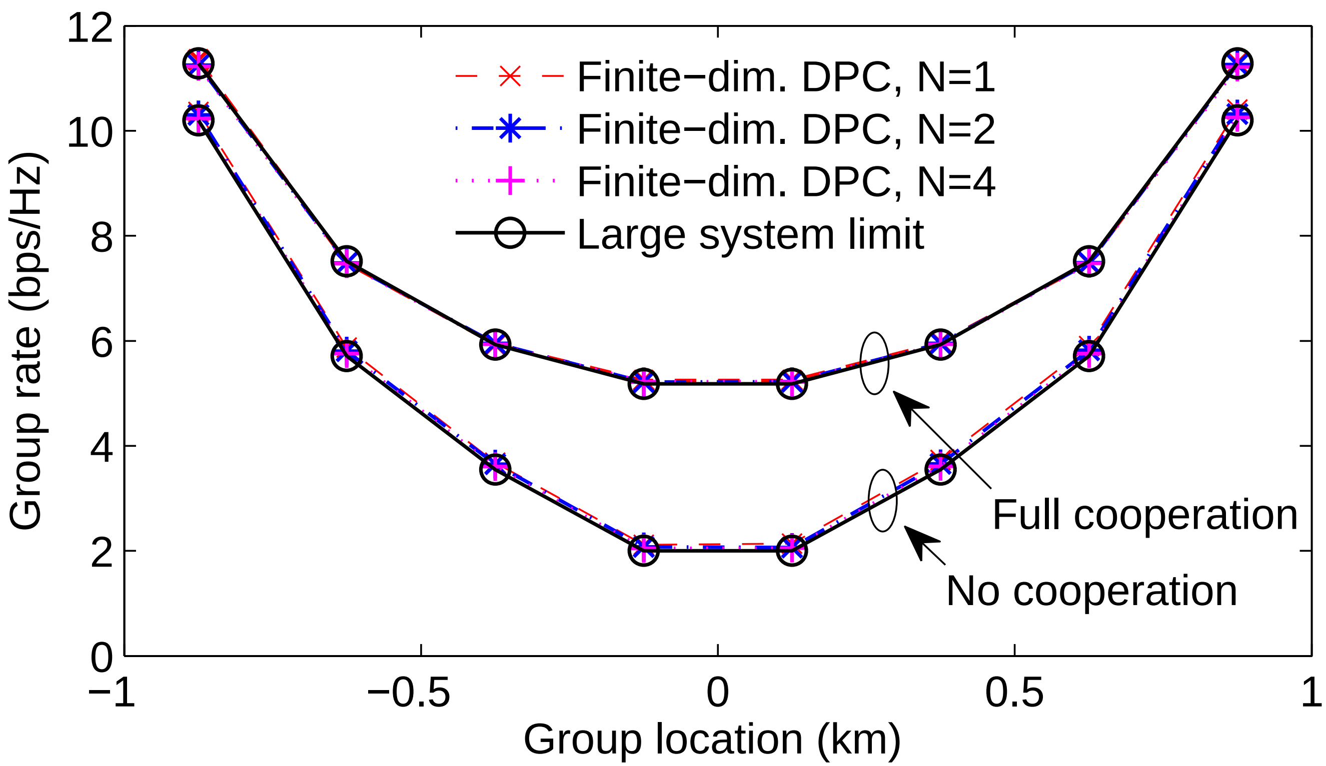
$$W_k(n+1) = W_k(n) - \mu(n) \left(R_k^*(\mathbf{W}(n)) - \delta^*(\mathbf{W}(n)) \right), \quad \forall \, k$$

- Summing up the update equations over $k = 1, \ldots, A$ and using the condition that $\sum_{k=1}^{A} W_k(n) = 1$ for all n, we obtain

$$r_k^*(\mathbf{W}(n)) = \delta^*(\mathbf{W}(n)) = \frac{1}{A} \sum_{j=1}^{A} R_j^*(\mathbf{W}(n)), \quad \forall\, k$$

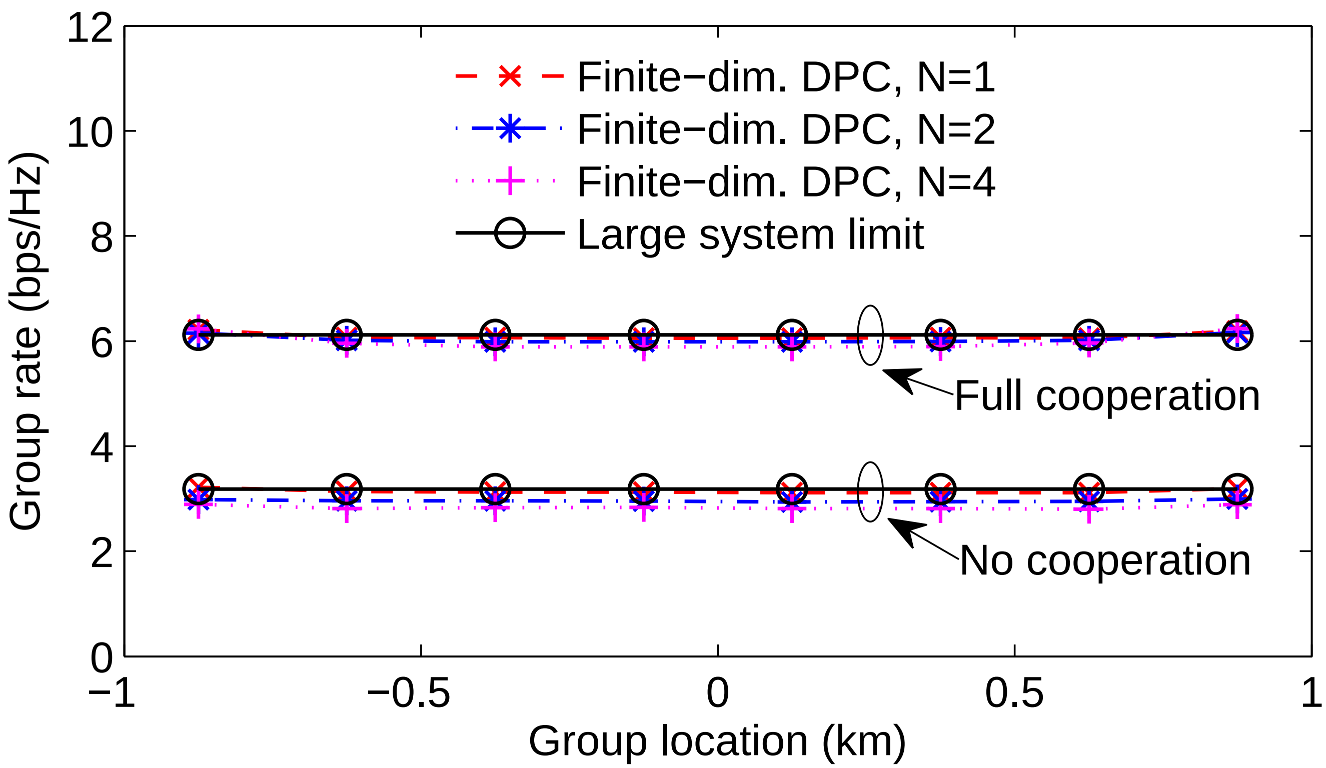
- Intuitively: this creates the same arrival process for all the virtual queues, which naturally yields the same service rate (for stability) and therefore the equal-rate point on the boundary of the ergodic capacity region.

Numerical Results: PFS, two cells



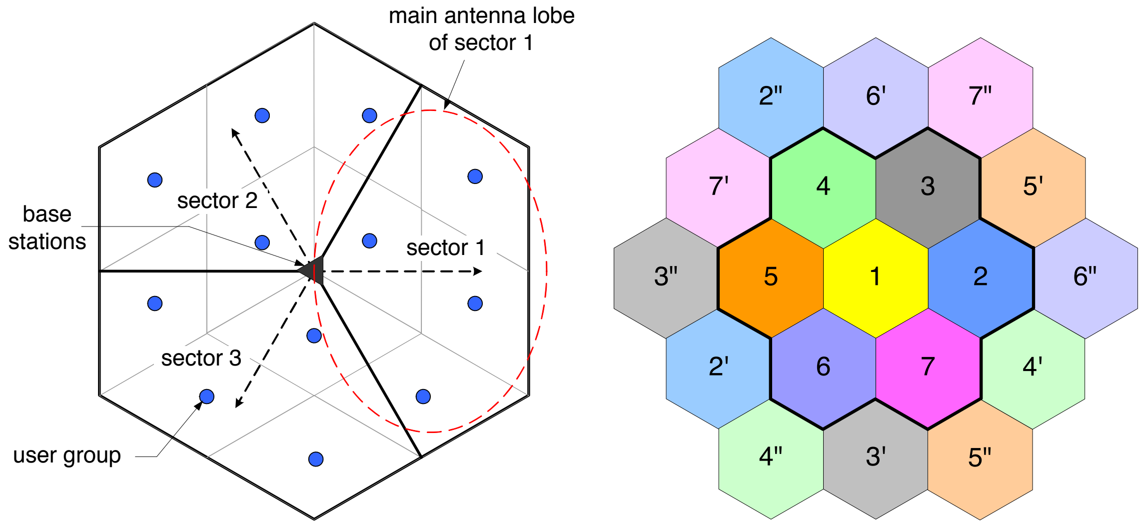
PFS with $\rho = 4$ and $K = 8$ in a 2-cell linear layout.

Numerical Results: HFS, two cells



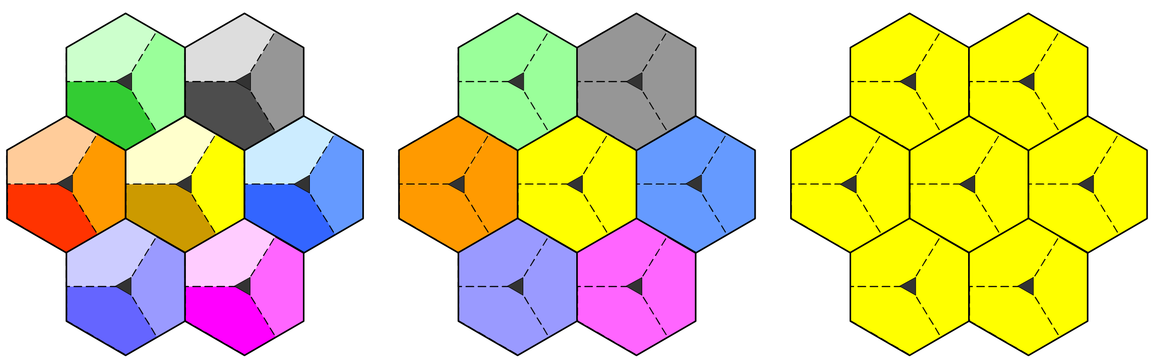
HFS with $\rho = 4$ and $K = 8$ in a 2-cell linear layout.

2-dimensional 7-cell, 21 sectors, 84 user groups model



(a) 3-sectored cell configuration

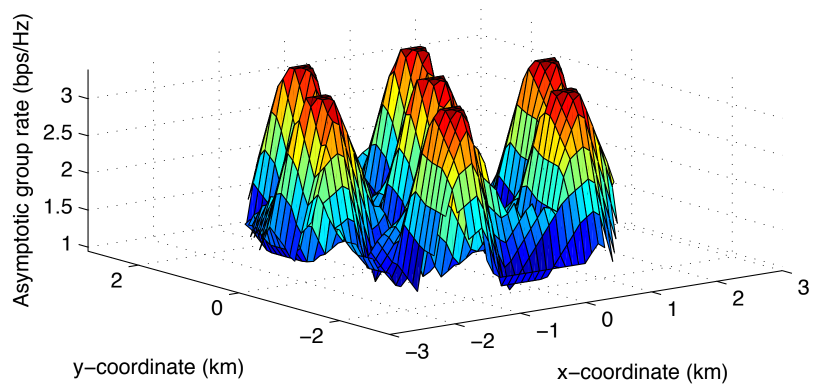
(b) Wrap-around torus topology



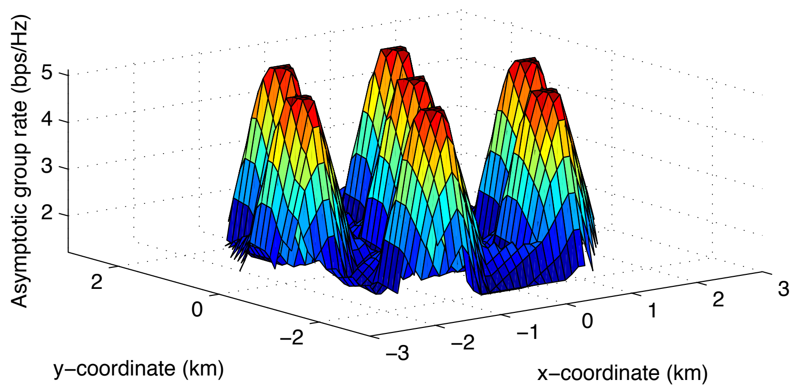
(a) No cooperation

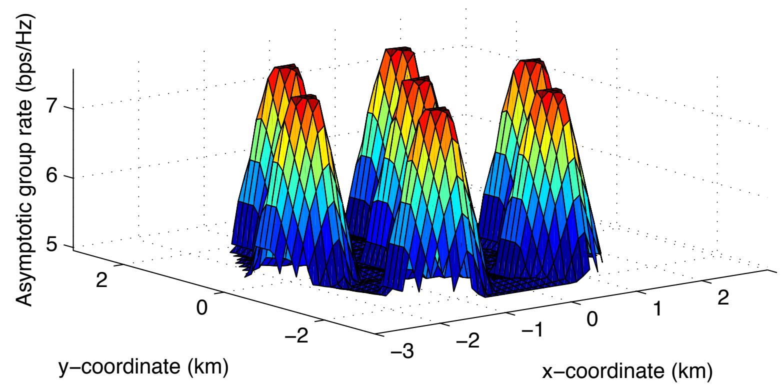
(b) three-sector cooperation within each cell

(c) Full cooperation over 7 cells



no cooperation





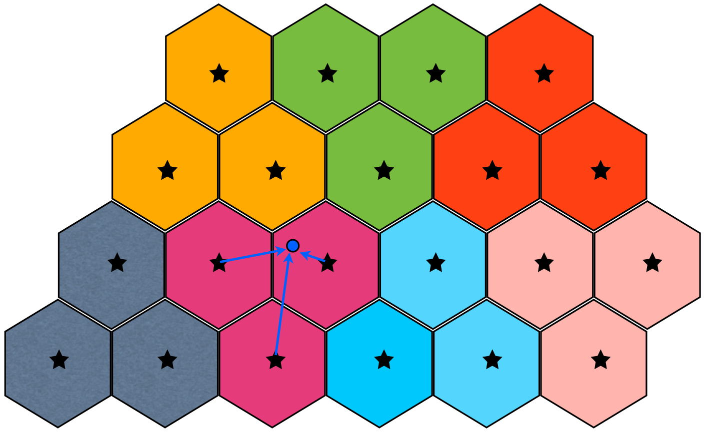
same-cell sector cooperation

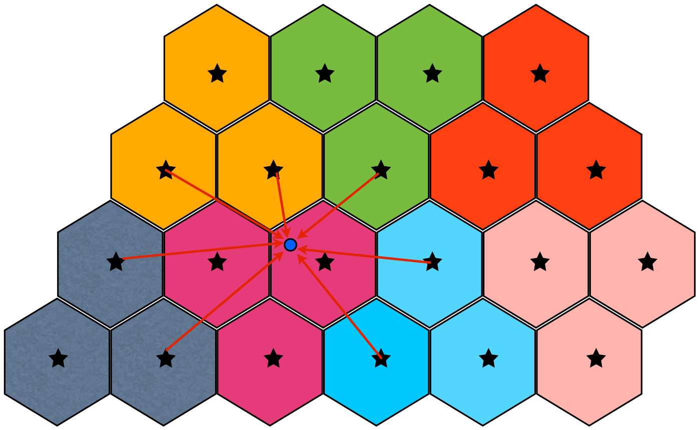
full 7 cell cooperation

End of Lecture 4

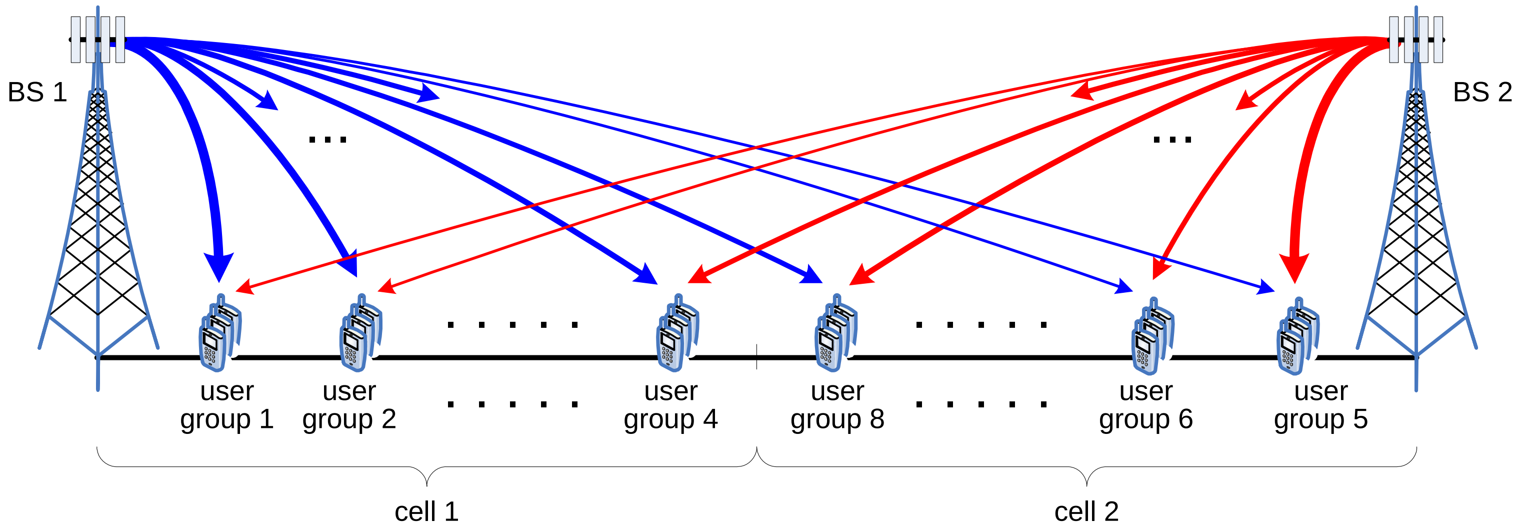
Lecture 5:
Downlink Beamforming

Multi-Cell Network Model





Discretization of the Users Distribution



- We assume that the users are partitioned in co-located groups with N single-antenna terminals each.
- We have A user groups per cluster, and clusters of B cells.
- We have $M = \rho N$ base station antennas per cell.

Multiuser MIMO Downlink

- Channel model for the cooperative cluster MU-MIMO model is exactly as already introduced before.

- Letting A and B denote the number of user locations and BSs in the reference cluster $\mathcal{M}$, and re-normalizing the signal at each user location k by σ_k, such that its ICI plus noise variance is 1, we obtain

$$\mathbf{H} = \left[\begin{array}{ccc} \beta_{1,1}\mathbf{H}_{1,1} & \cdots & \beta_{1,A}\mathbf{H}_{1,A} \\ \vdots & \ddots & \vdots \\ \beta_{B,1}\mathbf{H}_{B,1} & \cdots & \beta_{B,A}\mathbf{H}_{B,A} \end{array} \right],$$

  with $\beta_{m,k} = \frac{\alpha_{m,k}}{\sigma_k}$.

- It follows that from our reference cluster point of view, the relevant downlink channel model is given by

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{v}$$

  with $\mathbf{y} = \mathbb{C}^{AN}$, $\mathbf{x} = \mathbb{C}^{\rho BN}$, and $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

Linear Zero-Forcing Beamforming: A Primer

- A simple alternative to DPC: linear ZF beamforming (ZFBF).
- Assume $\mathbf{H}$ $M \times K$ tall and full column rank. Then, we let

$$\mathbf{x} = \mathbf{V}\mathbf{Q}^{1/2}\mathbf{u}$$

- $\mathbf{u} \in \mathbb{C}^K$ contains the users' information-bearing code symbols (downlink streams), with $\mathbb{E}[\mathbf{u}\mathbf{u}^{\mathsf{H}}] = \mathbf{I}$.
- $\mathbf{V}$ is the precoding matrix with unit-norm columns.
- $\mathbf{Q}$ is a diagonal weighting matrix that contains the power allocated for each downlink stream.

- The ZFBF precoding matrix is obtained as

$$\mathbf{V} = \mathbf{H}^{+}\mathbf{\Lambda}^{-1/2}$$

where

$$\mathbf{H}^{+} = \mathbf{H}(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}$$

is the Moore-Penrose pseudo-inverse of the downlink channel matrix $\mathbf{H}^{\mathsf{H}}$, and where the column-normalizing matrix $\mathbf{\Lambda}$ has elements

$$\Lambda_k = \frac{1}{\left[\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)^{-1}\right]_{k,k}}$$

- The resulting ZFBF-precoded downlink channel is given by

$$\mathbf{y} = \mathbf{\Lambda}^{1/2}\mathbf{Q}^{1/2}\mathbf{u} + \mathbf{v}$$

where inter-cluster multiuser interference is completely removed.

ZFBF in the Large System Regime

- Suppose that $\mathbf{H}$ is a matrix with variance profile, with $N \to \infty$ and $K/N = \beta$.

Theorem 14. *(Corollary of Theorem 9) Defining the effective dimension ratio as*

$$\beta' = \beta \frac{\mathbb{P}\left(\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}] \neq 0\right)}{\mathbb{P}\left(\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{X}] \neq 0\right)},$$

and let $\mathcal{F}(y,\gamma)$ be the function defined by Theorem 9. As γ goes to infinity, we have

$$\lim_{\gamma\to\infty} \mathcal{F}(y,\gamma) = \begin{cases} \Psi_\infty(y) & \textit{if } \beta' < 1 \\ 0 & \textit{if } \beta' \geq 1 \end{cases} \tag{39}$$

where, for $\nu' < 1$, $\Psi_\infty(y)$ is the positive solution to

$$\Psi_\infty(y) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1+\nu\mathbb{E}\left[\left.\frac{v(\mathsf{X},\mathsf{Y})}{\Psi_\infty(\mathsf{Y})}\right|\mathsf{X}\right]}\right] \tag{40}$$

- We wish to characterize the asymptotic expression (for large N) of the ZFBF channel coefficients Λ_k.

- Using the well-known formula for the inverse of a 2×2 block matrix, we can write the (k,k) diagonal element of the matrix $(\mathbf{I} + \gamma \mathbf{H}^{\mathsf{H}} \mathbf{H})^{-1}$ as

$$\left[\left(\mathbf{I} + \gamma \mathbf{H}^{\mathsf{H}} \mathbf{H} \right)^{-1} \right]_{k,k} = \frac{1}{1 + \gamma \mathbf{h}_k^{\mathsf{H}} \left(\mathbf{I} + \gamma \sum_{\ell \neq k} \mathbf{h}_\ell \mathbf{h}_\ell^{\mathsf{H}} \right)^{-1} \mathbf{h}_k}$$

- Furthermore, assuming that $\mathbf{H}$ has full rank, then

$$\begin{aligned}
\left[\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)^{-1}\right]_{k,k} &= \lim_{\gamma\to\infty} \gamma \left[\left(\mathbf{I} + \gamma\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)^{-1}\right]_{k,k} \\
&= \lim_{\gamma\to\infty} \frac{\gamma}{1 + \gamma\mathbf{h}_k^{\mathsf{H}}\left(\mathbf{I} + \gamma\sum_{\ell\neq k}\mathbf{h}_\ell\mathbf{h}_\ell^{\mathsf{H}}\right)^{-1}\mathbf{h}_k} \\
&= \frac{1}{\lim_{\gamma\to\infty}\mathbf{h}_k^{\mathsf{H}}\left(\mathbf{I} + \gamma\sum_{\ell\neq k}\mathbf{h}_\ell\mathbf{h}_\ell^{\mathsf{H}}\right)^{-1}\mathbf{h}_k}
\end{aligned}$$

- Comparing the definition of Λ_k with the above expression and using Theorem 14, we have that

$$\lim_{N\to\infty}\Lambda_k = \lim_{\gamma\to\infty} F(y,\gamma) = \Psi_\infty(y),$$

for $\frac{k-1}{K} \leq y < \frac{k}{N}$, i.e., for $k = \lceil yK \rceil$, with $y \in [0,1)$.

Network Utility Maximization (.... again)

- We follow the already mentioned approach of NUM.
- For a concave non-decreasing network utility function $U(\cdot)$ of the user average rates, we wish to operate the system at the point solution of:

$$\begin{array}{ll} \text{maximize} & U(\mathbf{R}) \\ \text{subject to} & \mathbf{R} \in \mathcal{R}_{\text{zfbf}}(P_1, \ldots, P_B) \end{array}$$

where $\mathcal{R}_{\text{zfbf}}(P_1, \ldots, P_B)$ is the ergodic rate region achievable by ZFBF.

- As before, we start by considering the instantaneous weighted rate sum maximization:

$$\begin{array}{ll} \text{maximize} & \sum_{k=1}^{A} \sum_{i=1}^{N} W_k^{(i)} R_k^{(i)} \\ \text{subject to} & \mathbf{R} \in \mathcal{R}_{\text{zfbf}}(\mathbf{H}) \end{array}$$

- The solution is generally combinatorial, since it requires a search over all user subsets of cardinality less or equal to $\rho B N$.

- Well-known approaches consider the selection of a user subset in some greedy fashion, by adding users to the active user set one by one, till the objective function cannot be improved further.

- User selection involves learning the channel from many users, and selecting a subset: very inefficient in terms of CSIT feedback.

- We shall develop a scheme where users are preselected statistically, and only the pre-selected users feed back their CSIT.

Some simplifying assumptions

- The scheduler picks a fraction μ_k of users in group k by random selection inside the group, independent from slot to slot.

- The ZFBF precoder is obtained by normalizing the columns of the Moore-Penrose pseudo-inverse of the channel matrix, although this choice is not necessarily optimal under the per-BS power constraint.

- Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_A)$ denote the fractions of active users in groups $1, \ldots, A$, respectively. For given $\boldsymbol{\mu}$, the corresponding effective channel matrix is given by

$$\mathbf{H}_{\boldsymbol{\mu}} = \begin{bmatrix} \beta_{1,1}\mathbf{H}_{1,1}(\mu_1) & \cdots & \beta_{1,A}\mathbf{H}_{1,A}(\mu_A) \\ \vdots & & \vdots \\ \beta_{B,1}\mathbf{H}_{B,1}(\mu_1) & \cdots & \beta_{B,A}\mathbf{H}_{B,A}(\mu_A) \end{bmatrix},$$

- The user fractions must satisfy $\mu_k \in [0, 1]$ for each $k = 1, \ldots, A$ and $\mu \triangleq \mu_{1:A} \leq \rho B$ where we introduce the notation $\mu_{1:k} = \sum_{j=1}^{k} \mu_j$.

ZFBF Channel for given user fractions μ

- Operating as before, we have

$$\mathbf{y}_{\boldsymbol{\mu}} = \mathbf{\Lambda}_{\boldsymbol{\mu}}^{1/2} \mathbf{Q}^{1/2} \mathbf{u} + \mathbf{z}_{\boldsymbol{\mu}}$$

where $\Lambda_k^{(i)}(\boldsymbol{\mu})$, the diagonal element of $\mathbf{\Lambda}_{\boldsymbol{\mu}}$ in position $\mu_{1:k-1}N + i$, for $i = 1, \ldots, \mu_k N$, is given by

$$\Lambda_k^{(i)}(\boldsymbol{\mu}) = \frac{1}{\left[\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}} \mathbf{H}_{\boldsymbol{\mu}} \right)^{-1} \right]_k^{(i)}}$$

- The optimization for the parallel channel model is still involved, since the channel coefficients $\Lambda_k^{(i)}(\boldsymbol{\mu})$ depend on the active user fractions $\boldsymbol{\mu}$ in a complicated and non-convex way.

Power allocation for fixed μ

- We divide all channel matrix coefficients by $\sqrt{N}$ and multiply the BS input power constraints P_m by N, thus obtaining an equivalent system where the channel coefficients have variance that scales as $1/N$.

- Let $q_k^{(i)}$ denote the diagonal element in position $\mu_{1:k-1}N + i$ of $\mathbf{Q}$, corresponding to the power allocated to the i-th user of group k.

- Sum-power constraint:

$$\frac{1}{N}\text{tr}(\mathbf{Q}) = \frac{1}{N}\sum_{k=1}^{A}\sum_{i=1}^{\mu_k N} q_k^{(i)} \leq P_{\text{sum}}$$

where $P_{\text{sum}} = \sum_{m=1}^{B} P_m$.

- Per-BS power constraint: let $\mathbf{\Phi}_m$ denote a diagonal matrix with all zeros, but for ρN consecutive ones, corresponding to positions from $(m-1)\rho N + 1$ to $m\rho N$ on the main diagonal.

- The per-BS power constraint is expressed in terms of the partial trace of the transmitted signal covariance matrix as

$$\frac{1}{N}\mathrm{tr}\left(\mathbf{\Phi}_m \mathbf{V}_{\boldsymbol{\mu}} \mathbf{Q} \mathbf{V}_{\boldsymbol{\mu}}^{\mathsf{H}}\right) \leq P_m, \quad m = 1, \ldots, B$$

or, more explicitly,

$$\sum_{k=1}^{A} \sum_{i=1}^{\mu_k N} q_k^{(i)} \theta_{m,k}^{(i)} \leq P_m, \quad m = 1, \ldots, B$$

where we define the coefficients

$$\theta_{m,k}^{(i)}(\boldsymbol{\mu}) = \frac{1}{N} \sum_{\ell=(m-1)\rho N+1}^{m\rho N} \left| \left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)} \right|^2$$

and where $\left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)}$ denotes the element of $\mathbf{V}_{\boldsymbol{\mu}}$ corresponding to the ℓ-th row and the $(\mu_{1:k-1}N+i)$-th column.

- With ZFBF precoding, optimization w.r.t. $\{q_k^{(i)}\}$ for fixed $\boldsymbol{\mu}$ and weights yields

$$\text{maximize} \quad \sum_{k=1}^{A}\sum_{i=1}^{\mu_k N} W_k^{(i)} \log(1+\Lambda_k^{(i)}(\boldsymbol{\mu})q_k^{(i)})$$

subject to either the sum-power or the per-BS power constraint.

- Sum-power $\Longrightarrow$ Waterfilling.

- Per-BS power $\Longrightarrow$ Easy Lagrangian dual/subgradient iteration solution.

- The Lagrangian is given by (dependency on $\boldsymbol{\mu}$ is dropped for notation simplicity)

$$\mathcal{L}(\mathbf{q},\boldsymbol{\lambda}) = \sum_{k=1}^{A}\sum_{i=1}^{\mu_k N} W_k^{(i)} \log(1+\Lambda_k^{(i)}q_k^{(i)}) - \boldsymbol{\lambda}^{\mathsf{T}}\left[\boldsymbol{\Theta}\mathbf{q}-\mathbf{P}\right]$$

where $\boldsymbol{\lambda} \geq 0$ is a vector of dual variables corresponding to the B BS power constraints, $\boldsymbol{\Theta}$ is the $B \times \mu N$ matrix containing the coefficients $\theta_{m,k}^{(i)}$ and $\mathbf{P} = (P_1, \ldots, P_B)^\mathsf{T}$.

- The KKT conditions are given by

$$\frac{\partial \mathcal{L}}{\partial q_k^{(i)}} = W_k^{(i)} \frac{\Lambda_k^{(i)}}{1 + \Lambda_k^{(i)} q_k^{(i)}} - \boldsymbol{\lambda}^\mathsf{T} \boldsymbol{\theta}_k^{(i)} \leq 0$$

where $\boldsymbol{\theta}_k^{(i)}$ is the column of $\boldsymbol{\Theta}$ containing the coefficients $\theta_{m,k}^{(i)}$ for $m = 1, \ldots, B$.

- Solving for $q_k^{(i)}$, we find

$$q_k^{(i)}(\boldsymbol{\lambda}) = \left[\frac{W_k^{(i)}}{\boldsymbol{\lambda}^\mathsf{T} \boldsymbol{\theta}_k^{(i)}} - \frac{1}{\Lambda_k^{(i)}} \right]_+$$

- Replacing this solution into $\mathcal{L}(\mathbf{q}, \boldsymbol{\lambda})$, we solve the dual problem by minimizing $\mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda} \geq 0$. It is immediate to check that for any $\boldsymbol{\lambda}' \geq 0$,

$$\mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}'), \boldsymbol{\lambda}') \geq \mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\lambda}') = (\boldsymbol{\lambda}' - \boldsymbol{\lambda})^{\mathsf{T}}(\mathbf{P} - \boldsymbol{\Theta}\mathbf{q}(\boldsymbol{\lambda})) + \mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$$

- Therefore, $(\mathbf{P} - \boldsymbol{\Theta}\mathbf{q}(\boldsymbol{\lambda}))$ is a subgradient for $\mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\nu})$.

- It follows that the dual problem can be solved by a simple B-dimensional subgradient iteration over the vector of dual variables $\boldsymbol{\lambda}$.

Large system limit for $\Lambda_k(\boldsymbol{\mu})$

- We let $N \to \infty$, when ρ, A, B, and $\boldsymbol{\mu}$ are fixed.

- The large system limit expression for the coefficients $\Lambda_k^{(i)}(\boldsymbol{\mu})$ and given by

Theorem 15. *For all $i = 1, \ldots, \mu_k N$, the following limit holds almost surely:*

$$\lim_{N \to \infty} \Lambda_k^{(i)}(\boldsymbol{\mu}) = \Lambda_k(\boldsymbol{\mu}) = \rho \sum_{m=1}^{B} \beta_{m,k}^2 \eta_m(\boldsymbol{\mu})$$

where $(\eta_1(\boldsymbol{\mu}), \ldots, \eta_B(\boldsymbol{\mu}))$ is the unique solution in $[0,1]^B$ of the fixed point equations

$$\eta_m = 1 - \sum_{q=1}^{A} \mu_q \frac{\eta_m \beta_{m,q}^2}{\rho \sum_{\ell=1}^{B} \eta_\ell \beta_{\ell,q}^2}, \quad m = 1, \ldots, B$$

with respect to the variables $\boldsymbol{\eta} = \{\eta_m\}$. $\square$

Proof:

- From Theorem 14 and the following observations, we have that

$$\lim_{N\to\infty} \Lambda_k^{(i)}(\boldsymbol{\mu}) = \Psi_\infty(y),$$

  evaluated at y such that

$$\frac{\mu_{1:k-1}N + i - 1}{\mu N} \le y < \frac{\mu_{1:k-1}N + i}{\mu N}$$

  after replacing the general matrix $\mathbf{H}$ with $\mathbf{H}_{\boldsymbol{\mu}}$ given by our problem.

- Notice that the dimensions of $\mathbf{H}_{\boldsymbol{\mu}}$ are $\rho BN \times \mu N$ and that $\mu \le \rho B$ by construction.

- The matrix $\mathbf{H}_{\boldsymbol{\mu}}$ is formed by independent blocks $\mathbf{H}_{m,k}(\mu_k)$ of dimension $\rho N \times \mu_k N$, such that each block has i.i.d. $\mathcal{CN}(0, \beta_{m,k}^2/N)$ elements.

- As $N \to \infty$, we have that the aspect ratio is $\nu = \frac{\mu}{\rho B} \le 1$.

- The asymptotic variance profile of $\mathbf{H}_{\boldsymbol{\mu}}$ is given by the piece-wise constant function

$$v(x,y) = \rho B \beta_{m,k}^2 \ \text{ for } \ (x,y) \in \left[\frac{m-1}{B}, \frac{m}{B}\right) \times \left[\frac{\mu_{1:k-1}}{\mu}, \frac{\mu_{1:k}}{\mu}\right)$$

with $m = 1, \ldots, B$ and $k = 1, \ldots, A$.

- Also, we find explicitly

$$\nu' = \nu \, \frac{\sum_{k=1}^{A} \frac{\mu_k}{\mu} 1 \left\{ \frac{1}{B} \sum_{m=1}^{B} \beta_{m,k} \neq 0 \right\}}{\frac{1}{B} \sum_{m=1}^{B} 1 \left\{ \frac{1}{\mu} \sum_{k=1}^{A} \mu_k \beta_{m,k} \neq 0 \right\}}$$

and notice that the case $\nu' < 1$ always holds since, by construction, rank$(\mathbf{H}_{\boldsymbol{\mu}}) = \mu N$.

- As a matter of fact, the piece-wise constant form of $v(x,y)$ yields that $\Lambda_k^{(i)}(\boldsymbol{\mu})$ converges to a limit that depends only on k (the user group) and not on i (the specific user in the group).

- This limit, indicated by $\Lambda_k(\boldsymbol{\mu}) = \Psi_\infty(y)$ for $y \in \left[\frac{\mu_{1:k-1}}{\mu}, \frac{\mu_{1:k}}{\mu}\right)$, is given by

$$\Lambda_k(\boldsymbol{\mu}) = \rho \sum_{m=1}^{B} \frac{\beta_{m,k}^2}{1 + \sum_{q=1}^{A} \mu_q \frac{\beta_{m,q}^2}{\Lambda_q(\boldsymbol{\mu})}}, \qquad k = 1, \ldots, A$$

- In order to obtain the more convenient expression of Theorem 15, we introduce the variables $\eta_m \in [0,1]$, for $m = 1, \ldots, B$, and replace $\Lambda_k(\boldsymbol{\mu}) = \rho \sum_{m=1}^{B} \beta_{m,k}^2 \eta_m$.

- Since η_m takes values in $[0,1]$, we can write $\eta_m = 1/(1+z_m)$ for $z_m \geq 0$, and solving for z_m, we obtain $z_m = \sum_{q=1}^{A} \mu_q \frac{\beta_{m,q}^2}{\Lambda_q(\boldsymbol{\mu})}$.

- Eliminating the variables z_m from the latter equation, we arrive at the desired fixed point equation.

$\blacksquare$

Constant power on each user group

- Since the users in group k have identical $\Lambda_k(\boldsymbol{\mu})$, independent of i, by symmetry we have that $q_k^{(i)} = q_k$ for all active users in group k.

- Using this in the per-BS constraint, we obtain

$$\sum_{k=1}^{A} q_k \theta_{m,k}(\boldsymbol{\mu}) \leq P_m, \quad m = 1, \ldots, B,$$

where

$$\theta_{m,k}(\boldsymbol{\mu}) = \sum_{i=1}^{\mu_k N} \theta_{m,k}^{(i)}(\boldsymbol{\mu}) = \frac{1}{N} \sum_{i=1}^{\mu_k N} \sum_{\ell=1+(m-1)\rho N}^{m\rho N} \left| \left[\mathbf{V} \boldsymbol{\mu} \right]_{\ell,k}^{(i)} \right|^2 .$$

Large system limit for $\theta_{m,k}(\boldsymbol{\mu})$

Theorem 16. *For all m, k, the following limit holds almost surely:*

$$\lim_{N\to\infty} \theta_{m,k}(\boldsymbol{\mu}) = \frac{\mu_k \eta_m^2(\boldsymbol{\mu}) \left(\beta_{m,k}^2 + \xi_{m,k}\right)}{\sum_{\ell=1}^{B} \eta_\ell(\boldsymbol{\mu}) \beta_{\ell,k}^2}$$

where $\boldsymbol{\xi}_m = (\xi_{m,1}, \ldots, \xi_{m,A})^\mathsf{T}$ is the solution to the linear system

$$\left[\mathbf{I} - \rho\mathbf{M}\right] \boldsymbol{\xi}_m = \rho\mathbf{M}\mathbf{b}_m$$

where $\mathbf{M}$ is the $A \times A$ matrix

$$\mathbf{M} = \left[\sum_{\ell=1}^{B} \eta_\ell^2(\boldsymbol{\mu}) \mathbf{b}_\ell \mathbf{b}_\ell^\mathsf{T}\right] \textit{diag}\left(\frac{\mu_1}{\Lambda_1^2(\boldsymbol{\mu})}, \ldots, \frac{\mu_A}{\Lambda_A^2(\boldsymbol{\mu})}\right)$$

and $\mathbf{b}_\ell = (\beta_{\ell,1}^2, \ldots, \beta_{\ell,A}^2)^\mathsf{T}$, and the coefficients $\{\eta_m(\boldsymbol{\mu})\}$ and $\{\Lambda_k(\boldsymbol{\mu})\}$ are provided by Theorem 15. $\square$

Proof:

- We start with the following auxiliary result:

  Let $\mathbf{x}$ be a n-dimensional vector with i.i.d. entries with variance $\frac{1}{n}$. Let $\mathbf{A}$ and $\mathbf{C}$ be $n \times n$ Hermitian symmetric matrices independent on $\mathbf{x}$. Finally let $\mathbf{D}$ be a $n \times n$ diagonal matrix independent on $\mathbf{x}$. Then:

$$\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}} + \mathbf{A})^{-1}\mathbf{C}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}} + \mathbf{A})^{-1}\mathbf{D}\mathbf{x} \rightarrow \frac{\phi(\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{D})}{\left(1 + \phi(\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D})\right)^2}$$

  where $\phi(\cdot) = \lim_{n\rightarrow\infty} \frac{1}{n}\mathrm{tr}(\cdot)$ and the convergence is almost surely.

- Using this lemma, we can proceed with the proof as follows:

$$\begin{aligned}
\theta_{m,k}(\boldsymbol{\mu}) &= \frac{1}{N}\sum_{i=1}^{\mu_k N} \sum_{\ell=1+(m-1)\rho N}^{m\rho N} \left| \left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)} \right|^2 \\
&= \frac{1}{N}\mathrm{tr}\left(\mathbf{\Phi}_m \mathbf{V}_{\boldsymbol{\mu}} \mathbf{\Theta}_k \mathbf{V}_{\boldsymbol{\mu}}^{\mathsf{H}}\right) \\
&= \frac{1}{N}\mathrm{tr}\left(\mathbf{\Phi}_m \mathbf{H}_{\boldsymbol{\mu}} (\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1} \mathbf{\Lambda}_{\boldsymbol{\mu}}^{1/2} \mathbf{\Theta}_k \mathbf{\Lambda}_{\boldsymbol{\mu}}^{1/2} (\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1} \mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}} \mathbf{\Phi}_m\right)
\end{aligned}$$

where $\mathbf{\Phi}_m$ is a diagonal matrix with all zeros, but for ρN consecutive ones, corresponding to positions from $(m-1)\rho N+1$ to $m\rho N$ on the main diagonal, and where $\mathbf{\Theta}_k$ denotes the μN-dimensional diagonal matrix with all zeros, but for $\mu_k N$ consecutive ones, corresponding to positions from $\mu_{1:k-1}N+1$ to $\mu_{1:k}N$ on the main diagonal.

- The submatrix of $\mathbf{\Phi}_m \mathbf{H}_{\boldsymbol{\mu}}$ corresponding to the non-zero rows, i.e., including rows from $(m-1)\rho N+1$ to $m\rho N$, can be written as

$$[\beta_{m,1}\mathbf{H}_{m,1}(\mu_1), \cdots, \beta_{m,A}\mathbf{H}_{m,A}(\mu_A)] = \mathbf{W}_m \mathbf{B}_m$$

where $\mathbf{W}_m$ is a $\rho N \times \mu N$ rectangular matrix with i.i.d. entries, with mean 0 and variance $1/N$, and

$$\mathbf{B}_m = \text{diag}\left(\underbrace{\beta_{m,1}, \ldots, \beta_{m,1}}_{\mu_1 N}, \ldots, \underbrace{\beta_{m,k}, \ldots, \beta_{m,k}}_{\mu_k N}, \ldots, \underbrace{\beta_{m,A}, \ldots, \beta_{m,A}}_{\mu_A N}\right)$$

- Also, we let

$$\begin{aligned}
\mathbf{C}_k &= \boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2} \\
&= \text{diag}\left(\underbrace{0,\ldots,0}_{\mu_{1:k-1}N}, \underbrace{\Lambda_k^{(1)}(\boldsymbol{\mu}),\ldots,\Lambda_k^{(\mu_k N)}(\boldsymbol{\mu})}_{\mu_k N}, \underbrace{0,\ldots,0}_{(\mu-\mu_{1:k})N}\right)
\end{aligned}$$

and notice that $\mathbf{B}_m$ and $\mathbf{C}_k$ have both dimension $\mu N \times \mu N$.

- Letting the ℓ-th row of $\mathbf{W}_m$ be denoted by $\mathbf{w}_{m,\ell}^{\mathsf{H}}$ we can write

$$\begin{aligned}
\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}} &= \sum_{m=1}^{B} \mathbf{B}_m\mathbf{W}_m^{\mathsf{H}}\mathbf{W}_m\mathbf{B}_m \\
&= \mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m + \sum_{j\neq\ell}\mathbf{B}_m\mathbf{w}_{m,j}\mathbf{w}_{m,j}^{\mathsf{H}}\mathbf{B}_m \\
&\quad + \sum_{q\neq m}\mathbf{B}_q\mathbf{W}_q^{\mathsf{H}}\mathbf{W}_q\mathbf{B}_q
\end{aligned}$$

- In order to be able to apply our lemma, we need that the variance of the elements of the i.i.d. vector $\mathbf{w}_{m,\ell}$ (playing the role of $\mathbf{x}$ in the lemma), is equal to the inverse of the vector length. Therefore, dividing by μ, we define

$$\mathbf{A} = \frac{1}{\mu} \sum_{q=1}^{B} \mathbf{B}_q \mathbf{W}_q^{\mathsf{H}} \mathbf{W}_q \mathbf{B}_q$$

and

$$\mathbf{A}_{m,\ell} = \mathbf{A} - \frac{1}{\mu} \mathbf{B}_m \mathbf{w}_{m,\ell} \mathbf{w}_{m,\ell}^{\mathsf{H}} \mathbf{B}_m$$

- Eventually, collecting all these expressions, we arrive at

$$
\begin{aligned}
\theta_{m,k}(\boldsymbol{\mu}) &= \frac{1}{N\mu}\mathrm{tr}\left(\frac{1}{\sqrt{\mu}}\mathbf{W}_m\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\mathbf{W}_m^{\mathsf{H}}\frac{1}{\sqrt{\mu}}\right) \\
&= \frac{1}{N\mu}\sum_{\ell=1}^{\rho N}\frac{1}{\sqrt{\mu}}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m\left(\frac{1}{\mu}\mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m+\mathbf{A}_{m,\ell}\right)^{-1}\mathbf{C}_k \\
&\qquad \cdot\left(\frac{1}{\mu}\mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m+\mathbf{A}_{m,\ell}\right)^{-1}\mathbf{B}_m\mathbf{w}_{m,\ell}\frac{1}{\sqrt{\mu}} \\
&\to \frac{\rho}{\mu}\frac{\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\right)}{\left(1+\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right)\right)^2}
\end{aligned} \tag{41}
$$

- At this point, our goal is to evaluate the two limit normalized traces.

- We start by the term in the denominator:

$$
\begin{aligned}
\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right) &= \lim_{N\to\infty}\frac{1}{\mu N}\mathrm{tr}\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right)\\
&= \lim_{N\to\infty}\frac{1}{\mu N}\mathrm{tr}\left(\left(\frac{1}{\mu}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\mathbf{B}_m^2\right)\\
&= \lim_{N\to\infty}\frac{1}{N}\mathrm{tr}\left(\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\mathbf{B}_m^2\right)\\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{A}\sum_{i=1}^{\mu_k N}\frac{\beta_{m,k}^2}{\Lambda_k^{(i)}(\boldsymbol{\mu})}\\
&= \sum_{k=1}^{A}\frac{\mu_k\beta_{m,k}^2}{\Lambda_k(\boldsymbol{\mu})}
\end{aligned}
$$

where we used the fact that, by definition,

$$\left[\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\right]_{k}^{(i)} = \frac{1}{\Lambda_k^{(i)}(\boldsymbol{\mu})}$$

for the diagonal elements of $\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}$ in position $\mu_{1:k-1}N + i$ for $i = 1, \ldots \mu_k N$, and the convergence result of Theorem 15.

- Also, comparing with the expression of z_m in the proof of Theorem 15 we have that

$$z_m = \sum_{k=1}^{A} \frac{\mu_k \beta_{m,k}^2}{\Lambda_k(\boldsymbol{\mu})}$$

- Since $\eta_m(\boldsymbol{\mu}) = 1/(1+z_m)$, where $\{\eta_m(\boldsymbol{\mu}) : m = 1, \ldots, B\}$ are the auxiliary variables defined in Theorem 15, we have that the denominator is given by

$$\left(1 + \phi\left(\mathbf{B}_m \mathbf{A}^{-1} \mathbf{B}_m\right)\right)^2 = \eta_m^{-2}(\boldsymbol{\mu})$$

- Next, we consider the numerator. For this purpose, let ζ be a dummy non-negative real variable and consider the identity:

$$\frac{-d}{d\zeta}\text{tr}\left(\left(\zeta\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right)=\text{tr}\left(\mathbf{B}_m(\zeta\mathbf{B}_m^2+\mathbf{A})^{-1}\mathbf{C}_k(\zeta\mathbf{B}_m^2+\mathbf{A})^{-1}\mathbf{B}_m\right)$$

- By almost-sure continuity of the trace with respect to $\zeta \geq 0$, it follows that the desired expression for the numerator can be calculated as

$$\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\right)=\lim_{\zeta\downarrow 0}\frac{-d}{d\zeta}\phi\left(\left(\zeta\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right)$$

- In order to compute the asymptotic normalized trace above, we use Theorem 8, that we recall here for convenience:

  Let $\mathbf{H}$ be $N_r \times N_c$ of the variance profile type. For any $a, b \in [0,1]$ with $a < b$,

$$\frac{1}{N_r} \sum_{i=\lfloor aN_r \rfloor}^{\lfloor bN_r \rfloor} \left[\left(s\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{I} \right)^{-1} \right]_{i,i} \rightarrow \int_a^b \Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(x,s)\, dx$$

  where $N_c/N_r \rightarrow \nu$ and where $\Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(x,s)$ and $\Upsilon_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(y,s)$ are functions defined implicitly by the fixed-point equation

$$\begin{aligned}
\Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(x,s) &= \frac{1}{1 + \nu s \mathbb{E}\left[v(x, \mathsf{Y}), \Upsilon_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\mathsf{Y}, s) \right]} \\
\Upsilon_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(y,s) &= \frac{1}{1 + s \mathbb{E}\left[v(\mathsf{X}, y), \Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\mathsf{X}, s) \right]}
\end{aligned}$$

  for $(x,y) \in [0,1] \times [0,1]$, where X and Y are i.i.d. uniform-$[0,1]$ RVs and where the variance profile function is $v(x,y)$.

- In order to use Theorem 8 in our case, we write

$$\begin{aligned}\mathrm{tr}\left(\left(\zeta\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) &= \mathrm{tr}\left(\left(\zeta\mathbf{I}+\mathbf{B}_m^{-1}\mathbf{A}\mathbf{B}_m^{-1}\right)^{-1}\mathbf{B}_m^{-1}\mathbf{C}_k\mathbf{B}_m^{-1}\right)\\ &= \frac{1}{\zeta}\mathrm{tr}\left(\left(\mathbf{I}+\frac{1}{\zeta}\mathbf{B}_m^{-1}\mathbf{A}\mathbf{B}_m^{-1}\right)^{-1}\mathbf{B}_m^{-1}\mathbf{C}_k\mathbf{B}_m^{-1}\right)\end{aligned}$$

- Noticing that, by definition, $\mathbf{A} = \frac{1}{\mu}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}$, we can identify the matrix $\frac{1}{\sqrt{\mu}}\mathbf{B}_m^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}$ with the matrix $\mathbf{H}$ of the Lemma.

- In this case, $N_r = \mu N$ and $N_c = \rho B N$. Using $\{\mathbf{B}_m\}$ and $\{\mathbf{W}_m\}$ defined before, we can write the block-matrix form

$$\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}} = \left[\mathbf{B}_1\mathbf{W}_1^{\mathsf{H}}, \mathbf{B}_2\mathbf{W}_2^{\mathsf{H}}, \ldots, \mathbf{B}_B\mathbf{W}_B^{\mathsf{H}}\right]$$

so that

$$\mathbf{B}_m^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}} = \left[\mathbf{B}_m^{-1}\mathbf{B}_1\mathbf{W}_1^{\mathsf{H}}, \mathbf{B}_m^{-1}\mathbf{B}_2\mathbf{W}_2^{\mathsf{H}}, \ldots, \mathbf{B}_m^{-1}\mathbf{B}_B\mathbf{W}_B^{\mathsf{H}}\right]$$

- It follows that the variance profile function of $\frac{1}{\sqrt{\mu}}\mathbf{B}_m^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}$ is given by

$$v_m(x,y) = \frac{\beta_{\ell,k}^2}{\beta_{m,k}^2}, \quad \text{for} \quad (x,y) \in \left[\frac{\mu_{1:k-1}}{\mu}, \frac{\mu_{1:k}}{\mu}\right) \times \left[\frac{\ell-1}{B}, \frac{\ell}{B}\right)$$

- Letting $1/\zeta = s$ and using Theorem 8, we find

$$\frac{1}{\mu N} \sum_{i=\mu_{1:k-1}N+1}^{\mu_{1:k}N} \left[\left(\mathbf{I} + s\mathbf{B}_m^{-1}\mathbf{A}\mathbf{B}_m^{-1}\right)^{-1}\right]_{i,i} \to \int_{\mu_{1:k-1}/\mu}^{\mu_{1:k}/\mu} \Gamma_m(x,s)\, dx$$

where $\Gamma_m(x,s)$ and $\Upsilon_m(y,s)$ are defined by

$$\begin{aligned}
\Gamma_m(x,s) &= \frac{1}{1 + \frac{\rho B s}{\mu}\mathbb{E}\left[v_m(x,\mathsf{Y}), \Upsilon_m(\mathsf{Y},s)\right]} \\
\Upsilon_m(y,s) &= \frac{1}{1 + s\mathbb{E}\left[v_m(\mathsf{X},y), \Gamma_m(\mathsf{X},s)\right]}
\end{aligned}$$

- Noticing that $v_m(x,y)$ is piecewise constant, we have that also the functions $\Gamma_m(x,s)$ and $\Upsilon_m(y,s)$ are piecewise constant.

- With some abuse of notation, we denote the values of these functions as $\{\Gamma_{m,q}(s), q=1,\dots,A\}$ and $\{\Upsilon_{m,\ell}(s), \ell=1,\dots,B\}$, respectively, we find that the fixed point equation can be re-written directly in terms of these values as

$$\begin{aligned}
\Gamma_{m,q}(s) &= \frac{1}{1+\frac{s}{\mu}\sum_{\ell=1}^{B}\frac{\rho\beta_{\ell,q}^2}{\beta_{m,q}^2}\Upsilon_{m,\ell}(s)}, \quad \text{for } q=1,\dots,A \\
\Upsilon_{m,\ell}(s) &= \frac{1}{1+\frac{s}{\mu}\sum_{q=1}^{A}\frac{\mu_q\beta_{\ell,q}^2}{\beta_{m,q}^2}\Gamma_{m,q}(s)} \quad \text{for } \ell=1,\dots,B
\end{aligned}$$

- Finally, using these results and noticing that the non-zero diagonal elements of $\mathbf{B}_m^{-1}\mathbf{C}_k\mathbf{B}_m^{-1}$ converge to the constant $\Lambda_k(\boldsymbol{\mu})\beta_{m,k}^{-2}$, we arrive at:

$$\phi\left(\left(\zeta\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) = \frac{\mu_k}{\zeta\mu}\Gamma_{m,k}(1/\zeta)\Lambda_k(\boldsymbol{\mu})\beta_{m,k}^{-2}$$

- It turns out that it is convenient to define the new variables

$$S_{m,q}(\zeta) = \frac{1}{\zeta \beta_{m,q}^2} \Gamma_{m,q}(1/\zeta), \quad \text{and} \quad G_{m,\ell}(\zeta) = \Upsilon_{m,\ell}(1/\zeta)$$

- Therefore, we can rewrite

$$\begin{aligned}
S_{m,q}(\zeta) &= \frac{1}{\zeta \beta_{m,q}^2 + \frac{\rho}{\mu} \sum_{\ell=1}^{B} \beta_{\ell,q}^2 G_{m,\ell}(\zeta)}, \quad \text{for } q = 1, \ldots, A \\
G_{m,\ell}(\zeta) &= \frac{1}{1 + \frac{1}{\mu} \sum_{q=1}^{A} \mu_q \beta_{\ell,q}^2 S_{m,q}(\zeta)}, \quad \text{for } \ell = 1, \ldots, B \\
\phi\left(\left(\zeta \mathbf{B}_m^2 + \mathbf{A}\right)^{-1} \mathbf{C}_k\right) &= \frac{\mu_k}{\mu} \Lambda_k(\boldsymbol{\mu}) S_{m,k}(\zeta)
\end{aligned}$$

- Taking the derivative, we obtain the desired numerator in the form

$$\begin{aligned}
\lim_{\zeta \downarrow 0} \frac{-d}{d\zeta} \phi \left(\left(\zeta \mathbf{B}_m^2 + \mathbf{A} \right)^{-1} \mathbf{C}_k \right) &= \frac{\mu_k}{\mu} \Lambda_k(\boldsymbol{\mu}) \lim_{\zeta \downarrow 0} \frac{-d}{d\zeta} S_{m,k}(\zeta) \\
&= \frac{\mu_k}{\mu} \Lambda_k(\boldsymbol{\mu}) \dot{S}_{m,k}(0)
\end{aligned}$$

where we define $\dot{S}_{m,k}(0) = \frac{-d}{d\zeta} S_{m,k}(\zeta)|_{\zeta=0}$ and, for later use, $\dot{G}_{m,\ell}(0) = \frac{d}{d\zeta} G_{m,\ell}(\zeta)|_{\zeta=0}$.

- Next, we wish to find a fixed-point equation that yields directly $\dot{S}_{m,k}(0)$.

- By continuity, we can replace directly $\zeta = 0$ into the fixed point equations after taking the derivatives. By doing so, we obtain:

$$\dot{S}_{m,q}(0) = \frac{\beta_{m,q}^2 + \frac{\rho}{\mu}\sum_{\ell=1}^{B} \beta_{\ell,q}^2 \dot{G}_{m,\ell}(0)}{\left(\frac{\rho}{\mu}\sum_{\ell=1}^{B} \beta_{\ell,q}^2 G_{m,\ell}(0)\right)^2}, \quad \text{for} \quad q = 1, \ldots, A$$

$$\dot{G}_{m,\ell}(0) = \frac{\frac{1}{\mu}\sum_{q=1}^{A} \mu_q \beta_{\ell,q}^2 \dot{S}_{m,q}(0)}{\left(1 + \frac{1}{\mu}\sum_{q=1}^{A} \mu_q \beta_{\ell,q}^2 S_{m,q}(0)\right)^2}, \quad \text{for} \quad \ell = 1, \ldots, B$$

- Also, the equations for $S_{m,q}(0)$ and $G_{m,\ell}(0)$, obtained by replacing $\zeta = 0$, read:

$$\begin{aligned}
S_{m,q}(0) &= \frac{1}{\frac{\rho}{\mu}\sum_{\ell=1}^{B}\beta_{\ell,q}^2 G_{m,\ell}(0)}, \quad \text{for } \; q = 1,\ldots,A \\
G_{m,\ell}(0) &= \frac{1}{1+\frac{1}{\mu}\sum_{q=1}^{A}\mu_q\beta_{\ell,q}^2 S_{m,q}(0)}, \quad \text{for } \; \ell = 1,\ldots,B
\end{aligned}$$

- Using these equations, we obtain, for all $\ell = 1,\ldots,B$,

$$G_{m,\ell}(0) = \frac{1}{1+\sum_{q'=1}^{A}\frac{\mu_{q'}\beta_{\ell,q'}^2}{\rho\sum_{\ell'=1}^{B}\beta_{\ell',q'}^2 G_{m,\ell'}(0)}}.$$

- By multiplying both sides by $\rho\beta_{\ell,q}^2$ and summing over ℓ, we find

$$U_{m,q} = \rho \sum_{\ell=1}^{B} \frac{\beta_{\ell,q}^2}{1 + \sum_{q'=1}^{A} \frac{\mu_{q'}\beta_{\ell,q'}^2}{U_{m,q'}}},$$

where we define $U_{m,q} = \rho \sum_{\ell=1}^{B} \beta_{\ell,q}^2 G_{m,\ell}(0)$.

- Comparing the fixed point equation with the expression for $\Lambda_k(\boldsymbol{\mu})$ from Theorem 15, we discover that $U_{m,q} = \Lambda_q(\boldsymbol{\mu})$, independent of m. Using this result we obtain

$$S_{m,q}(0) = \frac{\mu}{\Lambda_q(\boldsymbol{\mu})}$$

- Using the definition of $U_{m,q}$, we arrive at

$$\dot{S}_{m,q}(0) = \frac{\mu^2\beta_{m,q}^2 + \mu\dot{U}_{m,q}}{\Lambda_q^2(\boldsymbol{\mu})},$$

where, with some abuse of notation, we define $\dot{U}_{m,q} = \rho \sum_{\ell=1}^{B} \beta_{\ell,q}^2 \dot{G}_{m,\ell}(0)$.

- Multiplying both sides by $\rho\beta^2_{\ell,q}$, using the expression of $\dot{S}_{m,q}(0)$ and summing over ℓ, we obtain

$$
\begin{aligned}
\dot{U}_{m,q} &= \rho \sum_{\ell=1}^{B} \beta^2_{\ell,q} \frac{\frac{1}{\mu}\sum_{q'=1}^{A} \mu_{q'}\beta^2_{\ell,q'}\dot{S}_{m,q'}(0)}{\left(1+\frac{1}{\mu}\sum_{q'=1}^{A} \mu_{q'}\beta^2_{\ell,q'}S_{m,q'}(0)\right)^2} \\
&= \rho \sum_{\ell=1}^{B} \beta^2_{\ell,q} \frac{\frac{1}{\mu}\sum_{q'=1}^{A} \mu_{q'}\beta^2_{\ell,q'}\frac{\mu^2\beta^2_{m,q'}+\mu\dot{U}_{m,q'}}{\Lambda^2_{q'}(\boldsymbol{\mu})}}{\left(1+\sum_{q'=1}^{A} \frac{\mu_{q'}\beta^2_{\ell,q'}}{\Lambda_{q'}(\boldsymbol{\mu})}\right)^2} \\
&= \rho\mu \sum_{q'=1}^{A} \left[\sum_{\ell=1}^{B} \eta^2_\ell(\boldsymbol{\mu})\beta^2_{\ell,q}\beta^2_{\ell,q'}\right] \frac{\mu_{q'}}{\Lambda^2_{q'}(\boldsymbol{\mu})} \left(\beta^2_{m,q'} + \frac{1}{\mu}\dot{U}_{m,q'}(\boldsymbol{\mu})\right)
\end{aligned}
$$

- Somehow surprisingly, we notice that the last line is a system of A linear equations in the A unknown $\{\dot{U}_{m,q} : q = 1, \ldots, A\}$. Therefore, this can be solved explicitly (although not in closed form in general).

- In particular, we define the $A \times A$ matrix

$$\mathbf{M} = \left[\sum_{\ell=1}^{B} \eta_\ell^2(\boldsymbol{\mu})\mathbf{b}_\ell \mathbf{b}_\ell^\mathsf{T}\right] \text{diag}\left(\frac{\mu_1}{\Lambda_1^2(\boldsymbol{\mu})}, \ldots, \frac{\mu_A}{\Lambda_A^2(\boldsymbol{\mu})}\right)$$

where $\mathbf{b}_\ell = (\beta_{\ell,1}^2, \ldots, \beta_{\ell,A}^2)^\mathsf{T}$, and the vector of unknowns $\dot{\mathbf{U}}_m$, then, we the linear system is given by

$$[\mathbf{I} - \rho\mathbf{M}]\, \dot{\mathbf{U}}_m = \rho\mu\mathbf{M}\mathbf{b}_m$$

- Solving the system we obtain the sought numerator in the form

$$\frac{\mu_k}{\mu}\Lambda_k(\boldsymbol{\mu})\dot{S}_{m,k}(0) = \mu_k \frac{\mu\beta_{m,k}^2 + \dot{U}_{m,k}}{\Lambda_k(\boldsymbol{\mu})}.$$

- Finally, we obtain our final result:

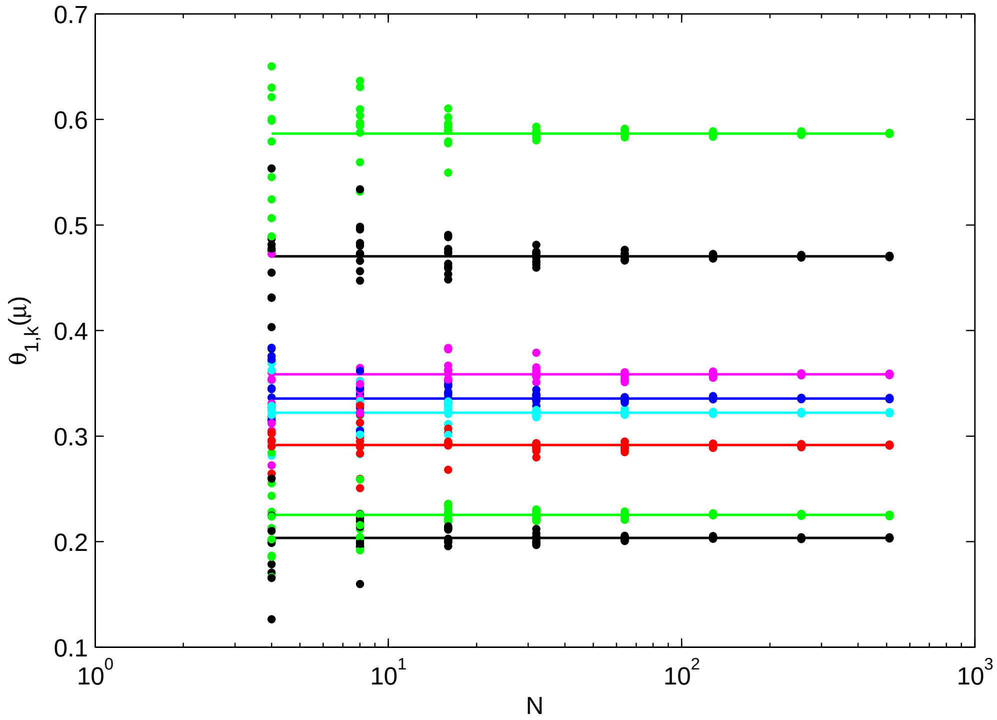
$$
\begin{aligned}
\theta_{m,k}(\boldsymbol{\mu}) &= \frac{\rho}{\mu}\frac{\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\right)}{\left(1+\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right)\right)^2} \\
&= \frac{\rho}{\mu}\frac{\mu_k(\mu\beta_{m,k}^2+\dot{U}_{m,k})}{\Lambda_k(\boldsymbol{\mu})}\eta_m^2(\boldsymbol{\mu}) \\
&= \frac{\mu_k\eta_m^2(\boldsymbol{\mu})\left(\beta_{m,k}^2+\dot{U}_{m,k}/\mu\right)}{\sum_{\ell=1}^{B}\eta_\ell(\boldsymbol{\mu})\beta_{\ell,k}^2}
\end{aligned}
$$

where in the last line we used Theorem 15.

- Comparing the expression of Theorem 16 with the above we see that the two expression coincide by letting $\xi_m = \dot{U}_m/\mu$.

■

See it to believe it!



Finite dimensional samples of $\theta_{m,k}(\boldsymbol{\mu})$ with $N = [4\ 8\ 16\ 32\ 64\ 128\ 256]$ (dots) and asymptotic values in the large system limit (lines) for $m = 1$, $k = 1, \ldots, 8$, and $\boldsymbol{\mu} = [0.5\ 0.5\ 0.75\ 1\ 1\ 0.75\ 0.5\ 0.5]$.

System symmetry

- For symmetric systems (same definition as before), choosing the same user fraction in each symmetric equivalence class of groups, yields

$$\theta_{m,k}(\boldsymbol{\mu}) = \frac{\mu_k}{B}$$

independently of m.

- As a consequence, if all the BSs in the cluster have the equal power constraint, i.e., $P_1 = \ldots = P_B = P$, then for a symmetric system the per-BS power constraint coincides with the sum power constraint with $P_{\mathrm{sum}} = BP$.

- This conclusion is analogous to what we have already found for the case of DPC downlink precoding.

Weighted sum-rate maximization

- Sum power constraint: using our large-system results, we arrive at:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{k=1}^{A} W_k \mu_k \log\left(1 + \rho\left(\sum_{m=1}^{B} \beta_{m,k}^2 \eta_m\right) q_k\right) \\
\text{subject to} \quad & \sum_{k=1}^{A} \mu_k q_k \le P_{\text{sum}}, \quad \sum_{k=1}^{A} \mu_k \le \rho B, \\
& \eta_m = 1 - \sum_{k=1}^{A} \mu_k \frac{\eta_m \beta_{m,k}^2}{\rho \sum_{\ell=1}^{B} \eta_\ell \beta_{\ell,k}^2}, \quad m = 1, \dots, B \\
& 0 \le \eta_m \le 1, \quad m = 1, \dots, B \\
& q_k \ge 0, \;\; 0 \le \mu_k \le 1, \quad k = 1, \dots, A
\end{aligned}
$$

- For per-BS constraint, the power constraint is replaced by

$$\sum_{k=1}^{A} q_k \theta_{m,k}(\boldsymbol{\mu}) \leq P_m, \quad m = 1, \ldots, B,$$

- These problems are generally non-convex in $\mathbf{q}$, $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$.

- For fixed $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$, they are convex in $\mathbf{q}$.

- For fixed $\boldsymbol{\eta}$ and $\mathbf{q}$, we have a linear program with respect to $\boldsymbol{\mu}$.

- Finally, for fixed $\boldsymbol{\mu}$ and $\mathbf{q}$ the problem is degenerate with respect to $\boldsymbol{\eta}$ because of the equality constraint that corresponds to the fixed-point equation of Theorem 15.

- We proposed a greedy search over the user fractions $\boldsymbol{\mu}$ that yields near-optimal results, inspired by the greedy user selection in finite dimension.

Network utility function maximization

- In general, the solution of the weighted sum-rate maximization problem for the case $A > \rho B$ (more users than antennas) yields an unbalanced distribution of instantaneous rates, where some user classes are not served at all (we have $\mu_k = 0$ for some k).
- This is true even in the large system limit, since the ZFBF precoder is limited by the rank of the channel matrix.
- This shows that, for a general strictly concave network utility function $U(\cdot)$, the ergodic rate region $\overline{\mathcal{R}}$ requires time-sharing even in the asymptotic large-system case.
- Finding the solution of the optimal network utility maximization is therefore extremely hard.
- Nevertheless, this solution can be computed to any level of accuracy by using a method inspired by the dynamic scheduling policy (stochastic optimization) approach.

Virtual queues and weight update

- For each user group $k = 1, \ldots, A$, define a *virtual queue* that evolves according to
$$W_k(t+1) = [W_k(t) - r_k(t)]_+ + a_k(t)$$
where $r_k(t)$ denotes the virtual service rate and $a_k(t)$ the virtual arrival process.

- The queues are initialized by $W_k(0) = r_k(0) = 0$. Then, at each iteration $t = 1, 2, \ldots$, the virtual arrival processes is given by $a_k(t) = a_k^\star$ where $\mathbf{a}^\star$ is the solution of

$$\begin{aligned} \text{maximize} \quad & VU(\mathbf{a}) - \sum_{k=1}^{A} W_k(t) a_k \\ \text{subject to} \quad & 0 \leq a_k \leq A_{\max}, \quad \forall\, k \end{aligned} \tag{42}$$

and where $V, A_{\max} > 0$ are some suitably chosen constants, that determine the convergence properties of the iterative algorithm.

- The service rates are given by

$$r_k(t) = \mu_k(t) \log \left(1 + \rho \left(\sum_{m=1}^{B} \beta_{m,k}^2 \eta_m(t) \right) q_k(t) \right)$$

where $(\boldsymbol{\mu}(t), \mathbf{q}(t), \boldsymbol{\eta}(t))$ is the solution of the joint power and user fraction optimization problem for weights $W_k = W_k(t)$.

- Let $\mathbf{r}(t)$ denote the vector of service rates generated by the above iterative algorithm. Then, we can show that

$$\liminf_{t \to \infty} U \left(\frac{1}{t} \sum_{\tau=0}^{t-1} \mathbf{r}(\tau) \right) \geq U(\overline{\mathbf{R}}^{\star}) - \frac{\mathcal{K}}{V}$$

where $\overline{\mathbf{R}}^{\star}$ is the optimal ergodic rate point and $\mathcal{K}$ is a constant.

Examples

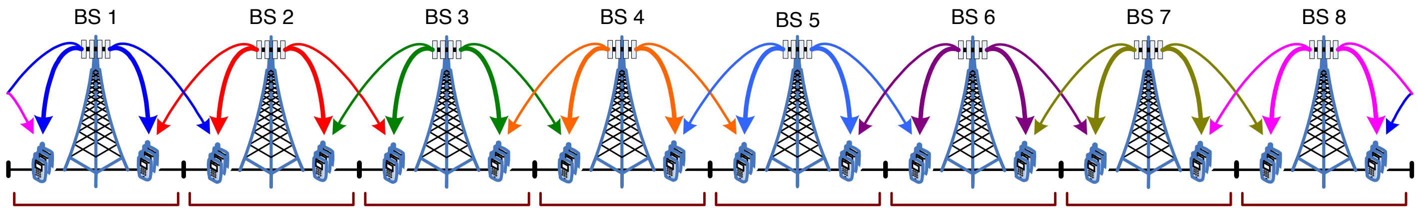
- We consider a linear cellular arrangement where M base stations are equally spaced on the segment $[-M, M]$ km, in positions $2m-M-1$ for $m = 1, \dots, M$ and K user groups are also equally spaced on the same segment, with K/M user groups uniformly spaced in each cell.

- The distance $d_{m,k}$ between BS m and user group k is defined modulo $[-M, M]$, i.e., we assume a wrap-around topology in order to eliminate boundary effects.

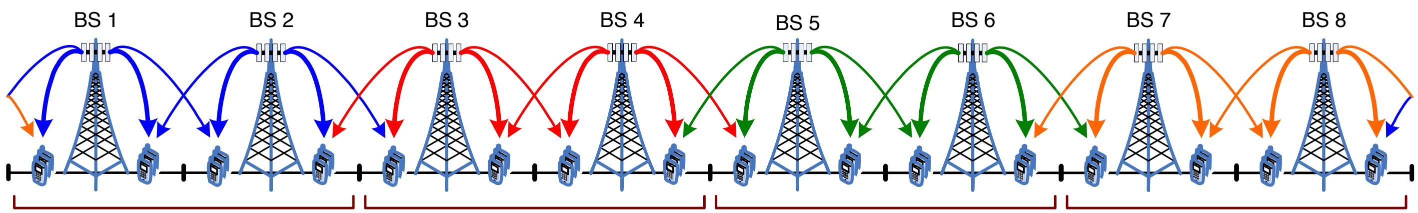
- We use a distance-dependent pathloss model given by $\alpha^2_{m,k} = G_0/(1 + (d_{m,k}/\delta)^\nu))$ and the pathloss parameters, G_0, ν, and δ follow the mobile WiMAX system evaluation specifications, such that the 3dB break point is $\delta = 36$m, the pathloss exponent is $\nu = 3.504$, the reference pathloss at $d_{m,k} = \delta$ is $G_0 = -91.64$ dB, and the per-BS transmit power normalized by the noise power at user terminals is $P = 154$ dB.

Linear cellular layout

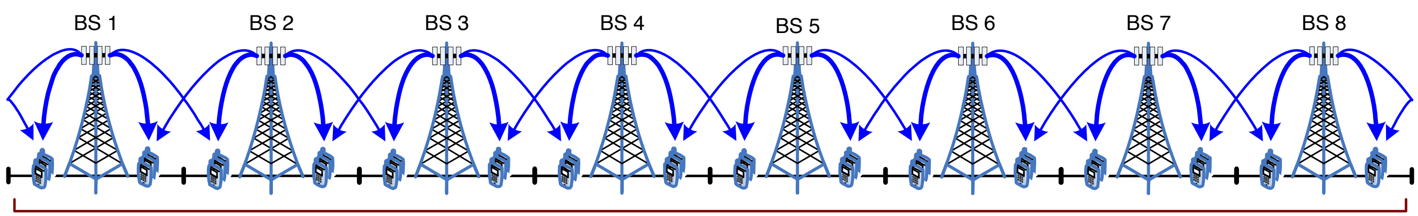
Example: linear cellular layout with $M = 8$ cells



$|\mathcal{M}_\ell| = B = 1$ cell cooperation

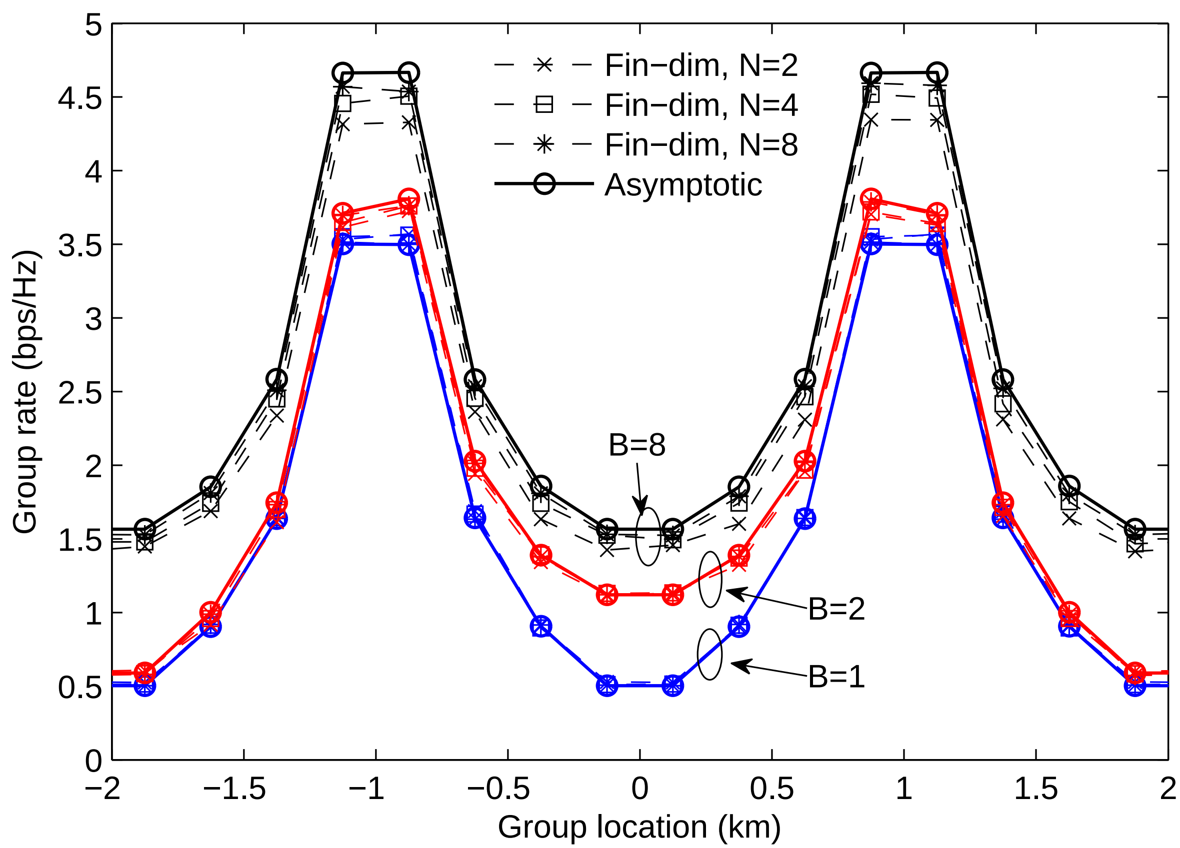


$|\mathcal{M}_\ell| = B = 2$ cell cooperation



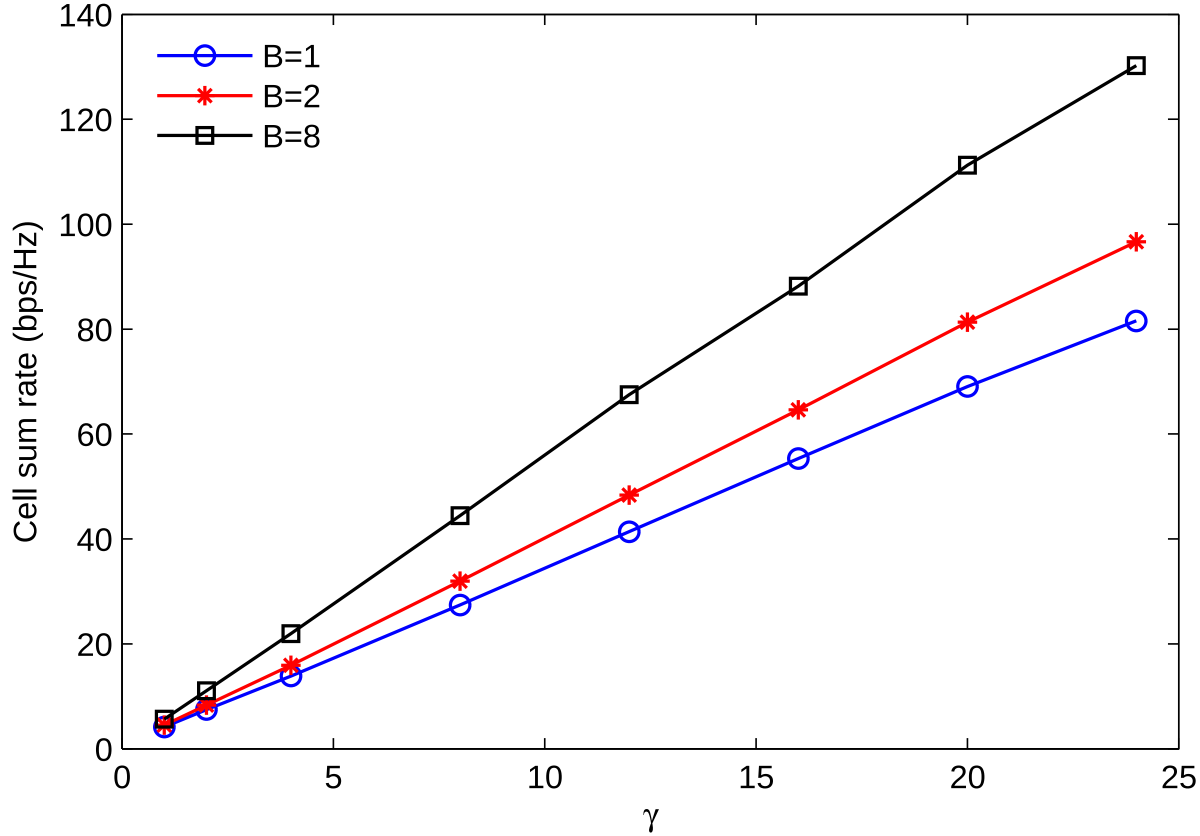
$|\mathcal{M}_\ell| = B = 8$ cell cooperation

Comparison with finite dimensional systems



User group rate in finite dimension (N = 2, 4, and 8) for cooperation clusters of size B=1, 2, and 8, with perfect CSIT. $M = 8$ cells and $K = 64$ user groups.

Increasing the number of antennas at each BC



Cell sum rate versus the antenna ratio ρ for cooperation clusters of size B=1, 2, and 8. $M = 8$ cells and $K = 192$ user groups.

End of Lecture 5

Lecture 6:
Deterministic Approximations

Stieltjes Transform

- One of the mast useful tools in Random Matrix Theory is the Stieltjes Transform, defined by

$$m_X(z) = \mathbb{E}\left[\frac{1}{X-z}\right] = \int_{-\infty}^{\infty} \frac{1}{x-z} dF_X(x), \qquad z \in \mathbb{C}$$

- For non-negative X, $m_X(z)$ is analytical in $\mathbb{C} - \mathbb{R}_+$.

- Stieltjes transform and moments:

$$m_X(z) = -\frac{1}{z}\sum_{k=1}^{\infty} \frac{\mathbb{E}[X^k]}{z^k}$$

- Inversion formula:

$$f_X(x) = \lim_{\omega \to 0_+} \frac{1}{\pi} \operatorname{Im}\left\{m_X(x + j\omega)\right\}$$

- Stieltjes transform and η-transform

$$\eta_X(\gamma) = \frac{1}{\gamma} m_X\left(-\frac{1}{\gamma}\right)$$

A Recent Useful Result

- The following result turns out to be very useful to analyze cases of structured channel matrices beyond the case of variance profile.

- For proofs, see [S. Wagner, R. Couillet, M. Debbah and D. T. M. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels under Limited Feedback," IT Trans. 2012.].

- Consider a matrix $\mathbf{B} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{\Xi}$, where $\mathbf{\Xi} \in \mathbb{C}^{N\times N}$ is Hermitian symmetric nonnegative definite, and $\mathbf{H} \in \mathbb{C}^{N\times K}$ is formed by columns

$$\mathbf{h}_k = \mathbf{\Psi}_k \mathbf{w}_k$$

with $\mathbf{\Psi}_k \in \mathbb{C}^{N\times r_k}$ and $\mathbf{w}_k$ has i.i.d. elements with variance $1/N$ and 8-th order moment that decreases as $O(1/N^4)$ (e.g., complex Gaussian will do).

- Define $\mathbf{\Theta}_k = \mathbf{\Psi}_k \mathbf{\Psi}_k^{\mathsf{H}}$ and $\mathbf{Q} \in \mathbb{C}^{N\times N}$ to be deterministic, with bounded spectral norm.

- Define the quantity

$$m_{\mathbf{B},\mathbf{Q}}(z) = \frac{1}{N}\text{tr}\left(\mathbf{Q}(\mathbf{B} - z\mathbf{I})^{-1}\right)$$

- Then, for $z \in \mathbb{C} - \mathbf{R}_+$, as $N \to \infty$ with $\beta = K/N$, and $\beta_k = r_k/N$, we have

$$m_{\mathbf{B},\mathbf{Q}}(z) - m^o_{\mathbf{B},\mathbf{Q}}(z) \xrightarrow{a.s.} 0$$

where $m^o_{\mathbf{B},\mathbf{Q}}(z)$ is the Stieltjes transform of a non-negative RV with compactly supported distribution, and is given by

$$m^o_{\mathbf{B},\mathbf{Q}}(z) = \frac{1}{N}\text{tr}\left(\mathbf{Q}\left(\frac{1}{N}\sum_{k=1}^{K}\frac{\mathbf{\Theta}_k}{1 + e_k(z)} + \mathbf{\Xi} - z\mathbf{I}\right)^{-1}\right)$$

with the terms $\{e_k(z)\}$ given by the unique non-negative solution of the system of coupled fixed point equations

$$e_k(z) = \frac{1}{N}\mathrm{tr}\left(\mathbf{\Theta}_k \left(\frac{1}{N}\sum_{j=1}^{K}\frac{\mathbf{\Theta}_j}{1+e_j(z)} + \mathbf{\Xi} - z\mathbf{I}\right)^{-1}\right)$$

Sketch of Proof

- The goal is to find a convergent deterministic approximation (usually referred to as "deterministic equivalent") to the sequence of random variables $m_{\mathbf{B},\mathbf{Q}}(z)$, for $N \to \infty$.

- To this purpose, let $\mathbf{D}$ denote a sequence of deterministic matrices and assume
$$\frac{1}{N}\text{tr}\left(\mathbf{Q}(\mathbf{B}-z\mathbf{I})^{-1}\right) - \frac{1}{N}\text{tr}\mathbf{D}^{-1} \xrightarrow{a.s.} 0$$

- We use the "resolvent formula": $\mathbf{U}^{-1} - \mathbf{V}^{-1} = -\mathbf{U}^{-1}(\mathbf{U}-\mathbf{V})\mathbf{V}^{-1}$, and write

$$\begin{aligned}
\mathbf{Q}(\mathbf{B}-z\mathbf{I})^{-1} - \mathbf{D}^{-1} &= \mathbf{D}^{-1}\left(\mathbf{D} - (\mathbf{B}-z\mathbf{I})\mathbf{Q}^{-1}\right)\mathbf{Q}(\mathbf{B}-z\mathbf{I})^{-1} \\
&= \mathbf{D}^{-1}\left(\mathbf{D} - (\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{\Xi} - z\mathbf{I})\mathbf{Q}^{-1}\right)\mathbf{Q}(\mathbf{B}-z\mathbf{I})^{-1}
\end{aligned}$$

- We postulate $\mathbf{D} = (\mathbf{R}+\boldsymbol{\Xi}-z\mathbf{I})\mathbf{Q}^{-1}$, where $\mathbf{R}$ is a deterministic approximation (in some sense) of the random matrix $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ to be specified later. Replacing, we have

$$\mathbf{Q}(\mathbf{B}-z\mathbf{I})^{-1} - \mathbf{D}^{-1} = \mathbf{D}^{-1}\mathbf{R}(\mathbf{B}-z\mathbf{I})^{-1} - \mathbf{D}^{-1}\mathbf{H}\mathbf{H}^{\mathsf{H}}(\mathbf{B}-z\mathbf{I})^{-1}$$

- Recalling that $\mathbf{H}\mathbf{H}^{\mathsf{H}} = \sum_{k=1}^{K} \boldsymbol{\Psi}_k \mathbf{w}_k \mathbf{w}_k^{\mathsf{H}} \boldsymbol{\Psi}_k$, we have

$$\begin{aligned}
\frac{1}{N}\mathrm{tr}\left(\mathbf{D}^{-1}\mathbf{H}\mathbf{H}^{\mathsf{H}}(\mathbf{B}-z\mathbf{I})^{-1}\right) &= \frac{1}{N}\mathrm{tr}\left(\mathbf{D}^{-1}\sum_{k=1}^{K}\boldsymbol{\Psi}_k\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}}\boldsymbol{\Psi}_k(\mathbf{B}-z\mathbf{I})^{-1}\right) \\
&= \frac{1}{N}\sum_{k=1}^{K}\mathbf{w}_k^{\mathsf{H}}\boldsymbol{\Psi}_k(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{D}^{-1}\boldsymbol{\Psi}_k\mathbf{w}_k
\end{aligned}$$

- Next, we write $\mathbf{B} = \mathbf{B}_k + \boldsymbol{\Psi}_k \mathbf{w}_k \mathbf{w}_k^{\mathsf{H}} \boldsymbol{\Psi}_k^{\mathsf{H}}$ and apply the matrix inversion lemma, to obtain

$$\frac{1}{N} \operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{H}\mathbf{H}^{\mathsf{H}}(\mathbf{B} - z\mathbf{I})^{-1}\right) = \frac{1}{N}\sum_{k=1}^{K} \frac{\mathbf{w}_k^{\mathsf{H}} \boldsymbol{\Psi}_k (\mathbf{B}_k - z\mathbf{I})^{-1} \mathbf{D}^{-1} \boldsymbol{\Psi}_k \mathbf{w}_k}{1 + \mathbf{w}_k^{\mathsf{H}} \boldsymbol{\Psi}_k (\mathbf{B}_k - z\mathbf{I})^{-1} \boldsymbol{\Psi}_k \mathbf{w}_k}$$

- Term should become similar to the second term

$$\frac{1}{N} \operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}(\mathbf{B} - z\mathbf{I})^{-1}\right)$$

- Furthermore, for the trace lemma, notice that

$$\mathbf{w}_k^{\mathsf{H}} \boldsymbol{\Psi}_k (\mathbf{B}_k - z\mathbf{I})^{-1} \boldsymbol{\Psi}_k \mathbf{w}_k \xrightarrow{a.s.} \frac{1}{N} \operatorname{tr}\left(\boldsymbol{\Theta}_k (\mathbf{B}_k - z\mathbf{I})\right) \approx \frac{1}{N} \operatorname{tr}\left(\boldsymbol{\Theta}_k (\mathbf{B} - z\mathbf{I})\right)$$

where the last approximate equality can be made asymptotically rigorous (finite rank perturbation).

- Our of good intuition, we choose

$$\mathbf{R} = \frac{1}{N}\sum_{k=1}^{K} \frac{\mathbf{\Theta}_k}{1 + \frac{1}{N}\mathrm{tr}(\mathbf{\Theta}_k(\mathbf{B} - z\mathbf{I})^{-1})}$$

and let $e_k(z) = \frac{1}{N}\mathrm{tr}\left(\mathbf{\Theta}_k(\mathbf{B} - z\mathbf{I})^{-1}\right) \approx \frac{1}{N}\mathrm{tr}\left(\mathbf{\Theta}_k(\mathbf{R} + \mathbf{\Xi} - z\mathbf{I})^{-1}\right)$.

- Therefore, if this is true, the functions $e_k(z)$ must satisfy

$$e_k(z) = \frac{1}{N}\mathrm{tr}\left(\mathbf{\Theta}_k\left(\frac{1}{N}\sum_{j=1}^{K}\frac{\mathbf{\Theta}_j}{1 + e_j(z)} + \mathbf{\Xi} - z\mathbf{I}\right)^{-1}\right)$$

- The rest of the proof is dedicated to developing rigorous bounds to show that this convergence actually occurs, and that the system of coupled equations defining $\{e_k(z)\}$ converges to a unique solution compatible with the property of Stieltjes transforms.

Sanity check

- Suppose $\boldsymbol{\Xi} = \mathbf{0}$, $\mathbf{Q} = \mathbf{I}$, $\boldsymbol{\Psi}_k = \sqrt{T_k}\mathbf{I}$. Then $\mathbf{B} = \mathbf{STS}^{\mathsf{H}}$, in the form we have already seen several times.

- Given the relation between Stieltjes transform and η-transform, we expect that

$$\frac{1}{\gamma} m_{\mathbf{B}}\left(-\frac{1}{\gamma}\right) = \eta_{\mathbf{STS}^{\mathsf{H}}}(\gamma) = \eta$$

solution of (from the key equation of Theorem 4)

$$\eta = \frac{1}{1 + \beta\gamma\mathbb{E}\left[\frac{\mathsf{T}}{1+\gamma\eta\mathsf{T}}\right]}$$

- We wish to recover this result from the general deterministic equivalent case. The iteration for $\{e_k(z)\}$ becomes

$$e_k(z) = \frac{T_k}{\frac{\beta}{K}\sum_{j=1}^{K}\frac{T_j}{1+e_j(z)} - z}$$

- Dividing by T_k we find

$$\frac{e_k(z)}{T_k} = \frac{1}{\frac{\beta}{K}\sum_{j=1}^{K} \frac{T_j}{1+e_j(z)} - z}$$

- We conclude that $e_k(z)/T_k$ does not depend on k. We call this quantity $\mu(z)$. Therefore:

$$\mu(z) = \frac{1}{\frac{\beta}{K}\sum_{j=1}^{K} \frac{T_j}{1+T_j\mu(z)} - z}$$

- Equivalently

$$\mu(-1/\gamma)/\gamma = \frac{1}{1 + \gamma\beta\frac{1}{K}\sum_{j=1}^{K} \frac{T_j}{1+\gamma T_j\mu(-1/\gamma)/\gamma}}$$

- Identifying terms, we have $\mu(-1/\gamma)/\gamma = \eta(\gamma)$, such that we recover the fixed point equation

$$\eta = \frac{1}{1 + \gamma\beta\frac{1}{K}\sum_{j=1}^{K}\frac{T_j}{1+\gamma\eta T_j}}$$

- Finally, we notice that in this special case we have

$$m_{\mathbf{B}}(z) = \frac{1}{\frac{\beta}{K}\sum_{k=1}^{K}\frac{T_k}{1+e_k(z)} - z}$$

- Replacing $z = -1/\gamma$, $e_k(-1/\gamma) = T_k\mu(-1/\gamma) = \gamma T_k\eta(\gamma)$, and multiplying both sides by $1/\gamma$, we find that $m_{\mathbf{B}}(-1/\gamma)/\gamma = \eta(\gamma)$ since it satisfies the same equation.

Application: massive MIMO with antenna correlations

- In FDD systems, “massive MIMO” is impractical since the downlink training and the CSIT feedback consume too many dimensions.

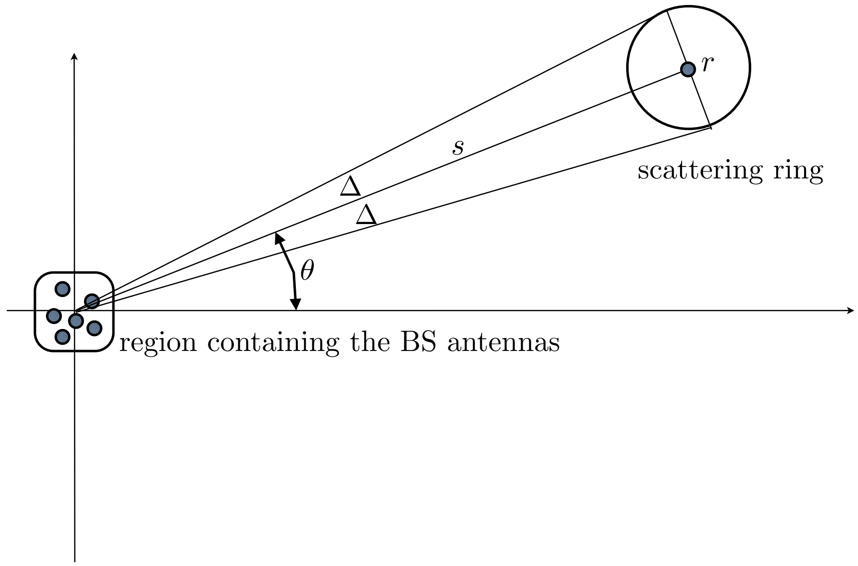
- Idea: we can exploit the channel correlation in order to achieve a channel dimensionality reduction, while retaining the benefits of massive MIMO.

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

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

- Two users separated by a few meters (say 10 λ) are practically uncorrelated.

- In contrast, the base station sees user groups at different AoAs under narrow AS $\Delta \approx \arctan(\mathsf{r}/\mathsf{s})$.



- This leads to the Tx antenna correlation model

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

with

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

Joint Space Division and Multiplexing (JSDM)

- K users selected to form G groups, with $\approx$ same channel correlation.

$$\underline{\mathbf{H}} = [\mathbf{H}_1, \ldots, \mathbf{H}_G], \text{ with } \mathbf{H}_g = \mathbf{U}_g \mathbf{\Lambda}_g^{1/2} \mathbf{W}_g.$$

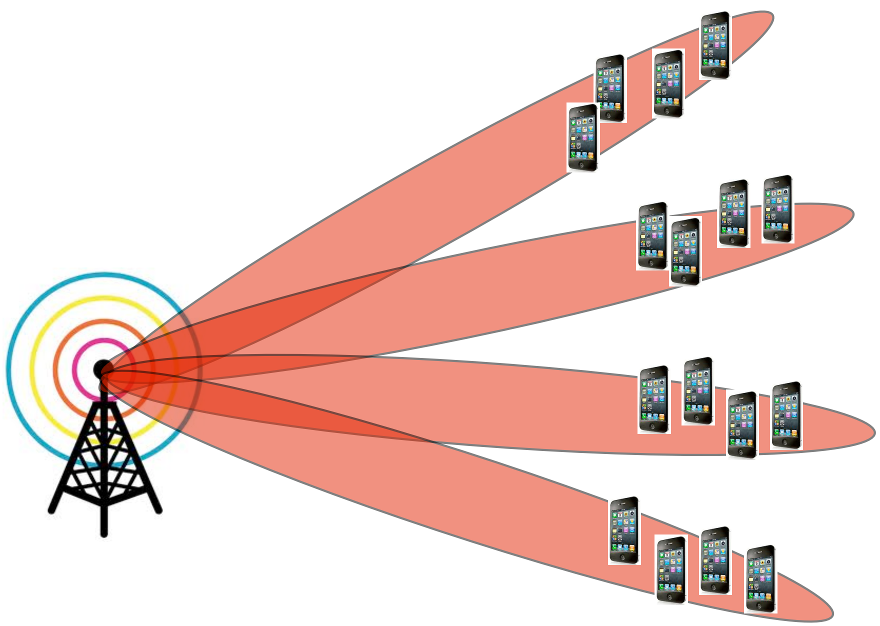
- Two-stage precoding: $\mathbf{V} = \mathbf{B}\mathbf{P}$.
- $\mathbf{B} \in \mathbb{C}^{M \times b_g}$ is a pre-beamforming matrix function of $\{\mathbf{U}_g, \mathbf{\Lambda}_g\}$ only.
- $\mathbf{P} \in \mathbb{C}^{b_g \times S_g}$ is a precoding matrix that depends on the effective channel.
- The effective channel matrix is given by

$$\underline{\mathsf{H}} = \begin{bmatrix} \mathbf{B}_1^{\mathsf{H}}\mathbf{H}_1 & \mathbf{B}_1^{\mathsf{H}}\mathbf{H}_2 & \cdots & \mathbf{B}_1^{\mathsf{H}}\mathbf{H}_G \\ \mathbf{B}_2^{\mathsf{H}}\mathbf{H}_1 & \mathbf{B}_2^{\mathsf{H}}\mathbf{H}_2 & \cdots & \mathbf{B}_2^{\mathsf{H}}\mathbf{H}_G \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_G^{\mathsf{H}}\mathbf{H}_1 & \mathbf{B}_G^{\mathsf{H}}\mathbf{H}_2 & \cdots & \mathbf{B}_G^{\mathsf{H}}\mathbf{H}_G \end{bmatrix}.$$

- Per-Group Processing: If estimation and feedback of the whole $\underline{\mathbf{H}}$ is still too costly, then each group estimates its own diagonal block $\mathbf{H}_g = \mathbf{B}_g^{\mathsf{H}} \mathbf{H}_g$, and $\mathbf{P} = \mathrm{diag}(\mathbf{P}_1, \cdots, \mathbf{P}_G)$.

- This results in

$$\mathbf{y}_g = \mathbf{H}_g^{\mathsf{H}} \mathbf{B}_g \mathbf{P}_g \mathbf{d}_g + \sum_{g' \neq g} \mathbf{H}_g^{\mathsf{H}} \mathbf{B}_{g'} \mathbf{P}_{g'} \mathbf{d}_{g'} + \mathbf{z}_g$$



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 $\underline{\mathbf{U}} = [\mathbf{U}_1, \cdots, \mathbf{U}_G]$ is $M \times r$ *tall unitary* (i.e., $r \leq M$ and $\underline{\mathbf{U}}^{\mathsf{H}}\underline{\mathbf{U}} = \mathbf{I}_r$).

- 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

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

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

Theorem 17. *For $\underline{\mathbf{U}}$ 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 $\mathbf{B}_g$ such that:

$$\mathbf{U}_{g'}^{\mathsf{H}} \mathbf{B}_g = \mathbf{0} \quad \forall \ g' \neq g, \quad \text{and} \ \ \mathrm{rank}(\mathbf{U}_g^{\mathsf{H}} \mathbf{B}_g) \geq S_g$$

- Necessary condition for exact BD

$$\mathrm{Span}(\mathbf{B}_g) \subseteq \mathrm{Span}^{\perp}(\{\mathbf{U}_{g'} : g' \neq g\}).$$

- When $\mathrm{Span}^{\perp}(\{\mathbf{U}_{g'} \ : \ g' \ \neq \ g\})$ 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 $r_g^{\star} < r_g$ dominant eigenmodes for each group g can be implemented.

Performance analysis with regularized ZF

- The transformed channel matrix $\underline{\mathbf{H}}$ 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

$$\mathbf{P}_{g,\mathrm{rzf}} = \bar{\zeta}_g \bar{\mathbf{K}}_g \mathbf{H}_g,$$

where

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

and where

$$\bar{\zeta}_g^2 = \frac{S'}{\mathrm{tr}(\mathbf{H}_g^{\mathsf{H}} \mathbf{K}_g^{\mathsf{H}} \mathbf{B}_g^{\mathsf{H}} \mathbf{B}_g \mathbf{K}_g \mathbf{H}_g)}.$$

- The SINR of user g_k given by

$$\gamma_{g_k,\mathrm{pgp}} = \frac{\frac{P}{S}\bar{\zeta}_g^2 |\mathbf{h}_{g_k}^{\mathsf{H}} \mathbf{B}_g \bar{\mathbf{K}}_g \mathbf{B}_g^{\mathsf{H}} \mathbf{h}_{g_k}|^2}{\frac{P}{S}\sum_{j\neq k} \bar{\zeta}_g^2 |\mathbf{h}_{g_k}^{\mathsf{H}} \mathbf{B}_g \bar{\mathbf{K}}_g \mathbf{B}_g^{\mathsf{H}} \mathbf{h}_{g_j}|^2 + \frac{P}{S}\sum_{g'\neq g}\sum_j \bar{\zeta}_{g'}^2 |\mathbf{h}_{g_k}^{\mathsf{H}} \mathbf{B}_{g'} \bar{\mathbf{K}}_{g'} \mathbf{B}_{g'}^{\mathsf{H}} \mathbf{h}_{g'_j}|^2 + 1}$$

- Using the "deterministic equivalent" method we can calculate $\gamma_{g_k,\mathrm{pgp}}^o$ such that

$$\gamma_{g_k,\mathrm{pgp}} - \gamma_{g_k,\mathrm{pgp}}^o \xrightarrow{M\to\infty} 0$$

- First, we consider the terms appearing in the numerator and denominator in $\gamma_{g_k,\mathrm{pgp}}$ and express them as Stieltjes transforms of the form $\frac{1}{M}\mathrm{tr}(\mathbf{Q}(\mathbf{B}-z\mathbf{I})^{-1})$ evaluated as some appropriate value of $z \in \mathbb{R}_-$.

- Then, we repeatedly use the deterministic equivalent result.

- Finally, we pull all these terms together and express them as a single system of fixed-point equations.

- For the sake of completeness, we include the final result (after many pages of calculation): letting

$$\bar{\mathbf{R}}_g = \mathbf{B}_g^{\mathsf{H}} \mathbf{R}_g \mathbf{B}_g$$

denote the covariance matrix of users in group g, we have

$$\gamma^o_{g_k,\mathrm{pgp},\mathrm{rzf}} = \frac{\frac{P}{S}\bar{\zeta}_g^2 (\bar{m}_g^o)^2}{\bar{\zeta}_g^2 \bar{\Upsilon}^o_{g,g} + (1 + \sum_{g' \neq g} \bar{\zeta}_{g'}^2 \bar{\Upsilon}^o_{g,g'})(1 + \bar{m}_g^o)^2}, \tag{43}$$

where $\bar{\zeta}_g^2 = \frac{P/G}{\bar{\Gamma}_g^o}$ and the quantities $\bar{m}_g^o$, $\bar{\Upsilon}^o_{g,g}$, $\bar{\Upsilon}^o_{g,g'}$ and $\bar{\Gamma}_g^o$ are given by

$$\bar{m}_g^o = \frac{1}{b'}\text{tr}\left(\bar{\mathbf{R}}_g \bar{\mathbf{T}}_g\right) \tag{44}$$

$$\bar{\mathbf{T}}_g = \left(\frac{S'}{b'}\frac{\bar{\mathbf{R}}_g}{1+\bar{m}_g^o} + \alpha \mathbf{I}_{b'}\right)^{-1} \tag{45}$$

$$\bar{\Gamma}_g^o = \frac{1}{b'}\frac{P}{G}\frac{\bar{n}_g}{(1+\bar{m}_g^o)^2} \tag{46}$$

$$\bar{\Upsilon}_{g,g}^o = \frac{1}{b'}\frac{S'-1}{S'}\frac{P}{G}\frac{\bar{n}_{g,g}}{(1+\bar{m}_g^o)^2} \tag{47}$$

$$\bar{\Upsilon}_{g,g'}^o = \frac{1}{b'}\frac{P}{G}\frac{\bar{n}_{g',g}}{(1+\bar{m}_{g'}^o)^2} \tag{48}$$

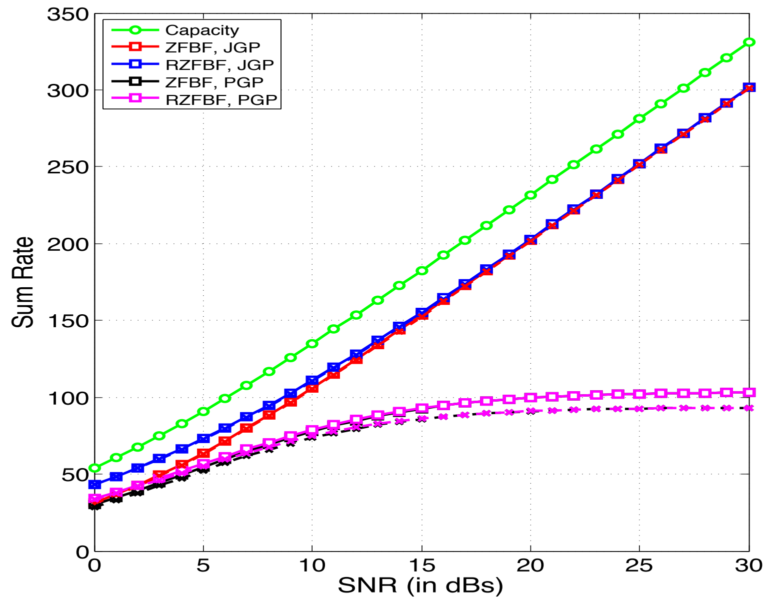
$$\bar{n}_g = \frac{\frac{1}{b'}\mathrm{tr}\left(\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\mathbf{B}_g^{\mathsf{H}}\mathbf{B}_g\bar{\mathbf{T}}_g\right)}{1-\frac{\frac{S'}{b'}\mathrm{tr}\left(\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\right)}{b'(1+\bar{m}_g^o)^2}} \tag{49}$$

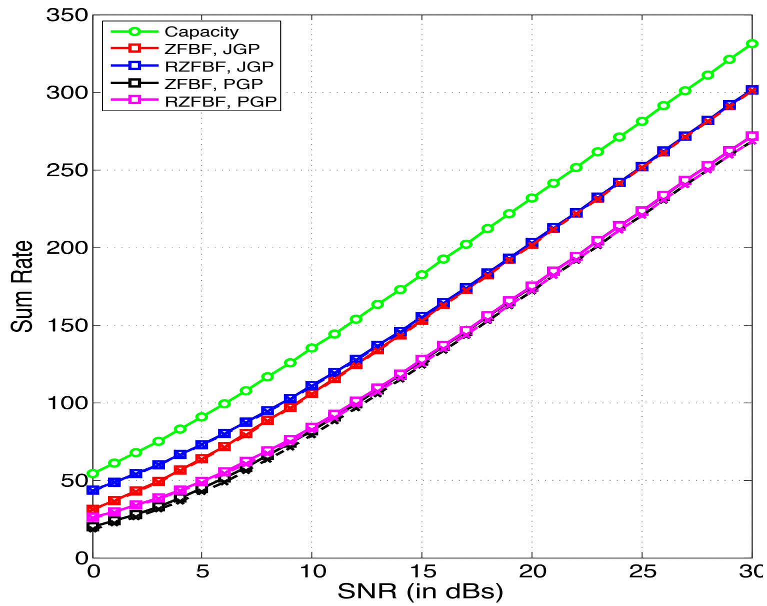
$$\bar{n}_{g,g} = \frac{\frac{1}{b'}\mathrm{tr}\left(\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\right)}{1-\frac{\frac{S'}{b'}\mathrm{tr}\left(\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\bar{\mathbf{R}}_g\bar{\mathbf{T}}_g\right)}{b'(1+\bar{m}_g^o)^2}} \tag{50}$$

$$\bar{n}_{g',g} = \frac{\frac{1}{b'}\mathrm{tr}\left(\bar{\mathbf{R}}_{g'}\bar{\mathbf{T}}_{g'}\mathbf{B}_{g'}^{\mathsf{H}}\mathbf{R}_g\mathbf{B}_{g'}\bar{\mathbf{T}}_{g'}\right)}{1-\frac{\frac{S'}{b'}\mathrm{tr}\left(\bar{\mathbf{R}}_{g'}\bar{\mathbf{T}}_{g'}\bar{\mathbf{R}}_{g'}\bar{\mathbf{T}}_{g'}\right)}{b'(1+\bar{m}_{g'}^o)^2}} \tag{51}$$

Example

- $M = 100$, $G = 6$ user groups, Rank$(\mathbf{R}_g) = 21$, effective rank $r_g^* = 11$.
- We serve $S' = 5$ users per group with $b' = 10$, $r^\star = 6$ and $r^\star = 12$.
- 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: 100×30 channel matrix $\Rightarrow$ 3000 complex channel coefficients per coherence block (CSI feedback), with 100×100 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 10×10 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 5×5, with respect to one matrix inversion of dimension 30×30.

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, θ, Δ.

- 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

$$S(\xi) = \frac{1}{2\Delta} \sum_{m \in [D\sin(-\Delta+\theta)+\xi, D\sin(\Delta+\theta)+\xi]} \frac{1}{\sqrt{D^2 - (m-\xi)^2}}.$$

Szego's Theorem: eigenvalues

Theorem 18. *The empirical spectral distribution of the eigenvalues of* $\mathbf{R}$*,*

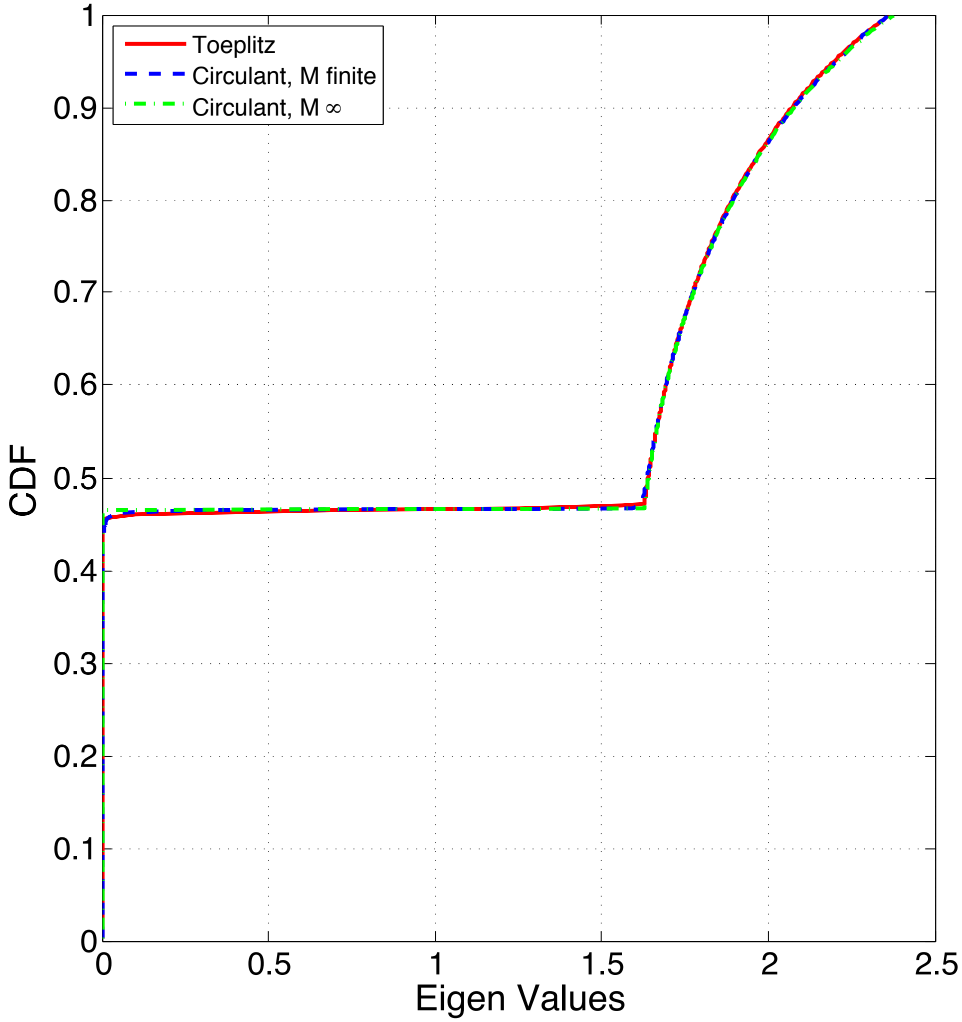
$$F_{\mathbf{R}}^{(M)}(\lambda) = \frac{1}{M} \sum_{m=1}^{M} 1\{\lambda_m(\mathbf{R}) \leq \lambda\},$$

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 19. *Let* $\lambda_0(\mathbf{R}) \leq \ldots, \leq \lambda_{M-1}(\mathbf{R})$ *and* $\lambda_0(\mathbf{C}) \leq \ldots, \leq \lambda_{M-1}(\mathbf{C})$ *denote the set of ordered eigenvalues of* $\mathbf{R}$ *and* $\mathbf{C}$*, and let* $\mathbf{U} = [\mathbf{u}_0, \ldots, \mathbf{u}_{M-1}]$ *and* $\mathbf{F} = [\mathbf{f}_0, \ldots, \mathbf{f}_{M-1}]$ *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* $\mathcal{I}_{[a,b]} = \{m : \lambda_m(\mathbf{R}) \in [a, b]\}$ *and* $\mathcal{J}_{[a,b]} = \{m : \lambda_m(\mathbf{C}) \in [a, b]\}$*, and define* $\mathbf{U}_{[a,b]} = (\mathbf{u}_m : m \in \mathcal{I}_{[a,b]})$ *and* $\mathbf{F}_{[a,b]} = (\mathbf{f}_m : m \in \mathcal{J}_{[a,b]})$ *be the submatrices of* $\mathbf{U}$ *and* $\mathbf{F}$ *formed by the columns whose indices belong to the sets* $\mathcal{I}_{[a,b]}$ *and* $\mathcal{J}_{[a,b]}$*, respectively. Then, the eigenvectors of* $\mathbf{C}$ *approximate the eigenvectors of* $\mathbf{R}$ *in the sense that*

$$\lim_{M \to \infty} \frac{1}{M} \left\| \mathbf{U}_{[a,b]} \mathbf{U}^{\mathsf{H}}_{[a,b]} - \mathbf{F}_{[a,b]} \mathbf{F}^{\mathsf{H}}_{[a,b]} \right\|_F^2 = 0.$$

■

Consequence 1: $\mathbf{U}_g$ is well approximated by a "slice" of the DFT matrix.

Consequence 2: DFT pre-beamforming is near optimal for large M.

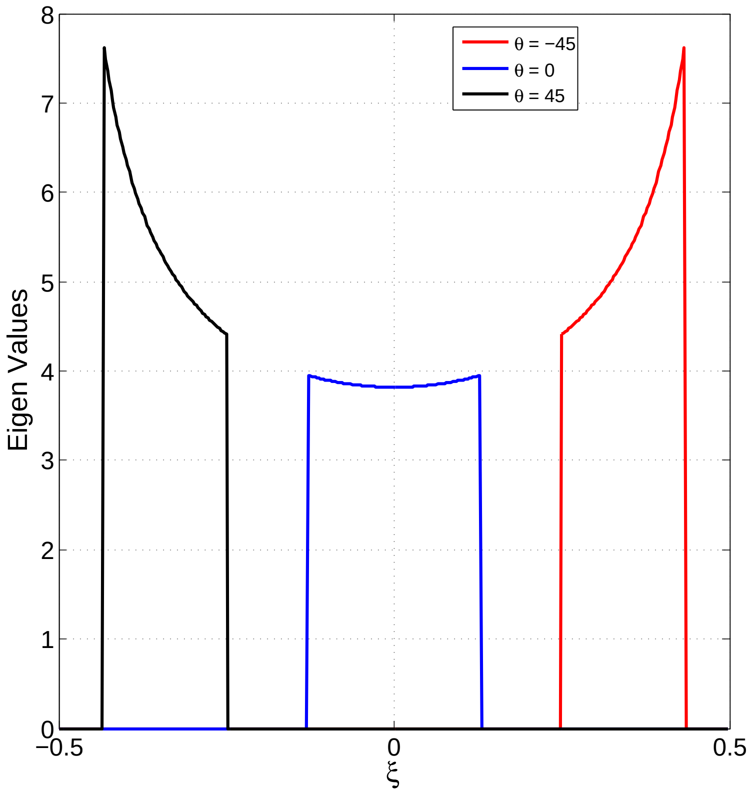
Theorem 20. *The asymptotic normalized rank of the channel covariance matrix $\mathbf{R}$, with antenna separation λD, AoA θ and AS Δ, is given by*

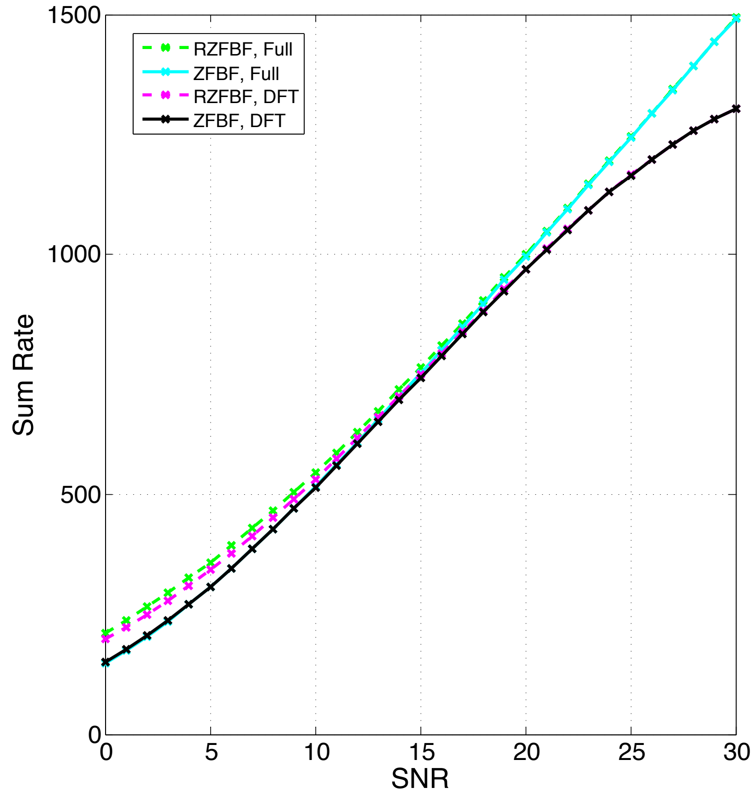
$$\rho = \min\{1, B(D, \theta, \Delta)\},$$

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

Theorem 21. *Groups g and g' with angle of arrival θ_g and $\theta_{g'}$ and common angular spread Δ have spectra with disjoint support if their AoA intervals $[\theta_g - \Delta, \theta_g + \Delta]$ and $[\theta_{g'} - \Delta, \theta_{g'} + \Delta]$ are disjoint.* ■

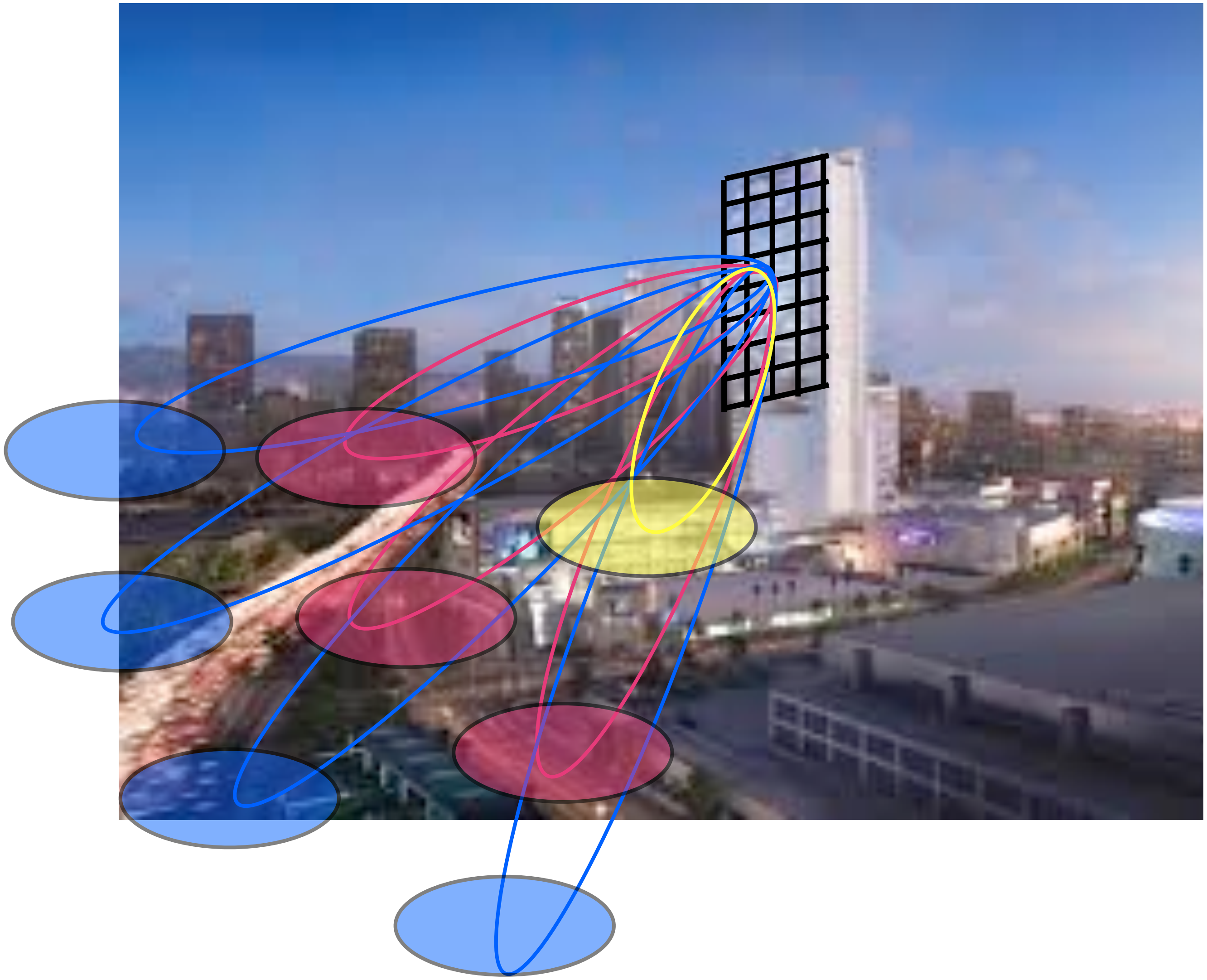
DFT Pre-Beamforming





- 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 ℓ-th region into G_ℓ 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.

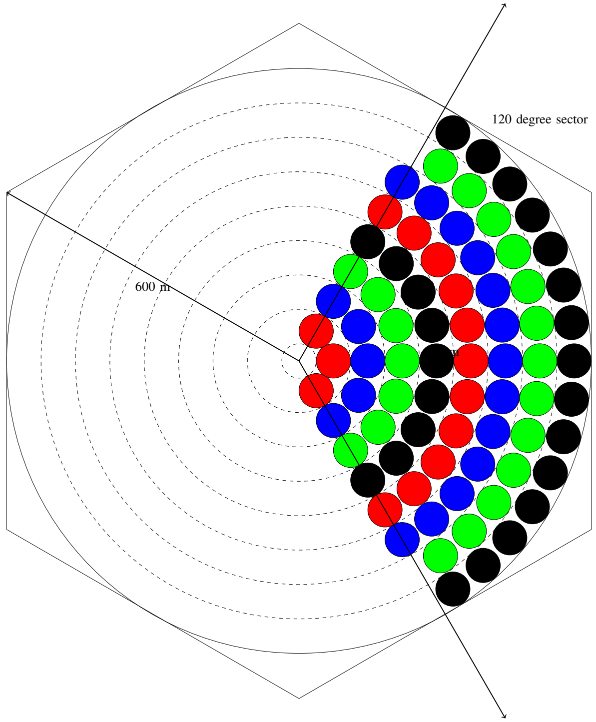
- A set of simultaneously served groups forms a “pattern”.

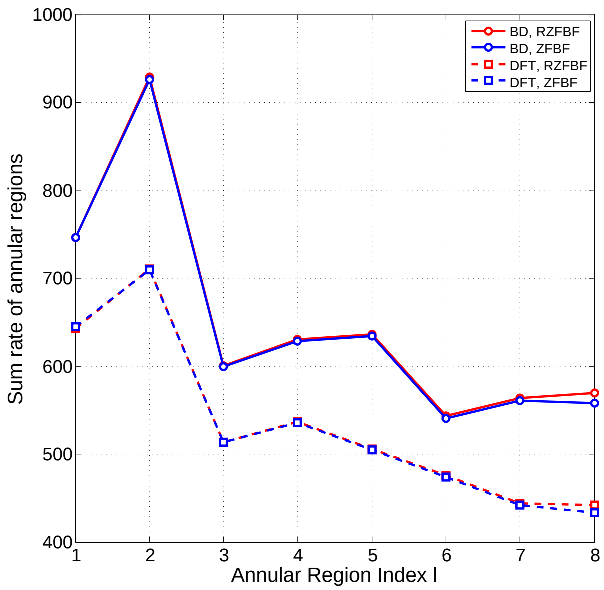
- Patterns need not cover the whole sector.

- 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 $g(x) = \frac{1}{1+(\frac{x}{d_0})^\delta}$ 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

Scheme	Approximate BD	DFT based
PFS, RZFBF	1304.4611	1067.9604
PFS, ZFBF	1298.7944	1064.2678
MAXMIN, RZFBF	1273.7203	1042.1833
MAXMIN, ZFBF	1267.2368	1037.2915

1000 bit/s/Hz $\times$ 40 MHz of bandwidth $= 40$ Gb/s per sector.

End of Lecture 6

Bibliography

- Fundamental references that contain the whole literature of results used in Communications, Signal Processing and Information Theory problems. The references below contain a large list of reference to important papers, as well as to the relevant results in the math literature.

1. A. Tulino and S. Verdu, “Random Matrix Theory and Wireless Communications,” Foundations and Trends in Communications and Information Theory, NOW Publisher, 2004.
2. R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, Cambridge University Press, 2012.

- The applications illustrated in this short course are taken from the following recent papers:

1. Huh, Hoon, Antonia M. Tulino, and Giuseppe Caire. ”Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling.” *IEEE Trans on Inform. Theory*, 58.5 (2012): 2911-2934.

2. Huh, Hoon, Sung-Hyun Moon, Young-Tae Kim, Inkyu Lee, and Giuseppe Caire. ”Multi-cell MIMO downlink with cell cooperation and fair scheduling: a large-system limit analysis.” *IEEE Trans on Inform. Theory*, 57, no. 12 (2011): 7771-7786.
3. Ansuman Adhikary, Junyoung Nam, Jae-Young Ahn, Giuseppe Caire, “Joint Spatial Division and Multiplexing,” Preprint: arXiv:1209.1402.

- Other relevant references (on deterministic equivalents):

1. Wagner, Sebastian, Romain Couillet, Mrouane Debbah, and Dirk TM Slock. ”Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback.” *IEEE Trans on Inform. Theory*, 58, no. 7 (2012): 4509-4537.
2. W. Hachem, P. Loubaton, and J. Najim, Deterministic Equivalents for Certain Functionals of Large Random Matrices, *Annals of Applied Probability*, vol. 17, no. 3, pp. 875930, Jun. 2007.

The End (Thank You)