Massive MIMO

Fundamentals, Trends and Recent Developments

Luca Sanguinetti, Emil Björnson

University of Pisa, Italy; luca.sanguinetti@unipi.it

Linköping University, Sweden; emil.bjornson@liu.se

ISWCS 2018, Lisbon, Aug. 28, 2018

Overview

Part I

- Definition of Massive MIMO
- Basic Channel and Signal Modeling
- Channel Estimation
- Spectral Efficiency in Uplink (and downlink)

Coffee break

Part II

- Spectral Efficiency in Uplink (cont'd)
- Spectral Efficiency: Asymptotic Analysis
- Practical Deployment Considerations
- Open Problems

Main Reference

"**Massive MIMO Networks: Spectral, Energy and Hardware Efficiency**" by E. Björnson, J. Hoydis, and L. Sanguinetti, Foundations and Trends® in Signal Processing: Vol. 11: No. 3-4, pp 154-655.

https://massivemimobook.com

- Monograph of 517 pages intended for PhD students and researchers;
- Printed books can be purchased, e-book freely available;
- Matlab code available online.

Additional material:

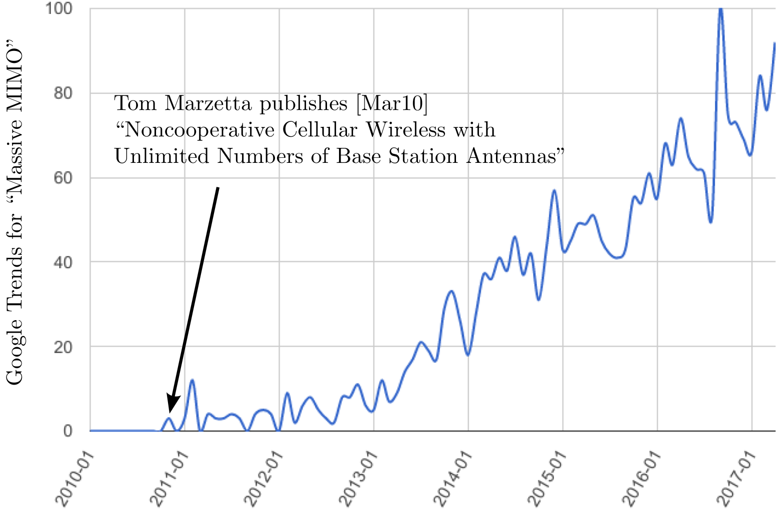
- "Fundamentals of Massive MIMO", by T. Marzetta, et al., Cambridge University, 2016
- Massive MIMO blog: http://massive-mimo.net/

Table of contents

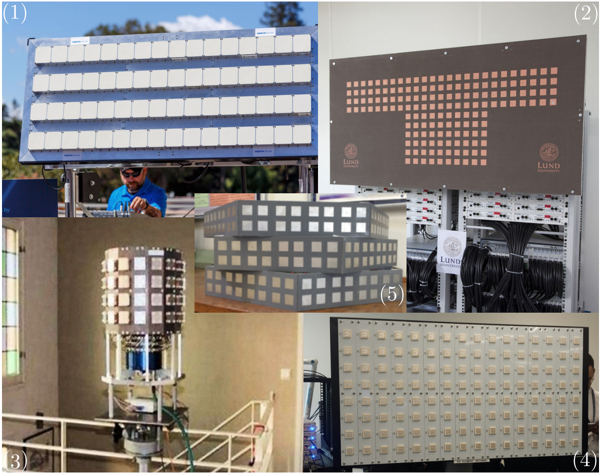
Contents

Table of contents (cont'd)

Massive MIMO is a rather novel technology...



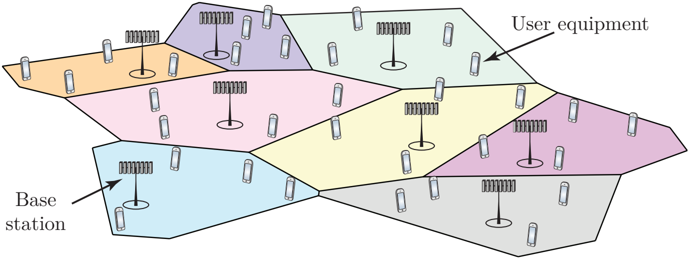
...but academia and industry are racing to build first prototypes



(1) https://www.cnet.com/pictures/bell-labs-historic-phones-and-next-gen-5g-networks-pictures/4/
(2) http://old.liu.se/elliit/Communication/highlights
(3) http://www.academia.edu/download/39443743/Large_MIMO.pdf
(4) http://the-mobile-network.com/2015/09/5gic-the-launch-the-goals-the-research-the-project/
(5) http://ieeexplore.ieee.org/document/7523981/

Introduction

Cellular Networks

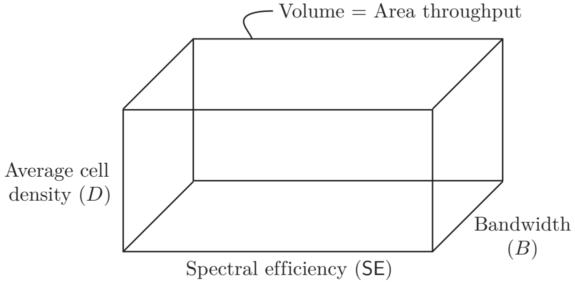


Definition (Cellular networks — A major breakthrough)

A cellular network consists of a set of *base stations (BSs)* and a set of *user equipments (UEs)*. Each UE is connected to one of the BS, which provides service to it.

- *Downlink (DL)* refers to signals sent from the BS to its UEs
- *Uplink (UL)* refers to signals sent from the UE to its respective BS

Area Throughput



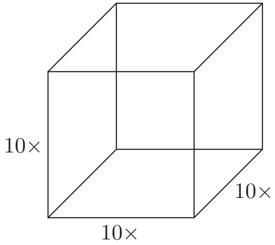
Definition (Area throughput)

The *area throughput* of a cellular network is measured in bit/s/km^2.

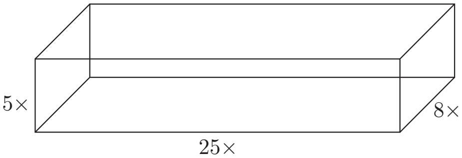
$$\text{Area throughput} = B \text{ [Hz]} \cdot D \text{ [cells/km}^2\text{]} \cdot \mathsf{SE} \text{ [bit/s/Hz/cell]}$$

where B is the bandwidth, D is the average cell density, and SE is the per-cell *spectral efficiency (SE)*. The SE is the amount of information transferred per second over a unit bandwidth.

How to Improve the Area Throughput?



(a) Equal improvement

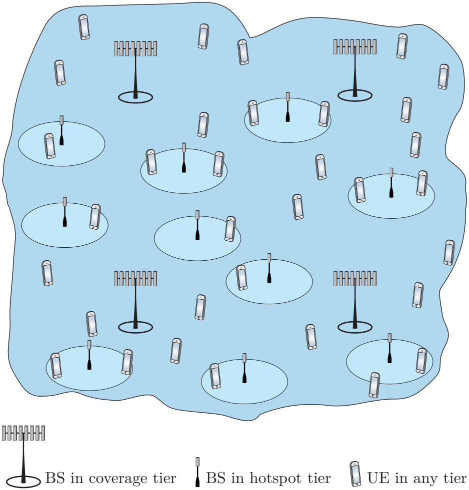


(b) Improving some factors more than others

Next generation networks: $1000\times$ higher area throughput [Qua12]

- Three main ways to achieve this:
    1. Allocate more bandwidth
    2. Densify the network by adding more BSs
    3. Improve the SE per cell
- Although there is an inherent dependence between the three factors, we can treat them as independent in a first-order approximation

Two Network Tiers



Definition (Hotspot tier)

BS offering high throughput in small local areas to a few UE.

- Very dense deployment possible
- Much bandwidth exist (mmWave)
- SE less important

Definition (Coverage tier)

BS providing wide-area coverage and mobility support to many UEs.

- Limited density and bandwidth
- Important to improve SE

Coverage tier is the most challenging – will be our focus

Spectral Efficiency

Nyquist-Shannon sampling theorem: A signal of bandwidth B Hz is determined by $2B$ real-valued equal-spaced samples per second.

- B complex-valued samples per second is the more natural quantity for the complex-baseband representation of the signal

Definition (Spectral efficiency)

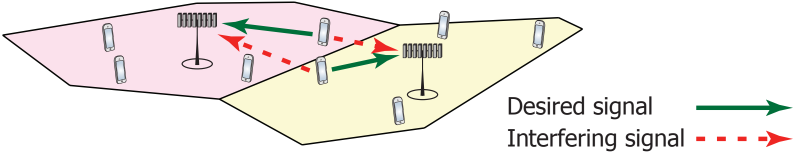
The *spectral efficiency (SE)* of an encoding/decoding scheme is a number of bits of information, per complex-valued sample, that can be reliably[1] transmitted over the channel under consideration.

Equivalent units:

- bit per complex-valued sample
- bit per second per Hertz (bit/s/Hz)

[1]With arbitrarily low error probability for sufficiently long signals

How to Improve Spectral Efficiency?



Two-cell Wyner model:

- Intra-cell signal-to-noise ratio (SNR): SNR.
- Inter-cell interference is $\bar{\beta} \leq 1$ weaker than intra-cell channels.
- M antennas per BS, K single-antenna UEs per cell

Sum SE with i.i.d. Rayleigh fading and Perfect Channel Knowledge

An achievable UL sum SE [bit/s/Hz/cell] is

$$\mathsf{SE} = K \log_2 \left(1 + \frac{M-1}{(K-1) + K\bar{\beta} + \frac{1}{\mathsf{SNR}}} \right).$$

- Grows logarithmically with M
- Pre-log grows linearly with K, but SINR decreases as $1/K$
- Avoid SINR reduction by increasing M, K jointly!

Canonical Definition and Notation

Canonical Massive MIMO Network

Definition (Canonical Massive MIMO Network)

A canonical Massive MIMO network is a multi-carrier cellular network with L cells that operate according to a synchronous TDD protocol.[2]

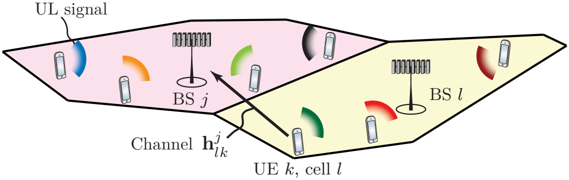
- BS j is equipped with $M_j \gg 1$ antennas, to achieve channel hardening
- BS j communicates with K_j single-antenna UEs on each time/frequency sample, where $M_j/K_j > 1$
- Each BS operates individually and processes its signals using linear transmit precoding and linear receive combining

[2] A synchronous TDD protocol refers to a protocol in which UL and DL transmissions within different cells are synchronized

Channel Notation

Numbering:

- L cells and BSs, numbered from 1 to L
- K_l UEs in cell l, numbered from 1 to K_l



Channel notation:

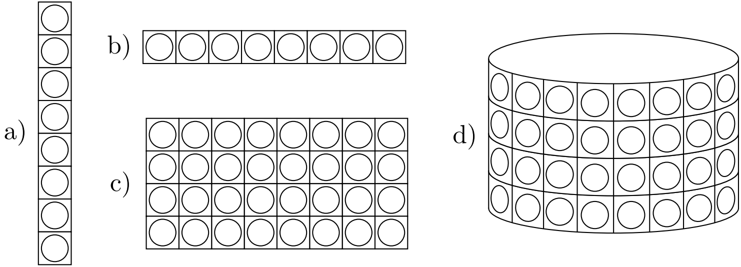
$$\mathbf{h}_{\text{UE's cell number, UE's number}}^{\text{BS's cell number}}$$

- Example: Channel between UE k in cell l and BS j:

$$\mathbf{h}_{lk}^{j}$$

  - This is an $M_j \times 1$ vector

Examples of Antenna Array Geometries

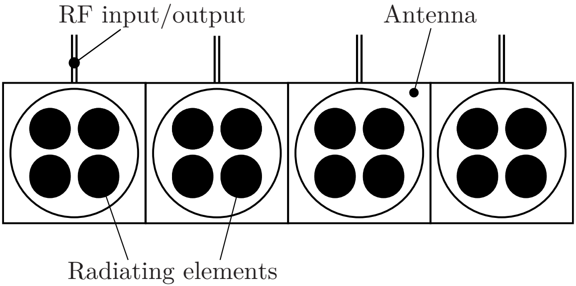


a) linear vertical; b) linear horizontal; c) planar; d) cylindrical.

- Deployment strategies
  - One or multiple cell sectors
  - One or multiple arrays per cell

- Massive in numbers, not in size
  - BSs in LTE have hundreds of radiating elements, but few RF chains
  - Novelty: Every radiating element is an antenna with an RF chain

Radiating Element, Antenna, and Antenna Array

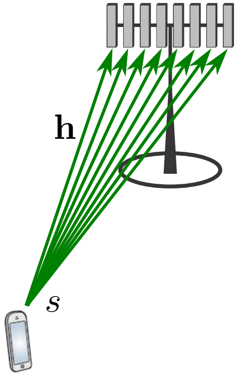


Definition (Radiating element, antenna, antenna array)

An *antenna* consists of one or more *radiating elements* (e.g., dipoles) which are fed by the same RF signal. An *antenna array* is composed of multiple antennas with individual RF chains.

CSI, Coherence block, TDD...

Example: Uplink Channel Estimation



- The UE sends a single pilot signal $s \in \mathbb{C}$ that is known at the BS

$$\mathbf{y} = \mathbf{h}s + \mathbf{n}$$

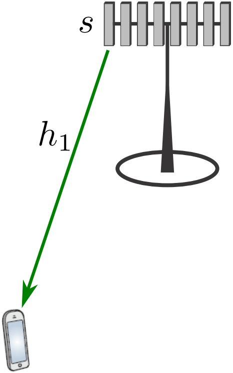
- Simple estimate of $\mathbf{h}$:

$$\hat{\mathbf{h}} = \frac{s^\star}{|s|^2}\mathbf{y}$$

In the uplink, the channel vector to an unlimited number of antennas can be learned from a single pilot transmission!

If there are K single-antenna UEs, then K pilot signals are required!

Example: Downlink Channel Estimation



- The BS sends a known pilot signal s subsequently from each antenna
- Received signal at the UE:

$$y_m = h_m s + n_m \quad m = 1, \ldots, M$$

- Simple estimate of h_m:

$$\hat{h}_m = \frac{s^\star}{|s|^2} y_m$$

- The UE feeds $\hat{\mathbf{h}}$ back to the BS[3]

M pilot transmissions (plus feedback) are needed to estimate the downlink channel!

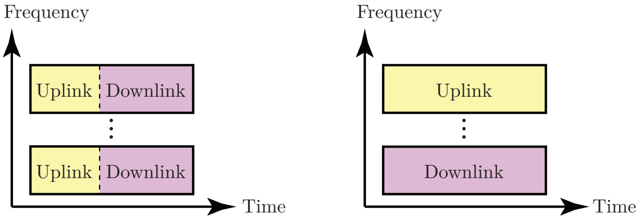
[3] Generally, a quantized version of $\hat{\mathbf{h}}$ is fed back which increases the estimation error.

Channel Coherence Block

Definition (Coherence block)

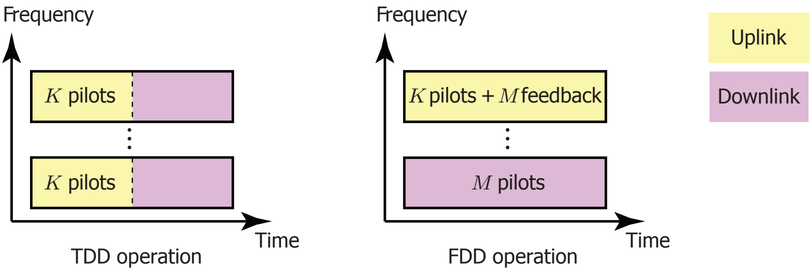
A coherence block consists of a number of subcarriers and time samples over which the channel response is approximately constant and flat-fading. If the coherence bandwidth is B_c and the coherence time is T_c, each coherence block contains $\tau_c = B_c T_c$ complex-valued samples.

- T_c and B_c depend on carrier frequency, UE speed, delay spread, etc.
- Typical values for T_c and B_c are in the range from 1–50 ms and 0.2–1 MHz: a coherence block contains 200–50000 samples



Different ways to assign UL and DL to coherence blocks

Overhead of CSI Acquisition



- Time-division duplex (TDD) — Overhead per block: K pilots
  - UL/DL channels are reciprocal
  - Only BS needs to know full channels

- Frequency-division duplex (FDD) — Overhead per block: $M + \frac{K}{2}$
  - K pilots + M feedback in UL
  - M pilots in DL

Feasible Operating Points

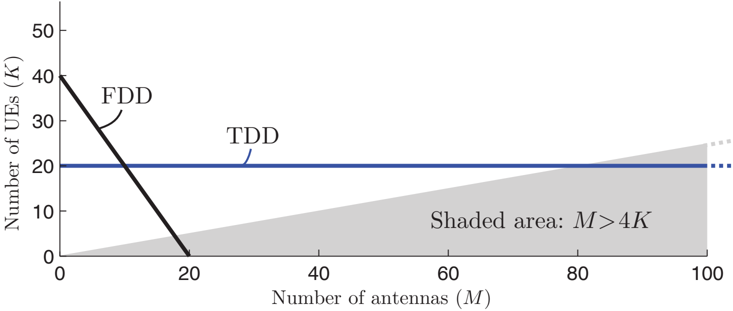


Illustration of operating points (M, K) supported by using $\tau_p = 20$ pilots, for different TDD and FDD protocols. The shaded area corresponds to operating points that are preferable in SDMA systems.

Only TDD and the resulting channel reciprocity allow for very large M!

Channel Parameterizations

In some propagation scenarios, the M-dimensional channels can be parameterized using much less than M parameters.

- Key example: LoS propagation
    - Mainly depends on the angle between the BS and the UE
- Instead of transmitting M DL pilots, select a set of equally spaced angles and send precoded DL pilot signals only in these directions
- If the number of angles is much smaller than M, then this method can enable FDD operation with potentially good estimation quality

But...

- LoS channel parameterizations depends on array geometry
- UE channels are likely a mixture of NLoS and LoS components

TDD operates efficiently in any kind of propagation environment!

Spatial Channel Correlation

What is Spatial Channel Correlation?

Definition (Spatial Channel Correlation)

A fading channel $\mathbf{h} \in \mathbb{C}^M$ is *spatially uncorrelated* if the channel gain $\|\mathbf{h}\|^2$ and the channel direction $\mathbf{h}/\|\mathbf{h}\|$ are independent random variables, and the channel direction is uniformly distributed over the unit-sphere in $\mathbb{C}^M$. The channel is otherwise *spatially correlated*.

Example of uncorrelated channel:

- Uncorrelated Rayleigh fading: $\mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \beta\mathbf{I})$
- All eigenvalues of correlation matrix are equal

Example of correlated channel:

- Any model with eigenvalue variations in the correlation matrix
  - Some spatial directions are statistically more likely to contain strong signal components than others
- Correlated Rayleigh fading: $\mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R})$
- More correlation: Larger eigenvalue variations

The Correlated Rayleigh Fading Channel Model

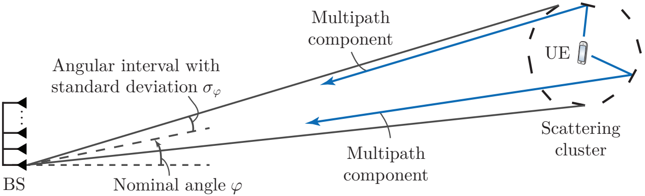
Definition (Correlated Rayleigh Fading)

Under the correlated Rayleigh fading channel model, the channel vectors $\mathbf{h}_{lk}^{j} \in \mathbb{C}^{M_j}$ are distributed as $\mathbf{h}_{lk}^{j} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_{M_j}, \mathbf{R}_{lk}^{j}\right)$, where $\mathbf{R}_{lk}^{j} \in \mathbb{C}^{M_j \times M_j}$ is the spatial channel correlation matrix.

- $\mathbf{h}_{lk}^{j}$ takes independent realizations in every coherence block
- Variations in $\mathbf{h}_{lk}^{j}$ describe *microscopic* effects due to movement
- $\mathbf{R}_{lk}^{j}$ is assumed to be known[4] at BS j
- The eigenvalues and eigenvectors of $\mathbf{R}_{lk}^{j}$ determine the *spatial channel correlation* of $\mathbf{h}_{lk}^{j}$
- Average channel gain is $\beta_{lk}^{j} = \frac{1}{M_j}\mathrm{tr}(\mathbf{R}_{lk}^{j})$ per antenna

[4] Estimation of $\mathbf{R}_{lk}^{j}$ is a very important topic, but will not be covered in this course.

Local Scattering Correlation Model



- NLoS channel between a UE and a uniform linear array (ULA)
  - Nominal angle φ

$$[\mathbf{R}]_{l,m} = \beta \int e^{2\pi \mathrm{j} d_{\mathrm{H}}(l-m)\sin(\bar{\varphi})} f(\bar{\varphi}) d\bar{\varphi} \quad , 1 \leq l, m \leq M$$

- Can be numerically computed for any angle distribution $f(\bar{\varphi})$

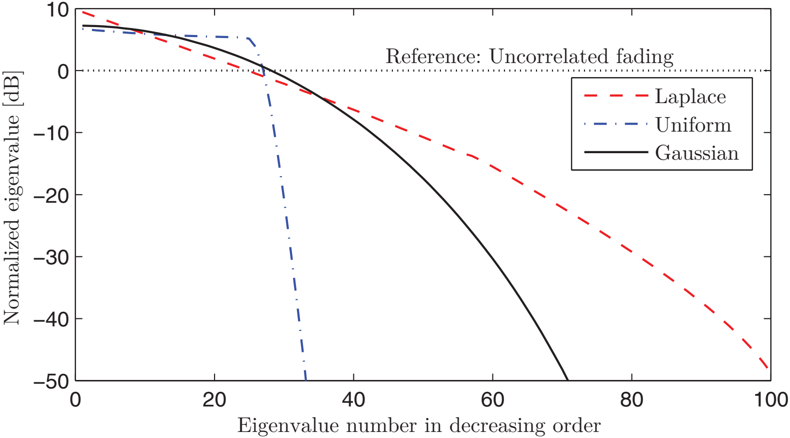
- Local scattering model: $\bar{\varphi} = \varphi + \Delta$ with only small Δ.
  - Several distributions of Δ in the literature:

    $\Delta \sim \mathcal{N}(0, \sigma_\varphi^2)$ (Normal distribution)

    $\Delta \sim \mathrm{Lap}(0, \sigma_\varphi/\sqrt{2})$ (Laplace distribution)

    $\Delta \sim U[-\sqrt{3}\sigma_\varphi, \sqrt{3}\sigma_\varphi]$ (Uniform distribution)

Local Scattering Correlation Model: Eigenvalue Distribution



$M = 100$, $\varphi = 30^\circ$, $\sigma_\varphi = 10^\circ$

Channel Hardening and Favorable Propagation

Channel Hardening (1/2)

Definition (Channel hardening)

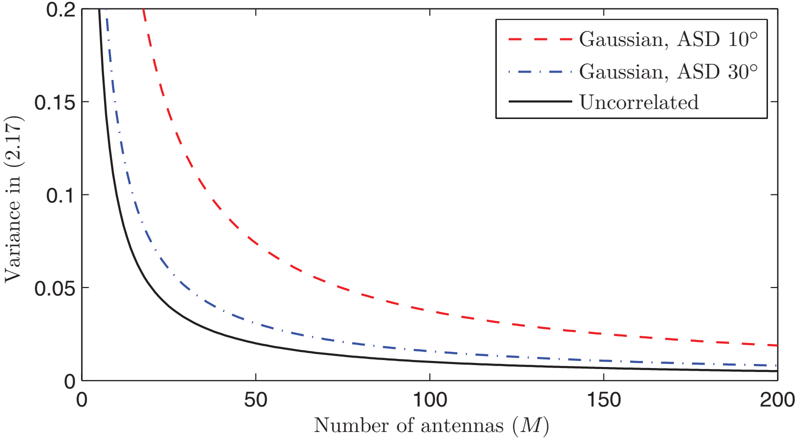
A propagation channel $\mathbf{h}^j_{jk}$ provides asymptotic channel hardening if

$$\frac{\|\mathbf{h}^j_{jk}\|^2}{\mathbb{E}\{\|\mathbf{h}^j_{jk}\|^2\}} \to 1 \quad \text{almost surely as } M_j \to \infty.$$

- Channel gain $\|\mathbf{h}^j_{jk}\|^2$ is close to its mean value $\mathbb{E}\{\|\mathbf{h}^j_{jk}\|^2\}$
  - Implies that fading has little impact on communication performance
  - Does not imply that $\|\mathbf{h}^j_{jk}\|^2$ becomes deterministic
- For uncorrelated fading, this follows from the law of large numbers
- For finite M_j and correlated fading, we want a small value of

$$\mathbb{V}\left\{\frac{\|\mathbf{h}^j_{jk}\|^2}{\mathbb{E}\{\|\mathbf{h}^j_{jk}\|^2\}}\right\} = \frac{\mathrm{tr}\left((\mathbf{R}^j_{jk})^2\right)}{(M_j \beta^j_{lk})^2} \tag{2.17}$$

Channel Hardening (2/2)



Variance of the channel hardening metric

Uncorrelated fading compared with local scattering model ($\varphi = 30^\circ$)

Spatial correlation leads to less channel hardening

Favorable Propagation (1/2)

Definition (Favorable propagation)

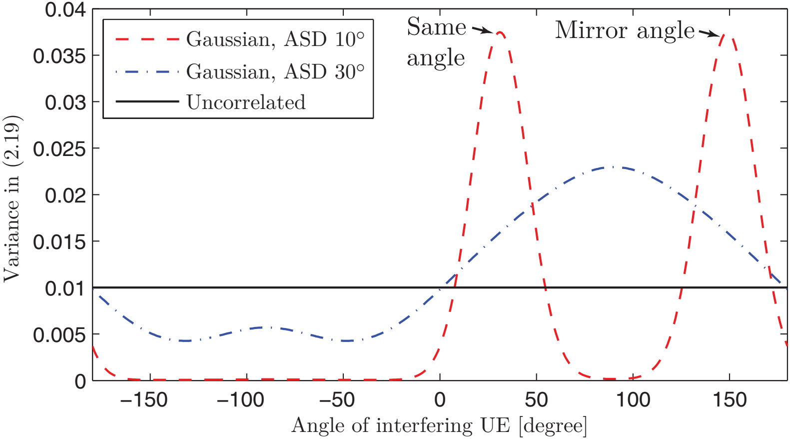
The pair of channels $\mathbf{h}_{li}^{j}$ and $\mathbf{h}_{jk}^{j}$ to BS j provide asymptotically favorable propagation if

$$\frac{(\mathbf{h}_{li}^{j})^{\mathrm{H}}\mathbf{h}_{jk}^{j}}{\sqrt{\mathbb{E}\{\|\mathbf{h}_{li}^{j}\|^2\}\mathbb{E}\{\|\mathbf{h}_{jk}^{j}\|^2\}}} \to 0 \quad \text{almost surely as } M_j \to \infty.$$

- Channel directions become orthogonal asymptotically
  - Implies less interference between the UEs
  - Does not imply that $(\mathbf{h}_{li}^{j})^{\mathrm{H}}\mathbf{h}_{jk}^{j} \to 0$
- For uncorrelated fading, this follows from the law of large numbers
- For finite M_j and correlated fading, we want a small value of

$$\mathbb{V}\left\{\frac{(\mathbf{h}_{li}^{j})^{\mathrm{H}}\mathbf{h}_{jk}^{j}}{\sqrt{\mathbb{E}\{\|\mathbf{h}_{li}^{j}\|^2\}\mathbb{E}\{\|\mathbf{h}_{jk}^{j}\|^2\}}}\right\} = \frac{\mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{R}_{jk}^{j}\right)}{M_j^2\beta_{li}^{j}\beta_{jk}^{j}} \tag{2.19}$$

Favorable Propagation (2/2)



Variance of the favorable propagation metric

Uncorrelated fading compared with local scattering model (desired UE: $\varphi = 30^\circ$)

Depends strongly on the UEs' correlation matrices

Physical Limits of Large Arrays

The channel hardening and favorable propagation phenonema have been validated experimentally for practical antenna numbers [GERT11, HHWtB12]...

- Physics prevent us from letting $M \to \infty$ and collecting more energy than was transmitted.
- This is not an issue when we deal with hundreds or thousands of antennas, since a "small" channel gain of -60 dB in cellular communications requires $M = 10^6$ to collect all power.

In conclusion...

The limit $M \to \infty$ is not physically achievable, but it is an analytical tool to explain what happens at practically large antenna numbers

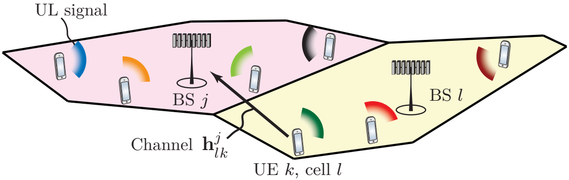
Five Differences Between Multiuser MIMO and Massive MIMO

- Massive MIMO is a refined form of multiuser MIMO
  - Has its roots in the 1980s [Win87] and 1990s [SBEM90, AMVW91].

	Multiuser MIMO	**Massive MIMO**
M_j and K_j	$M \approx K$ and both are small (e.g., < 10)	$M \gg K$ and typically large (e.g., $M = 100$, $K = 20$).
Duplexing	Designed to work in both TDD and FDD	Designed for TDD and exploits channel reciprocity
CSI acquisition	Mainly based on codebooks with set of predefined angular beams	Based on sending uplink pilots and exploiting channel reciprocity
Link quality	Varies rapidly due to frequency-selective and small-scale fading	Small variations over time and frequency, thanks to channel hardening
Resource allocation	Changes rapidly due to link quality variations	Can be planned since the link quality varies slowly

Uplink System Model

Uplink Transmission



Received UL signal $\mathbf{y}_j \in \mathbb{C}^{M_j}$ at BS j:

$$\mathbf{y}_j = \underbrace{\sum_{k=1}^{K_j} \mathbf{h}^j_{jk} s_{jk}}_{\text{Desired signals}} + \underbrace{\sum_{\substack{l=1 \\ l\neq j}}^{L} \sum_{i=1}^{K_l} \mathbf{h}^j_{li} s_{li}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{n}_j}_{\text{Noise}}$$

- UL signal of UE k in cell l: $s_{lk} \in \mathbb{C}$ with $p_{lk} = \mathbb{E}\{|s_{lk}|^2\}$, irrespective of whether it is a random payload data signal $s_{lk} \sim \mathcal{N}_{\mathbb{C}}(0, p_{lk})$ or a deterministic pilot signal with $p_{lk} = |s_{lk}|^2$
- Receiver noise: $\mathbf{n}_j \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{M_j}, \sigma^2_{\mathrm{UL}} \mathbf{I}_{M_j})$

Linear Receive Combining in the Uplink

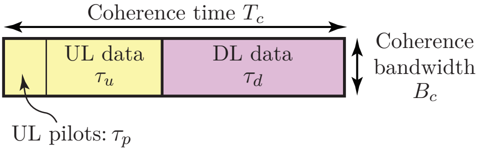
During payload transmission, the BS in cell j uses the *receive combining vector*[5] $\mathbf{v}_{jk} \in \mathbb{C}^{M_j}$ to separate the signal from its kth desired UE from the interference as

$$\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{y}_j = \underbrace{\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jk}^{j}s_{jk}}_{\text{Desired signal}} + \underbrace{\sum_{\substack{i=1 \\ i\neq k}}^{K_j}\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{ji}^{j}s_{ji}}_{\text{Intra-cell signals}} + \underbrace{\sum_{\substack{l=1 \\ l\neq j}}^{L}\sum_{i=1}^{K_l}\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{li}^{j}s_{li}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{n}_j}_{\text{Noise}}$$

The selection of combining (and precoding) vectors, based on estimated channels, and the corresponding SEs will be discussed in depth later

[5]Linear receive combining is also known as linear detection

Received Uplink Signal During Pilot Transmission



Received UL signal $\mathbf{Y}_j^p \in \mathbb{C}^{M_j \times \tau_p}$ at BS j:

$$\mathbf{Y}_j^p = \underbrace{\sum_{k=1}^{K_j} \sqrt{p_{jk}}\mathbf{h}_{jk}^j \boldsymbol{\phi}_{jk}^{\mathrm{T}}}_{\text{Desired pilots}} + \underbrace{\sum_{\substack{l=1 \\ l \neq j}}^{L} \sum_{i=1}^{K_l} \sqrt{p_{li}}\mathbf{h}_{li}^j \boldsymbol{\phi}_{li}^{\mathrm{T}}}_{\text{Inter-cell pilots}} + \underbrace{\mathbf{N}_j^p}_{\text{Noise}}$$

- UE k in cell j transmits the *pilot sequence* $\boldsymbol{\phi}_{jk} \in \mathbb{C}^{\tau_p}$
- $\|\boldsymbol{\phi}_{jk}\|^2 = \boldsymbol{\phi}_{jk}^{\mathrm{H}} \boldsymbol{\phi}_{jk} = \tau_p$ (scaled by UE's transmit power as $\sqrt{p_{jk}}$)
- $\mathbf{N}_j^p \in \mathbb{C}^{M_j \times \tau_p}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{UL}}^2)$ elements

Pilot Book and Pilot Allocation

- BS j correlates $\mathbf{Y}_j^p$ with $\boldsymbol{\phi}_{jk}$ to estimate $\mathbf{h}_{jk}^j$.
- The network uses τ_p mutually orthogonal UL pilot sequences
- These sequences form the *pilot book* $\boldsymbol{\Phi}^u \in \mathbb{C}^{\tau_p \times \tau_p}$:

$$(\boldsymbol{\Phi}^u)^{\mathrm{H}} \boldsymbol{\Phi}^u = \tau_p \mathbf{I}_{\tau_p}$$

- If $\tau_p \geq \max_l K_l$, each BS can allocate a different pilot to each UE
- Define the set of UEs utilizing the same pilot as UE k in cell j:

$$\mathcal{P}_{jk} = \left\{(l,i) : \boldsymbol{\phi}_{li} = \boldsymbol{\phi}_{jk}, \quad l = 1, \ldots, L, i = 1, \ldots, K_l\right\}$$

This leads to the simplified expression:

$$\mathbf{y}_{jjk}^p = \mathbf{Y}_j^p \boldsymbol{\phi}_{jk}^\star = \underbrace{\sqrt{p_{jk}}\tau_p \mathbf{h}_{jk}^j}_{\text{Desired pilot}} + \underbrace{\sum_{(l,i)\in\mathcal{P}_{jk}\setminus(j,k)} \sqrt{p_{li}}\tau_p \mathbf{h}_{li}^j}_{\text{Interfering pilots}} + \underbrace{\mathbf{N}_j^p \boldsymbol{\phi}_{jk}^\star}_{\text{Noise}}$$

where $\mathbf{N}_j^p \boldsymbol{\phi}_{jk}^\star \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{M_j}, \sigma_{\mathrm{UL}}^2 \tau_p \mathbf{I}_{M_j})$ since $\boldsymbol{\phi}_{jk}$ is deterministic

MMSE Channel Estimation

MMSE Channel Estimation

Theorem

The MMSE estimate of $\mathbf{h}_{li}^{j}$ based on the observation $\mathbf{Y}_{j}^{p}$ at BS j is

$$\hat{\mathbf{h}}_{li}^{j} = \sqrt{p_{li}}\mathbf{R}_{li}^{j}\mathbf{\Psi}_{li}^{j}\mathbf{y}_{jli}^{p}$$

where $\mathbf{\Psi}_{li}^{j} = \left(\sum_{(l',i')\in\mathcal{P}_{li}} p_{l'i'}\tau_p\mathbf{R}_{l'i'}^{j} + \sigma_{\mathrm{UL}}^{2}\mathbf{I}_{M_j}\right)^{-1}$.

The estimation error $\tilde{\mathbf{h}}_{li}^{j} = \mathbf{h}_{li}^{j} - \hat{\mathbf{h}}_{li}^{j}$ has the correlation matrix

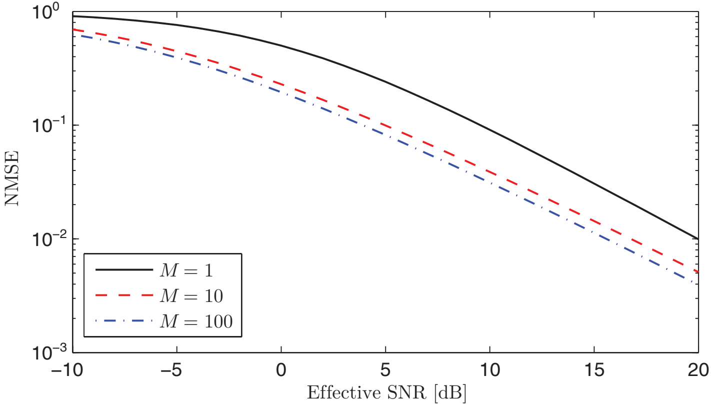
$$\mathbf{C}_{li}^{j} = \mathbb{E}\{\tilde{\mathbf{h}}_{li}^{j}(\tilde{\mathbf{h}}_{li}^{j})^{\mathrm{H}}\} = \mathbf{R}_{li}^{j} - p_{li}\tau_p\mathbf{R}_{li}^{j}\mathbf{\Psi}_{li}^{j}\mathbf{R}_{li}^{j}.$$

Corollary

The estimate $\hat{\mathbf{h}}_{li}^{j}$ and the estimation error $\tilde{\mathbf{h}}_{li}^{j}$ are independent random variables, distributed as follows:

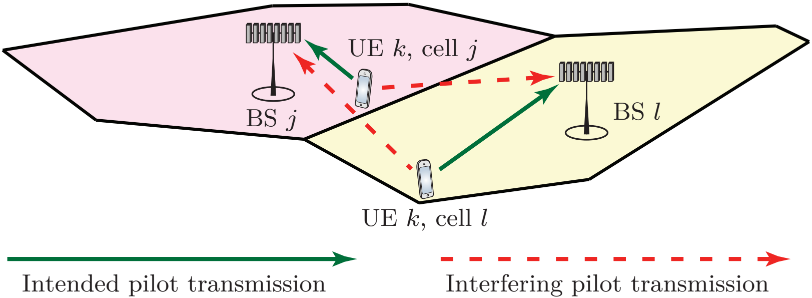
$$\hat{\mathbf{h}}_{li}^{j} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_{M_j}, \mathbf{R}_{li}^{j} - \mathbf{C}_{li}^{j}\right), \quad \tilde{\mathbf{h}}_{li}^{j} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_{M_j}, \mathbf{C}_{li}^{j}\right).$$

Impact of SNR on Estimation Quality



- One UE with effective SNR $p_{jk}\tau_p\beta_{jk}/\sigma^2_{\mathrm{UL}}$
  - Processing gain: SNR grows with τ_p
- Normalized MSE (NMSE): $\mathrm{tr}(\mathbf{C}^j_{jk})/\mathrm{tr}(\mathbf{R}^j_{jk}) \in [0, 1]$
- Local scattering channel model, Gaussian distribution ($\sigma_\varphi = 10^\circ$)
  - NMSE decays with M: Easier to estimate correlated channels

Example of Interfering Pilot Transmissions



$$\mathbf{y}_{jjk}^{p} = \underbrace{\sqrt{p_{jk}}\tau_p \mathbf{h}_{jk}^{j}}_{\text{Desired pilot}} + \underbrace{\sqrt{p_{lk}}\tau_p \mathbf{h}_{lk}^{j}}_{\text{Interfering pilot}} + \underbrace{\mathbf{N}_j^p \boldsymbol{\phi}_{jk}^{\star}}_{\text{Noise}}$$

Pilot Contamination

Corollary

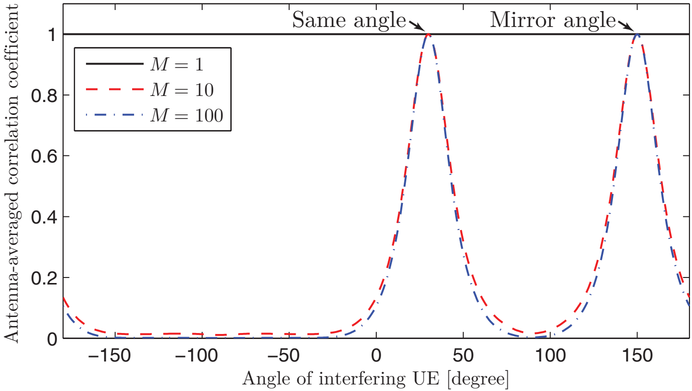
Consider UE k in cell j and UE i in cell l. It holds that

$$\frac{\mathbb{E}\{(\hat{\mathbf{h}}_{li}^{j})^{\mathrm{H}}\hat{\mathbf{h}}_{jk}^{j}\}}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}_{jk}^{j}\|^2\}\mathbb{E}\{\|\hat{\mathbf{h}}_{li}^{j}\|^2\}}} = \begin{cases} \frac{\mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{R}_{jk}^{j}\boldsymbol{\Psi}_{li}^{j}\right)}{\sqrt{\mathrm{tr}\left(\mathbf{R}_{jk}^{j}\mathbf{R}_{jk}^{j}\boldsymbol{\Psi}_{li}^{j}\right)\mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{R}_{li}^{j}\boldsymbol{\Psi}_{li}^{j}\right)}} & (l,i) \in \mathcal{P}_{jk} \\ 0 & (l,i) \notin \mathcal{P}_{jk} \end{cases}$$

despite the fact that $\mathbb{E}\left\{(\mathbf{h}_{li}^{j})^{\mathrm{H}}\mathbf{h}_{jk}^{j}\right\}/M_j = 0$ for all UE combinations with $(l,i) \neq (j,k)$.

- This corollary describes the phenomenon of *pilot contamination*
- Interfering UEs reduce estimation quality, but also *makes channel estimates statistically dependent*, despite the independent channels
- Less contamination if $\mathbf{R}_{li}^{j}\mathbf{R}_{jk}^{j}$ is small
  - Large pathloss difference or different supports.
- Pilot contamination makes it harder for the BS to mitigate interference between UEs that use the same pilot sequence

Pilot Contamination: Numerical Results



- Two UEs using the same pilot sequence
  - Desired UE has nominal angle of $30°$
  - Nominal angle of undesired UE varies from $-180°$ to $180°$
  - Local scattering channel model with Gaussian angular distribution
  - 10 dB SNR to desired UE, 0 dB to interfering UE

Pilot Contamination: Additional Remarks

- Pilot contamination exists because of the practical necessity to reuse the time-frequency resources across cells
- It is often described as a main characteristic of Massive MIMO [Mar10, GJ11, JAMV11], but it is not unique for Massive MIMO

Pilot contamination has a greater impact on Massive MIMO than on conventional systems because the aggressive spatial multiplexing requires more frequent spatial reuse of pilot sequences

- The eigenstructure of the spatial correlation matrices determines the strength of the pilot contamination
- Pilot sequence assignment to UEs with very "different" correlation matrices can hence help reduce this effect, e.g., [HCPR12, YGFL13]

Channel Estimation: Key Points

- Channel estimation based on **UL pilot sequences** is key
  - One orthogonal sequence per UE in the cell
  - Effective SNR is proportional to pilot length
- **MMSE estimation** uses channel statistics to obtain good estimates
  - Alternatives: Element-wise MMSE, least-square, data-aided
- Limited channel coherence makes **pilot reuse** across cells necessary:
  - Inter-cell interference **reduces estimation quality**
  - Channel estimates of UEs that use the same pilot are correlated; phenomenon called **pilot contamination**
  - Correlation small for UEs with sufficiently different correlation matrices; differences in large pathloss or spatial characteristics
  - Pilot contamination lead to **coherent interference**, hard to mitigate

Uplink Spectral Efficiency

Received Uplink Signal with Estimated Channels

The BS in cell j decodes UE k's signal s_{jk} based on:

$$\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{y}_j = \mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jk}^{j}s_{jk} + \underbrace{\sum_{l=1}^{L}\sum_{\substack{i=1\\(l,i)\neq(j,k)}}^{K_l}\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{li}^{j}s_{li} + \mathbf{v}_{jk}^{\mathrm{H}}\mathbf{n}_j}_{\text{Interference plus noise}}$$

Using the MMSE estimator, all channels can be decomposed as

$$\mathbf{h}_{li}^{j} = \underbrace{\hat{\mathbf{h}}_{li}^{j}}_{\text{Known}} + \underbrace{\tilde{\mathbf{h}}_{li}^{j}}_{\text{Unknown}}$$

Thus,

$$\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{y}_j = \underbrace{\mathbf{v}_{jk}^{\mathrm{H}}\hat{\mathbf{h}}_{jk}^{j}s_{jk}}_{\text{Desired signal over known channel}} + \underbrace{z_{jk}}_{\text{Everything else}}$$

An Achievable Uplink Spectral Efficiency

Theorem

If MMSE channel estimation is used, then the UL channel capacity of UE k in cell j is lower bounded by $\mathsf{SE}_{jk}^{\mathrm{UL}}$ [bit/s/Hz] given by

$$\mathsf{SE}_{jk}^{\mathrm{UL}} = \frac{\tau_u}{\tau_c} \mathbb{E}\left\{ \log_2 \left(1 + \mathsf{SINR}_{jk}^{\mathrm{UL}} \right) \right\}$$

with *instantaneous* SINR

$$\mathsf{SINR}_{jk}^{\mathrm{UL}} = \frac{p_{jk} |\mathbf{v}_{jk}^{\mathrm{H}} \hat{\mathbf{h}}_{jk}^{j}|^2}{\sum\limits_{l=1}^{L} \sum\limits_{\substack{i=1 \\ (l,i)\neq(j,k)}}^{K_l} p_{li} |\mathbf{v}_{jk}^{\mathrm{H}} \hat{\mathbf{h}}_{li}^{j}|^2 + \mathbf{v}_{jk}^{\mathrm{H}} \left(\sum\limits_{l=1}^{L} \sum\limits_{i=1}^{K_l} p_{li} \mathbf{C}_{li}^{j} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right) \mathbf{v}_{jk}}$$

and where the expectation is with respect to the channel estimates.

- The prelog factor arises because only a fraction $\frac{\tau_u}{\tau_c}$ of all samples are used for UL data transmission
- The result holds for any receive combining vector $\mathbf{v}_{jk}$

The Optimal Receive Combining Vector

Corollary: Multicell MMSE (M-MMSE) Combining Vector

$\mathsf{SINR}_{jk}^{\mathrm{UL}}$ is maximized by the combining vector

$$\mathbf{v}_{jk} = p_{jk} \left(\sum_{l=1}^{L} \sum_{i=1}^{K_l} p_{li} \left(\hat{\mathbf{h}}_{li}^{j} (\hat{\mathbf{h}}_{li}^{j})^{\mathrm{H}} + \mathbf{C}_{li}^{j} \right) + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right)^{-1} \hat{\mathbf{h}}_{jk}^{j}$$

which leads to

$$\mathsf{SINR}_{jk}^{\mathrm{UL}} = p_{jk} (\hat{\mathbf{h}}_{jk}^{j})^{\mathrm{H}} \left(\sum_{l=1}^{L} \sum_{\substack{i=1 \\ (l,i)\neq(j,k)}}^{K_l} p_{li} \hat{\mathbf{h}}_{li}^{j} (\hat{\mathbf{h}}_{li}^{j})^{\mathrm{H}} + \sum_{l=1}^{L} \sum_{i=1}^{K_l} p_{li} \mathbf{C}_{li}^{j} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right)^{-1} \hat{\mathbf{h}}_{jk}^{j}.$$

Remark

The M-MMSE combining vector minimizes the conditional MSE

$$\mathbb{E} \left\{ \left| s_{jk} - \mathbf{v}_{jk}^{\mathrm{H}} \mathbf{y}_j \right|^2 \middle| \{ \hat{\mathbf{h}}_{li}^{j} \} \right\}.$$

Other Combining Schemes

$$\mathbf{V}_j^{\text{M-MMSE}} = \left(\sum_{l=1}^{L} \hat{\mathbf{H}}_l^j \mathbf{P}_l (\hat{\mathbf{H}}_l^j)^{\mathrm{H}} + \sum_{l=1}^{L} \sum_{i=1}^{K_l} p_{li} \mathbf{C}_{li}^j + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right)^{-1} \hat{\mathbf{H}}_j^j \mathbf{P}_j$$

Single-cell MMSE (S-MMSE):

$$\mathbf{V}_j^{\text{S-MMSE}} = \left(\hat{\mathbf{H}}_j^j \mathbf{P}_j (\hat{\mathbf{H}}_j^j)^{\mathrm{H}} + \sum_{i=1}^{K_j} p_{ji} \mathbf{C}_{ji}^j + \sum_{\substack{l=1 \\ l \neq j}}^{L} \sum_{i=1}^{K_l} p_{li} \mathbf{R}_{li}^j + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right)^{-1} \hat{\mathbf{H}}_j^j \mathbf{P}_j$$

Regularized Zero-Forcing (RZF):

$$\mathbf{V}_j^{\mathrm{RZF}} = \left(\hat{\mathbf{H}}_j^j \mathbf{P}_j (\hat{\mathbf{H}}_j^j)^{\mathrm{H}} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right)^{-1} \hat{\mathbf{H}}_j^j \mathbf{P}_j = \hat{\mathbf{H}}_j^j \left((\hat{\mathbf{H}}_j^j)^{\mathrm{H}} \hat{\mathbf{H}}_j^j + \sigma_{\mathrm{UL}}^2 \mathbf{P}_j^{-1} \right)^{-1}$$

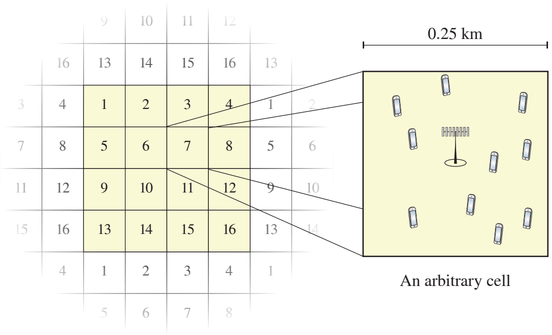
Zero-Forcing (ZF):

$$\mathbf{V}_j^{\mathrm{ZF}} = \hat{\mathbf{H}}_j^j \left((\hat{\mathbf{H}}_j^j)^{\mathrm{H}} \hat{\mathbf{H}}_j^j \right)^{-1}$$

Maximum Ratio (MR):

$$\mathbf{V}_j^{\mathrm{MR}} = \hat{\mathbf{H}}_j^j$$

Running Example: Geometry



- 16 cells in square pattern (with wrap-around)
  - M antennas per BS, K users randomly deployed per cell
  - Large-scale fading coefficient β_{lk}^{j} for UE at distance d_{lk}^{j} is[6]

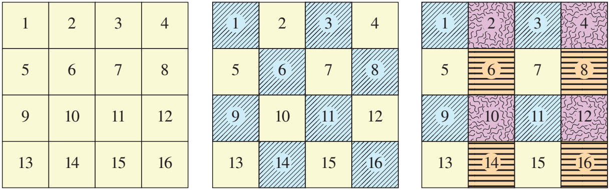
$$\beta_{lk}^{j}\,[\mathrm{dB}] = \Upsilon - 10\alpha \cdot \log_{10}\left(\frac{d_{lk}^{j}}{1\,\mathrm{km}}\right) + F_{lk}^{j}$$

with $\Upsilon = -148.1\,\mathrm{dB}$, $\alpha = 3.76$, $F_{lk}^{j} \sim \mathcal{N}(0, 7^2)$

[6] Remember $\beta_{lk}^{j} = M_j^{-1}\mathrm{tr}(\mathbf{R}_{lk}^{j})$. We make sure that $\beta_{jk}^{j} \geq \beta_{lk}^{j}$ for all l.

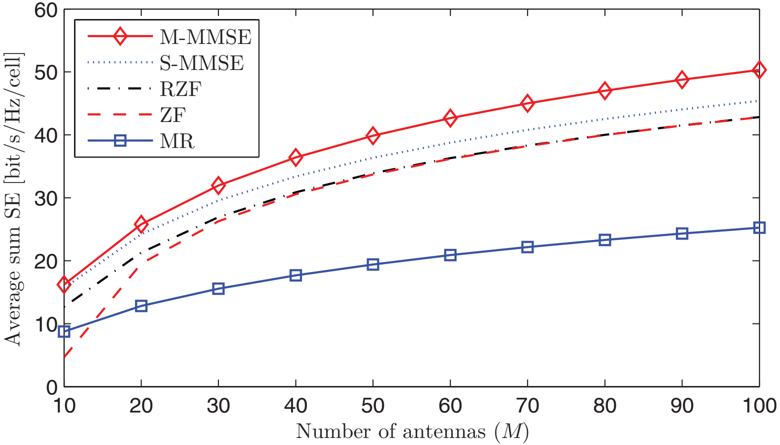
Running Example: Power and Pilot Reuse

- Bandwidth $B = 20$ MHz
  - UL/DL transmit power: 20 dBm per UE
  - Total noise power: -94 dBm
  - SNR: $20.5\,$dB (cell center), $-5.8\,$dB (cell corner), before shadowing
- Comparison of channel models
  - Gaussian local scattering: ASD σ_φ
  - Uncorrelated Rayleigh fading: $\mathbf{R}_{lk}^{j} = \beta_{lk}^{j}\mathbf{I}_M$
- Pilot reuse factor $f \in \{1, 2, 4\}$
  - $\tau_p = fK$ UL pilot sequences
  - K pilot sequences per cell, reused in $1/f$ of the cells



Pilot reuse f=1 Pilot reuse f=2 Pilot reuse f=4

Uplink SE Simulations: Universal Pilot Reuse



$K = 10$ UEs per cell, $\tau_c = 200$, $\tau_p = K$, $\tau_d = 0$ (UL only)

Gaussian local scattering model: ASD $\sigma_\varphi = 10^\circ$

LTE: for a TDD system, the UL is 2.8 bit/s/Hz/cell

Uplink SE Simulations: Insights ($M = 100$, $K = 10$)

Scheme	$f = 1$	$f = 2$	$f = 4$
M-MMSE	50.32	55.10	**55.41**
S-MMSE	45.39	**45.83**	42.41
RZF	42.83	**43.37**	39.99
ZF	42.80	**43.34**	39.97
MR	**25.25**	24.41	21.95

Average sum SE for different receive combining and pilot reuse factors

- Three schemes useful in practice:
  - M-MMSE: Highest SE, highest complexity
  - MR: Lowest SE, lowest complexity
  - RZF: Good balance between SE and complexity
- M-MMSE benefits most from $f > 1$
  (since improved channel estimation outweighs pre-log loss)
- MR does not gain from $f > 1$

Use-and-then-Forget (UatF) Bound

Theorem

The UL channel capacity of UE k in cell j is lower bounded by $\underline{\mathsf{SE}}_{jk}^{\mathrm{UL}} = \frac{\tau_u}{\tau_c} \log_2(1 + \underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}})$ [bit/s/Hz] with *effective* SINR

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}} = \frac{p_{jk} |\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2}{\sum_{l=1}^{L} \sum_{i=1}^{K_l} p_{li} \mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}} \mathbf{h}_{li}^{j}|^2\} - p_{jk} |\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2 + \sigma_{\mathrm{UL}}^2 \mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}}$$

where the expectations are with respect to the channel realizations.

- Less tight than previous bound
- Valid for any estimation and receive combining scheme[7]
- Each expectation can be computed separately
- Can allow for closed-form expressions

[7] It is also valid for any channel distribution!

UatF Bound for MR Combining

Lemma (UatF Bound for MR Combining)

If MR combining with $\mathbf{v}_{jk} = \hat{\mathbf{h}}_{jk}^{j}$ is used, then (nice exercise)

$$\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jk}^{j}\} = \mathbb{E}\{\|\mathbf{v}_{jk}\|^2\} = p_{jk}\tau_p \mathrm{tr}\left(\mathbf{R}_{jk}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)$$

$$\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{li}^{j}|^2\} = p_{jk}\tau_p \mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{R}_{jk}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right) + \begin{cases} p_{li}p_{jk}(\tau_p)^2\left|\mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)\right|^2 & (l,i) \in \mathcal{P}_{jk} \\ 0 & (l,i) \notin \mathcal{P}_{jk} \end{cases}$$

The SE expression becomes $\underline{\mathsf{SE}}_{jk}^{\mathrm{UL}} = \frac{\tau_u}{\tau_c}\log_2(1+\underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}})$ with

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}} = \frac{p_{jk}^2\tau_p \mathrm{tr}\left(\mathbf{R}_{jk}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)}{\sum\limits_{l=1}^{L}\sum\limits_{i=1}^{K_l} p_{li}\frac{\mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{R}_{jk}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)}{\mathrm{tr}\left(\mathbf{R}_{jk}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)} + \sum\limits_{(l,i)\in\mathcal{P}_{jk}\setminus(j,k)} \frac{p_{li}^2\tau_p\left|\mathrm{tr}\left(\mathbf{R}_{li}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)\right|^2}{\mathrm{tr}\left(\mathbf{R}_{jk}^{j}\mathbf{\Psi}_{jk}^{j}\mathbf{R}_{jk}^{j}\right)} + \sigma_{\mathrm{UL}}^2}$$

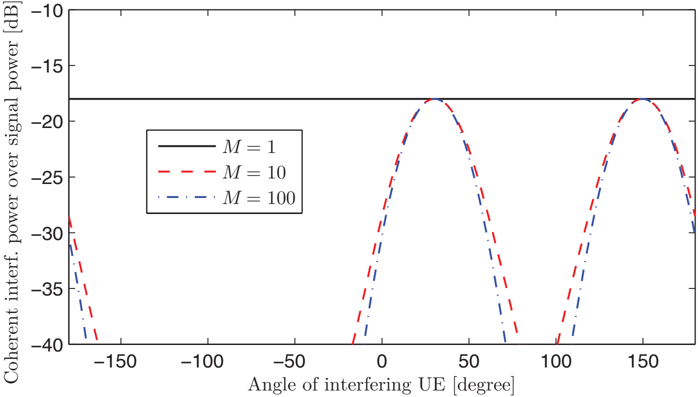
Insights from the UatF Bound with MR Combining

$$\frac{\overbrace{p_{jk}^2 \tau_p \mathrm{tr}\left(\mathbf{R}_{jk}^j \boldsymbol{\Psi}_{jk}^j \mathbf{R}_{jk}^j\right)}^{\text{Signal}}}{\underbrace{\sum_{l=1}^{L}\sum_{i=1}^{K_l} p_{li} \frac{\mathrm{tr}\left(\mathbf{R}_{li}^j \mathbf{R}_{jk}^j \boldsymbol{\Psi}_{jk}^j \mathbf{R}_{jk}^j\right)}{\mathrm{tr}\left(\mathbf{R}_{jk}^j \boldsymbol{\Psi}_{jk}^j \mathbf{R}_{jk}^j\right)}}_{\text{Non-coherent interference}} + \underbrace{\sum_{(l,i)\in\mathcal{P}_{jk}\setminus(j,k)} \frac{p_{li}^2 \tau_p \left|\mathrm{tr}\left(\mathbf{R}_{li}^j \boldsymbol{\Psi}_{jk}^j \mathbf{R}_{jk}^j\right)\right|^2}{\mathrm{tr}\left(\mathbf{R}_{jk}^j \boldsymbol{\Psi}_{jk}^j \mathbf{R}_{jk}^j\right)}}_{\text{Coherent interference}} + \underbrace{\sigma_{\mathrm{UL}}^2}_{\text{Noise}}}$$

- Signal $\sim M_j$ (trace of channel estimate's correlation matrix)
- Non-coherent interference + noise do not increase with M_j
- Coherent interference $\sim M_j$ (due to pilot contamination)

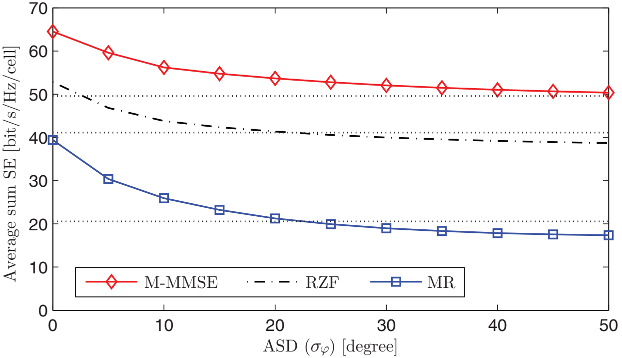
The relations between the correlation matrices $\mathbf{R}_{li}^j$ and $\mathbf{R}_{jk}^j$ determine the strength of the interference terms

Pilot Contamination: Coherent Interference



- Two UEs using the same pilot sequence
  - Desired UE has nominal angle of 30°
  - Nominal angle of undesired UE varies from -180° to 180°
  - Local scattering channel model with Gaussian angular distribution
  - 10 dB SNR to desired UE, 0 dB to interfering UE

Impact of Spatial Correlation

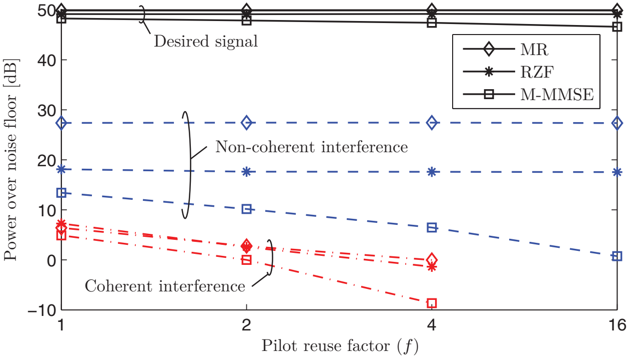


Local scattering channel model with varying ASD ($M = 100$, $K = 10$)[8]

Spatial channel correlation increases the sum SE since it reduces interference. For very small ASDs, the scenario is almost LoS.

[8] Dotted lines represent results for uncorrelated Rayleigh fading.

Coherent Interference at Strongest UE in the Cell

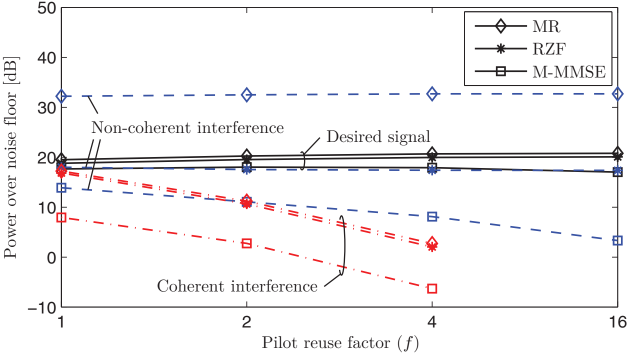


Gaussian local scattering model with $\sigma_\varphi = 10^\circ$, $M = 100$, $K = 10$

- Useful signal is barely affected by pilot contamination
- Coherent interference small compared to non-coherent interference[9]
- Coherent interference can be reduced by increasing f for any scheme
- M-MMSE can better suppress non-coherent interference for $f > 1$

[9] Picture can change depending on σ_φ

Coherent Interference at Weakest UE in the Cell



Gaussian local scattering model with $\sigma_\varphi = 10^\circ$, $M = 100$, $K = 10$

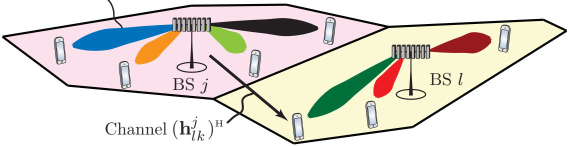
- Useful signal is much lower due to the additional pathloss
- Coherent interference comparable to non-coherent interference; with MR, 10 dB stronger than useful signal due to intra-cell interference
- All schemes can better suppress coherent interference for $f > 1$

Uplink Spectral Efficiency: Key Points

- **Lower bound on UL capacity** based on MMSE channel estimation
  - An achievable SE, maximized by M-MMSE combining
- **Combining schemes**: **M-MMSE**, S-MMSE, **RZF**, ZF, **MR**
- **Factors that affect SE**
  - Transmit powers
  - Pilot reuse factor
  - Spatial channel correlation
  - Pilot contamination
- **Insights from SE analysis** and running example
  - Received signal power and coherent interference linear in M
  - Non-coherent interference and noise independent of M
  - Coherent interference negligible for large pilot reuse factors
- **UatF bound** based on "average" channel:
  - Gives closed-form SE expressions with MR
  - Only tight with significant channel hardening

Downlink Spectral Efficiency

Linear Transmit Precoding in the Downlink



Received DL signal $y_{jk} \in \mathbb{C}$ at UE k in cell j:

$$y_{jk} = \underbrace{(\mathbf{h}_{jk}^{j})^{\mathrm{H}} \mathbf{w}_{jk} \varsigma_{jk}}_{\text{Desired signal}} + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{K_j} (\mathbf{h}_{jk}^{j})^{\mathrm{H}} \mathbf{w}_{ji} \varsigma_{ji}}_{\text{Intra-cell interference}} + \underbrace{\sum_{\substack{l=1 \\ l \neq j}}^{L} \sum_{i=1}^{K_l} (\mathbf{h}_{jk}^{l})^{\mathrm{H}} \mathbf{w}_{li} \varsigma_{li}}_{\text{Inter-cell interference}} + \underbrace{n_{jk}}_{\text{Noise}}$$

- BS l transmits the signal $\mathbf{x}_l = \sum_{i=1}^{K_l} \mathbf{w}_{li} \varsigma_{li}$
- Precoding vectors: $\mathbf{w}_{lk} \in \mathbb{C}^{M_l}$ with $\mathbb{E}\{\|\mathbf{w}_{lk}\|^2\} = 1$
- Data signals: $\varsigma_{lk} \sim \mathcal{N}_{\mathbb{C}}(0, \rho_{lk})$
- Receiver noise: $n_{jk} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{DL}}^2)$

Received Downlink Signal with Transmit Precoding

The UE k in cell j decodes its signal ς_{jk} based on:

$$y_{jk} = \underbrace{(\mathbf{h}_{jk}^{j})^{\mathrm{H}} \mathbf{w}_{jk} \varsigma_{jk}}_{\text{Desired signal}} + \underbrace{\sum_{l=1}^{L} \sum_{\substack{i=1 \\ (l,i) \neq (j,k)}}^{K_l} (\mathbf{h}_{jk}^{l})^{\mathrm{H}} \mathbf{w}_{li} \varsigma_{li} + n_{jk}}_{\text{Interference plus noise}}$$

- Efficient decoding requires:
  - Realization of precoded channel $(\mathbf{h}_{jk}^{j})^{\mathrm{H}} \mathbf{w}_{jk}$
  - Interference plus noise power $\sum_{(l,i) \neq (j,k)} |(\mathbf{h}_{jk}^{l})^{\mathrm{H}} \mathbf{w}_{li}|^2 \rho_{li} + \sigma_{\mathrm{DL}}^2$
- How to acquire this information?
  - Estimate current realizations from received DL signals
  - Exploit channel hardening

$$(\mathbf{h}_{jk}^{j})^{\mathrm{H}} \mathbf{w}_{jk} \approx \mathbb{E}\{(\mathbf{h}_{jk}^{j})^{\mathrm{H}} \mathbf{w}_{jk}\}$$

$$\sum_{(l,i) \neq (j,k)} |(\mathbf{h}_{jk}^{l})^{\mathrm{H}} \mathbf{w}_{li}|^2 \rho_{li} \approx \sum_{(l,i) \neq (j,k)} \mathbb{E}\{|(\mathbf{h}_{jk}^{l})^{\mathrm{H}} \mathbf{w}_{li}|^2\} \rho_{li}$$

A Downlink Spectral Efficiency (Hardening Bound)

Theorem

The DL channel capacity of UE k in cell j is lower bounded by $\underline{\mathsf{SE}}_{jk}^{\mathrm{DL}} = \frac{\tau_d}{\tau_c}\log_2(1+\underline{\mathsf{SINR}}_{jk}^{\mathrm{DL}})$ [bit/s/Hz] with *effective* SINR

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{DL}} = \frac{\rho_{jk}|\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}}\mathbf{h}_{jk}^{j}\}|^2}{\sum\limits_{l=1}^{L}\sum\limits_{i=1}^{K_l}\rho_{li}\mathbb{E}\{|\mathbf{w}_{li}^{\mathrm{H}}\mathbf{h}_{jk}^{l}|^2\} - \rho_{jk}|\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}}\mathbf{h}_{jk}^{j}\}|^2 + \sigma_{\mathrm{DL}}^2}$$

where the expectations are with respect to the channel realizations.

- The prelog factor $\frac{\tau_d}{\tau_c}$ is fraction of all samples used for DL data
- The result holds for any set of transmit precoding vectors $\{\mathbf{w}_{li}\}$
- Valid for any channel distribution and any estimation scheme
- Derived similarly to the UatF bound in UL

$\underline{\mathsf{SE}}_{jk}^{\mathrm{DL}}$ depends on all precoding vectors in entire network.
Not obvious how to design the precoding.

Insights from the SE Bound with MR

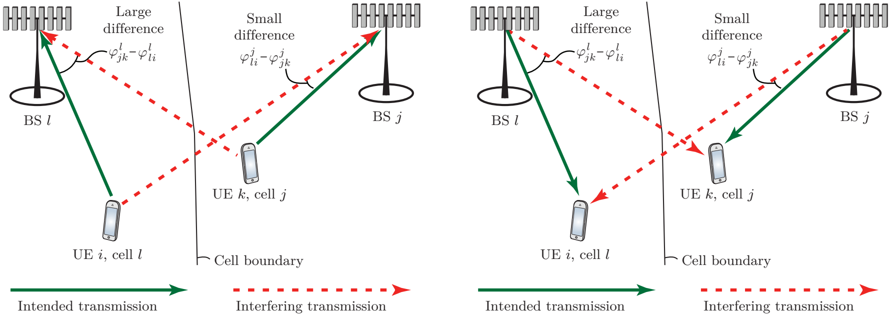
$$\frac{\overbrace{\rho_{jk} p_{jk} \tau_p \mathrm{tr}\left(\mathbf{R}_{jk}^{j} \mathbf{\Psi}_{jk}^{j} \mathbf{R}_{jk}^{j}\right)}^{\text{Signal}}}{\underbrace{\sum_{l=1}^{L} \sum_{i=1}^{K_l} \rho_{li} \frac{\mathrm{tr}\left(\mathbf{R}_{jk}^{l} \mathbf{R}_{li}^{l} \mathbf{\Psi}_{li}^{l} \mathbf{R}_{li}^{l}\right)}{\mathrm{tr}\left(\mathbf{R}_{li}^{l} \mathbf{\Psi}_{li}^{l} \mathbf{R}_{li}^{l}\right)}}_{\text{Non-coherent interference}} + \underbrace{\sum_{(l,i) \in \mathcal{P}_{jk} \setminus (j,k)} \rho_{li} \frac{p_{jk} \tau_p \left| \mathrm{tr}\left(\mathbf{R}_{jk}^{l} \mathbf{\Psi}_{li}^{l} \mathbf{R}_{li}^{l}\right)\right|^2}{\mathrm{tr}\left(\mathbf{R}_{li}^{l} \mathbf{\Psi}_{li}^{l} \mathbf{R}_{li}^{l}\right)}}_{\text{Coherent interference}} + \underbrace{\sigma_{\mathrm{DL}}^2}_{\text{Noise}}}$$

Similar interpretation as in uplink:

- Signal $\sim M_j$ (trace of channel estimate's correlation matrix)
- Non-coherent interference + noise do not increase with M_j
- Coherent interference $\sim M_l$ from BS l (due to pilot contamination)

The relations between the correlation matrices $\mathbf{R}_{li}^{l}$ and $\mathbf{R}_{jk}^{l}$ determine the strength of the interference terms

Different Correlation Matrices Affect DL and UL



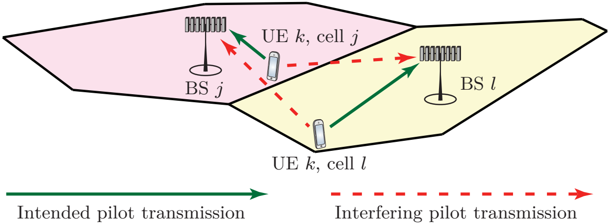
UE i in cell l interferes differently with UE k in cell j in the UL and DL

- Uplink: Interference comes directly from UE i in cell l
- Downlink: Interference comes from the BS in cell l

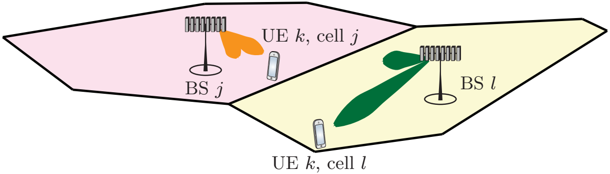
Different sets of correlation matrices affect UL and DL interference.

In the example, the UEs are well separated in angle in DL, but not in UL.

Pilot Contamination with MR Precoding



Two UEs transmit the same UL pilot sequence, causing the channel estimates at the respective BSs to be correlated



When a BS attempts to direct a signal towards its own UE using MR precoding, it will partially direct it towards the pilot-interfering UE in the other cell

Comparing Downlink and Uplink Expressions

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{DL}} = \frac{\rho_{jk} |\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2}{\sum_{l=1}^{L} \sum_{i=1}^{K_l} \rho_{li} \mathbb{E}\{|\mathbf{w}_{li}^{\mathrm{H}} \mathbf{h}_{jk}^{l}|^2\} - \rho_{jk} |\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2 + \sigma_{\mathrm{DL}}^2}$$

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}} = \frac{p_{jk} \frac{|\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2}{\mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}}}{\sum_{l=1}^{L} \sum_{i=1}^{K_l} p_{li} \frac{\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}} \mathbf{h}_{li}^{j}|^2\}}{\mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}} - p_{jk} \frac{|\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2}{\mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}} + \sigma_{\mathrm{UL}}^2}$$

Similar structure if $\mathbf{w}_{jk} = \mathbf{v}_{jk} / \sqrt{\mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}}$:

	Downlink	Uplink
Transmit power	ρ_{li}	p_{li}
Channel gain	$\lvert\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}\rvert^2$	Same value
Interference gain from UE i, cell l	$\mathbb{E}\{\lvert\mathbf{w}_{li}^{\mathrm{H}} \mathbf{h}_{jk}^{l}\rvert^2\}$	$\mathbb{E}\{\lvert\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{li}^{j}\rvert^2\}$ $(j \leftrightarrow l,\ k \leftrightarrow i)$
Noise power	σ_{DL}^2	σ_{UL}^2

Uplink-Downlink Duality

Theorem

Consider a given set of receive combining vectors $\{\mathbf{v}_{li}\}$ and UL powers $\{p_{li}\}$, which achieves $\underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}}$ for all j and k.

If the precoding vectors are selected as $\mathbf{w}_{jk} = \mathbf{v}_{jk} / \sqrt{\mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}}$, then there exist DL powers $\{\rho_{li}\}$ such that

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{DL}} = \underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}} \quad j = 1, \ldots, L, \ k = 1, \ldots, K_j.$$

The sum transmit power in the DL and UL is related as

$$\frac{1}{\sigma_{\mathrm{DL}}^2} \sum_{j=1}^{L} \sum_{k=1}^{K_j} \rho_{jk} = \frac{1}{\sigma_{\mathrm{UL}}^2} \sum_{j=1}^{L} \sum_{k=1}^{K_j} p_{jk}.$$

- Main insight: Use receive combining vectors for transmit precoding!
- Less important: DL powers can be computed in closed-form

Transmit Precoding Schemes

Implication from the uplink-downlink duality:

- Select precoding vectors based on receive combining vectors as

$$\mathbf{w}_{jk} = \frac{\mathbf{v}_{jk}}{\|\mathbf{v}_{jk}\|}$$

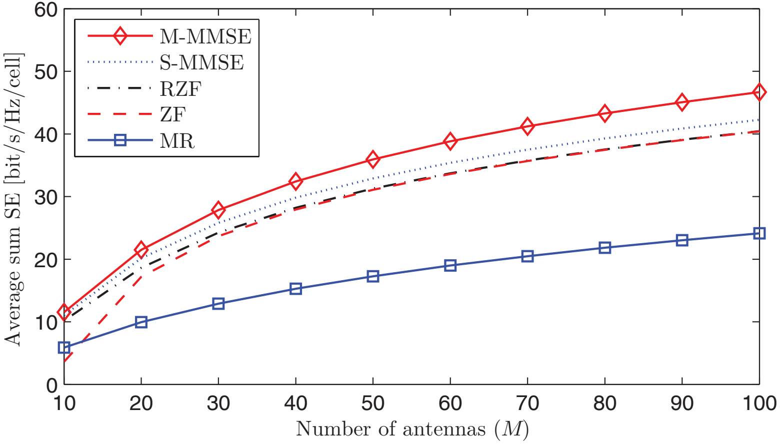
where

$$\left[\mathbf{v}_{j1} \, \ldots \, \mathbf{v}_{jK_j}\right] = \begin{cases} \mathbf{V}_j^{\text{M-MMSE}} & \text{with M-MMSE precoding} \\ \mathbf{V}_j^{\text{S-MMSE}} & \text{with S-MMSE precoding} \\ \mathbf{V}_j^{\text{RZF}} & \text{with RZF precoding} \\ \mathbf{V}_j^{\text{ZF}} & \text{with ZF precoding} \\ \mathbf{V}_j^{\text{MR}} & \text{with MR precoding} \end{cases}$$

- Note: These are all heuristic schemes

Normalize by $\|\mathbf{v}_{jk}\|$ instead of $\sqrt{\mathbb{E}\{\|\mathbf{v}_{jk}\|^2\}}$ to reduce variations in precoded channel $(\mathbf{h}_{jk}^{j})^{\text{H}}\mathbf{w}_{jk}$

Downlink SE Simulations: Universal Pilot Reuse

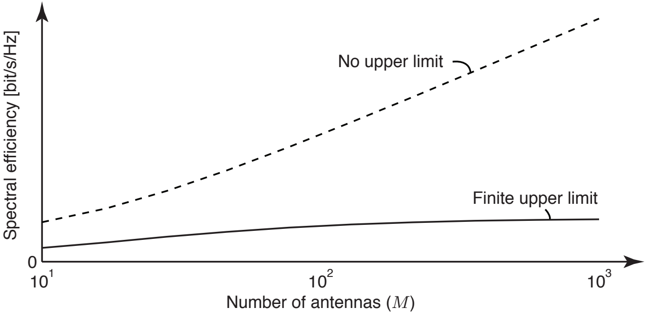


$K = 10$ UEs per cell, $\tau_c = 200$, $\tau_p = K$, $\tau_u = 0$ (DL data only)

Running example, Gaussian local scattering model: ASD $\sigma_\varphi = 10°$

Asymptotic Analysis

What is it Known as $M_j \to \infty$?



- Finite upper limit — uncorrelated Rayleigh fading [Mar10]
- No upper limit
  - Pilot contamination precoding, all base stations serve all users
  - Channels in different eigenspaces
  - Using semi-blind estimation and $\tau_c \to \infty$

We will prove there is *no upper limit* under general, practical conditions

Preliminary Assumptions for Asymptotic Analysis

Assumption 1

The spatial channel correlation matrix $\mathbf{R}_{li}^{j}$ satisfies

1. $\liminf_{M_j} \frac{1}{M_j} \mathrm{tr}(\mathbf{R}_{li}^{j}) > 0$
2. $\limsup_{M_j} \|\mathbf{R}_{li}^{j}\|_2 < \infty$

for $l = 1, \ldots, L$ and $i = 1, \ldots, K_l$.

- First condition: Expected channel gain grows proportionally to M_j
- Second condition: No eigenvalue grows without bound
- These are necessary conditions for channel hardening

Linearly Independent Matrices

Definition (Linearly independent correlation matrices)

Consider the correlation matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$. This matrix is *linearly independent* of the correlation matrices $\mathbf{R}_1, \ldots, \mathbf{R}_N \in \mathbb{C}^{M \times M}$ if

$$\left\| \mathbf{R} - \sum_{i=1}^{N} c_i \mathbf{R}_i \right\|_F^2 > 0$$

for all scalars $c_1, \ldots, c_N \in \mathbb{R}$.

We say that $\mathbf{R}$ is *asymptotically linearly independent* of $\mathbf{R}_1, \ldots, \mathbf{R}_N$ if

$$\liminf_{M} \frac{1}{M} \left\| \mathbf{R} - \sum_{i=1}^{N} c_i \mathbf{R}_i \right\|_F^2 > 0$$

for all scalars $c_1, \ldots, c_N \in \mathbb{R}$.

Practical Correlation Matrices are Linearly Independent

Consider the two matrices

$$\mathbf{R} = \begin{bmatrix} \epsilon_1 & 0 & \dots \\ 0 & \ddots & 0 \\ \dots & 0 & \epsilon_M \end{bmatrix} \quad \text{and} \quad \mathbf{R}_1 = \mathbf{I}_M$$

where $\epsilon_1, \dots, \epsilon_M$ are i.i.d. positive random variables

- From the law of large numbers:

$$\frac{1}{M} \|\mathbf{R} - c_1 \mathbf{R}_1\|_F^2 = \frac{1}{M} \sum_{m=1}^{M} (\epsilon_m - c_1)^2 \geq \frac{1}{M} \sum_{m=1}^{M} \left(\epsilon_m - \frac{1}{M} \sum_{n=1}^{M} \epsilon_n \right)^2$$

$$\to \mathbb{E}\{(\epsilon_m - \mathbb{E}\{\epsilon_m\})^2\} = \text{Variance} > 0$$

Take any linearly dependent matrices (e.g., uncorrelated Rayleigh fading). Add perturbations: they become asymptotically linearly independent.

- Nature will only create linearly independent correlation matrices

Asymptotic Behavior of MR

Theorem (MR combining)

Under Assumption 1, if MR combining with $\mathbf{v}_{jk} = \hat{\mathbf{h}}_{jk}^{j}$ is used, it follows that

$$\underline{\mathsf{SINR}}_{jk}^{\mathrm{UL}} - \frac{p_{jk}^2 \mathrm{tr}\left(\mathbf{R}_{jk}^{j} \mathbf{\Psi}_{jk}^{j} \mathbf{R}_{jk}^{j}\right)}{\sum\limits_{(l,i) \in \mathcal{P}_{jk} \setminus (j,k)} p_{li}^2 \frac{\left|\mathrm{tr}\left(\mathbf{R}_{li}^{j} \mathbf{\Psi}_{jk}^{j} \mathbf{R}_{jk}^{j}\right)\right|^2}{\mathrm{tr}\left(\mathbf{R}_{jk}^{j} \mathbf{\Psi}_{jk}^{j} \mathbf{R}_{jk}^{j}\right)}} \to 0$$

as $M_j \to \infty$.[10]

- Impact of noise and non-coherent interference vanishes
- Coherent signal and interference terms remain
  - There is a finite upper SE limit
- Similar result can be proved for the downlink

[10] Except in special cases when $\mathrm{tr}(\mathbf{R}_{jk}^{j} \mathbf{R}_{li}^{j})/M_j \to 0$ for all $(l,i) \in \mathcal{P}_{jk} \setminus (j,k)$

Asymptotic Behavior of M-MMSE

Theorem (M-MMSE combining)

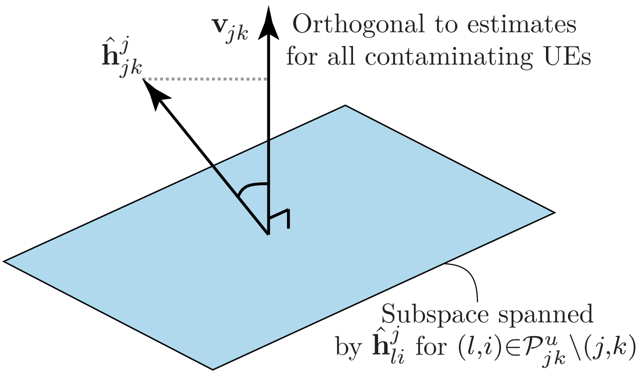
If BS j uses M-MMSE combining with MMSE channel estimation, then the uplink SE of UE k in cell j grows without bound as $M_j \to \infty$, if

- Assumption 1 holds
- The correlation matrix $\mathbf{R}_{jk}^{j}$ is asymptotically linearly independent of the set of correlation matrices $\mathbf{R}_{li}^{j}$ with $(l,i) \in \mathcal{P}_{jk} \setminus (j,k)$.

- Impact of noise, coherent, and non-coherent interference vanishes
- Asymptotic linear independence is key
  - Does not hold under uncorrelated Rayleigh fading
  - Practical correlation matrices satisfy this condition
- Channel estimates are linearly independent since

$$\hat{\mathbf{h}}_{jk}^{j} - c\hat{\mathbf{h}}_{li}^{j} = \left(\sqrt{p_{jk}}\mathbf{R}_{jk}^{j} - c\sqrt{p_{li}}\mathbf{R}_{li}^{j}\right)\boldsymbol{\Psi}_{jk}^{j}\mathbf{y}_{jjk}^{p}$$

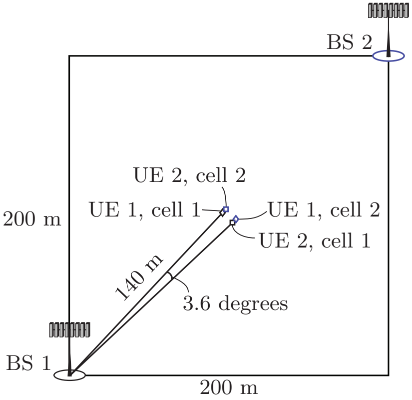
Asymptotic Behavior of M-MMSE: Geometric Illustration



UEs that share a pilot have linearly independent channel estimates

The indicated $\mathbf{v}_{jk}$ rejects the coherent interference: $\mathbf{v}_{jk}^{\mathrm{H}} \hat{\mathbf{h}}_{li}^{j} = 0$
The desired signal remains: $\mathbf{v}_{jk}^{\mathrm{H}} \hat{\mathbf{h}}_{jk}^{j} \neq 0$

Simulation Setup for Asymptotic Behavior

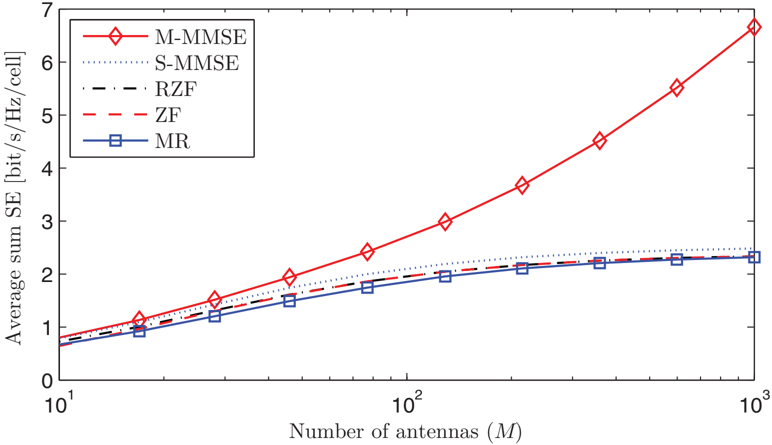


Uplink scenario with very strong coherent interference:

- $L = 2$ cells
- $K = 2$ UEs per cell, $\tau_p = 2$.
- SNR -2 dB from serving BS, -2.3 dB from interfering BS
- Gaussian local scattering model with 10° ASD

- Channels modeled as in running example (but no shadow fading)

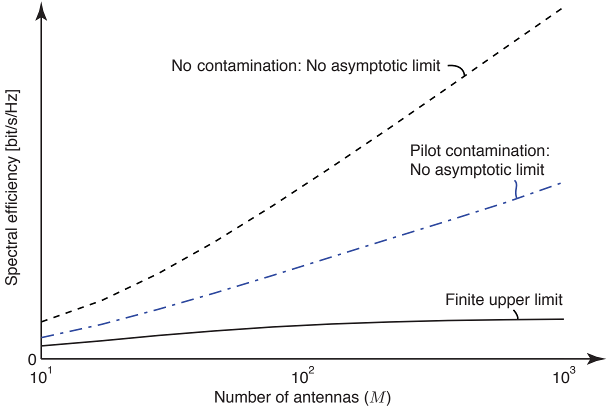
Asymptotic SE Behavior



Sum SE as a function of the number of BS antennas (logarithmic scale).

SE grows unboundedly as $\log_2(M)$ with M-MMSE combining
Convergence to finite limits with other combining schemes

Is Pilot Contamination a Fundamental Limitation?



No! Unlimited capacity is achieved using the following ingredients

- Spatial correlated channels – only a minor amount is needed
- MMSE channel estimation – not least-square
- Optimal linear combining – not MR, ZF, or S-MMSE

Key Points

- Asymptotic behavior
  - **Impact of noise and non-coherent interference always vanish**
  - Coherent interference caused by pilot contamination is a challenge
  - Impact of **coherent interference vanish with M-MMSE**
  - SE grows as $\log_2(M)$ when using M-MMSE
- **Spatial channel correlation** is important in asymptotic analysis
  - Enables **unbounded SE when using M-MMSE**
  - Determines the upper limit when using S-MMSE, RZF, ZF, MR
- **Knowing the channel correlation matrices is key**
  - Only diagonals are needed if element-wise MMSE estimation is used (details found in [BHS18])
  - Correlation matrices can be estimated from pilots

Frequently Asked Questions

1. What happens in the downlink?
    - Same thing: SE grows without bound as $M \to \infty$
2. How do you estimate covariance matrices?
    - See "Massive MIMO with imperfect channel covariance information"
3. Is it critical to know the full covariance matrices?
    - No, you only need the diagonals
      (and that these are linearly independent)
4. Can you have an infinite number of users/cells?
    - Maybe if the covariance matrices are *asymptotically* linearly independent
    - Current proof does not support this case.

...if you want to know more!

Massive MIMO Has Unlimited Capacity

Emil Björnson, *Member, IEEE*, Jakob Hoydis, *Member, IEEE*, Luca Sanguinetti, *Senior Member, IEEE*

***Abstract*—The capacity of cellular networks can be improved by the unprecedented array gain and spatial multiplexing offered by Massive MIMO. Since its inception, the coherent interference caused by pilot contamination has been believed to create a finite capacity limit, as the number of antennas goes to infinity. In this paper, we prove that this is incorrect and an artifact from using simplistic channel models and suboptimal precoding/combining schemes. We show that with multicell MMSE precoding/combining and a tiny amount of spatial channel correlation or large-scale fading variations over the array, the capacity increases without bound as the number of antennas increases, even under pilot contamination. More precisely, the result holds when the channel covariance matrices of the contaminating users are asymptotically linearly independent, which is generally the case. If also the diagonals of the covariance matrices are linearly independent, it is sufficient to know these diagonals (and not the full covariance matrices) to achieve an unlimited asymptotic capacity.**

***Index Terms*—Massive MIMO, ergodic capacity, asymptotic analysis, spatial correlation, multi-cell MMSE processing, pilot contamination.**

I. Introduction

The Shannon capacity of a channel manifests the spectral efficiency (SE) that it supports. Massive MIMO (multiple-input multiple-output) improves the sum SE of cellular networks by spatial multiplexing of a large number of user equipments (UEs) per cell [1]. It is therefore considered a key time-division duplex (TDD) technology for the next generation of cellular networks [2]–[4]. The main difference between Massive MIMO and classical multiuser MIMO is the large number of antennas, M, at each base station (BS) whose signals are processed by individual radio-frequency chains. By exploiting channel estimates for coherent receive combining, the uplink signal power of a desired UE is reinforced by a factor M, while the power of the noise and independent interference does not increase. The same principle holds for the transmit precoding in the downlink. Since the channel estimates are obtained by uplink pilot signaling and the pilot resources are limited by the channel coherence time, the same pilots must be reused in multiple cells. This leads to pilot contamination which has two main consequences: the channel estimation quality is reduced due to pilot interference and the channel estimate of a desired UE is correlated with the channels to the interfering UEs that use the same pilot. Marzetta showed in his seminal paper [1] that the interference from these UEs during data transmission is also reinforced by a factor M, under the assumptions of maximum ratio (MR) combining/precoding and independent and identically distributed (i.i.d.) Rayleigh fading channels. This means that pilot contamination creates a finite SE limit as $M \to \infty$.

The large-antenna limit has also been studied for other combining/precoding schemes, such as the minimum mean squared error (MMSE) scheme. Single-cell MMSE (S-MMSE) was considered in [5]–[7], while multicell MMSE (M-MMSE) was considered in [8], [9]. The difference is that with M-MMSE, the BS makes use of estimates of the channels from the UEs in all cells, while with S-MMSE, the BS only uses channel estimates of the UEs in the own cell. In both cases, the SE was proved to have a finite limit as $M \to \infty$, under the assumption of i.i.d. Rayleigh fading channels (i.e., no spatial correlation). In contrast, there are special cases of spatially correlated fading that give rise to rank-deficient covariance matrices [10]–[12]. If the UEs that share a pilot have rank-deficient covariance matrices with orthogonal support, then pilot contamination vanishes and the SE can increase without bound. The covariance matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ have orthogonal support if $\mathbf{R}_1\mathbf{R}_2 = \mathbf{0}$. To understand this condition, note that for arbitrary covariance matrices

$$\mathbf{R}_1 = \begin{bmatrix} a & c \\ c^\star & b \end{bmatrix} \quad \mathbf{R}_2 = \begin{bmatrix} d & f \\ f^\star & e \end{bmatrix} \tag{1}$$

every element of $\mathbf{R}_1\mathbf{R}_2$ must be zero. The first element is $ad + cf^\star$. If we model the practical covariance matrices of

[BHS18]: Available on `arxiv.org/abs/1705.00538`

Energy Efficiency

Definition of Energy Efficiency

Energy Efficiency

The EE of a wireless communication system is the number of bits that can be reliably transmitted per unit of energy [bit/Joule][11]

$$\mathsf{EE} = \frac{\text{Throughput}\,[\text{bit/s/cell}]}{\text{Power consumption}\,[\text{W/cell}]}$$

Power Consumption

$$\underbrace{\mathsf{PC}}_{\text{Power consumption}} = \underbrace{\mathsf{ETP}}_{\text{Effective transmit power}} + \underbrace{\mathsf{CP}}_{\text{Circuit power}}$$

[11] Some works on EE consider erroneous EE metrics in bit/Joule/Hz. The EE does not scale linearly with the bandwidth!

Circuit Power Model

Circuit Power for Massive MIMO

Transceiver hardware at the BS and UE are not the only contributions

A CP model for a generic BS j in a Massive MIMO network is

$$\mathrm{CP}_j = \underbrace{P_{\mathrm{FIX},j}}_{\text{Fixed power}} + \underbrace{P_{\mathrm{TC},j}}_{\text{Transceiver chains}} + \underbrace{P_{\mathrm{CE},j}}_{\text{Channel estimation}} + \underbrace{P_{\mathrm{C/D},j}}_{\text{Coding/Decoding}} + \underbrace{P_{\mathrm{BH},j}}_{\text{Load-dependent backhaul}} + \underbrace{P_{\mathrm{LP},j}}_{\text{Linear processing}}$$

- Before Massive MIMO, the CP of these operations was negligible compared to $P_{\mathrm{FIX},j}$
- A **tractable** and "realistic" model for each term is now needed
- Economical expenses can be potentially added — by dividing the network cost rate (in $\$/s$) with the energy price (in $\mathrm{Joule}/\$$).

Circuit Power for Massive MIMO: Combining/Precoding (2/2)

Scheme	$P_{\mathrm{SP-R/T},j}$	$P_{\mathrm{SP-C},j}^{\mathrm{UL}}$	$P_{\mathrm{SP-C},j}^{\mathrm{DL}}$
M-MMSE (MMSE estimation)	$\frac{3B}{\tau_c L_{\mathrm{BS}}} K_j M_j (\tau_u + \tau_d)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} \left(\sum_{l=1}^{L} \frac{(3M_j^2+M_j)K_l}{2} + \frac{M_j^3}{3} + 2M_j + M_j \tau_p (\tau_p - K_j) \right)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} M_j K_j$
M-MMSE (EW-MMSE estimation)	$\frac{3B}{\tau_c L_{\mathrm{BS}}} K_j M_j (\tau_u + \tau_d)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} \left(\sum_{l=1}^{L} \frac{(M_j^2+3M_j)K_l}{2} + (M_j^2 - M_j) K_j + \frac{M_j^3}{3} + 2M_j + M_j \tau_p (\tau_p - K_j) \right)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} M_j K_j$
S-MMSE	$\frac{3B}{\tau_c L_{\mathrm{BS}}} K_j M_j (\tau_u + \tau_d)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} \left(\frac{3M_j^2 K_j}{2} + \frac{M_j K_j}{2} + \frac{M_j^3 - M_j}{3} + \frac{7}{3} M_j \right)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} M_j K_j$
RZF	$\frac{3B}{\tau_c L_{\mathrm{BS}}} K_j M_j (\tau_u + \tau_d)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} \left(\frac{3K_j^2 M_j}{2} + \frac{3K_j M_j}{2} + \frac{K_j^3 - K_j}{3} + \frac{7}{3} K_j \right)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} M_j K_j$
ZF	$\frac{3B}{\tau_c L_{\mathrm{BS}}} K_j M_j (\tau_u + \tau_d)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} \left(\frac{3K_j^2 M_j}{2} + \frac{K_j M_j}{2} + \frac{K_j^3 - K_j}{3} + \frac{7}{3} K_j \right)$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} M_j K_j$
MR	$\frac{3B}{\tau_c L_{\mathrm{BS}}} K_j M_j (\tau_u + \tau_d)$	$\frac{7B}{\tau_c L_{\mathrm{BS}}} K_j$	$\frac{3B}{\tau_c L_{\mathrm{BS}}} M_j K_j$

Once again: A linear model in M_j and K_j is not accurate enough!

CP with Different Schemes — Parameters for CP model

Parameter	Value set 1	Value set 2
Fixed power: P_{FIX}	$10\,\mathrm{W}$	$5\,\mathrm{W}$
Power for BS LO: P_{LO}	$0.2\,\mathrm{W}$	$0.1\,\mathrm{W}$
Power per BS antennas: P_{BS}	$0.4\,\mathrm{W}$	$0.2\,\mathrm{W}$
Power for antenna at UE: P_{UE}	$0.2\,\mathrm{W}$	$0.1\,\mathrm{W}$
Power for data coding: P_{COD}	$0.1\,\mathrm{W/(Gbit/s)}$	$0.01\,\mathrm{W/(Gbit/s)}$
Power for data decoding: P_{DEC}	$0.8\,\mathrm{W/(Gbit/s)}$	$0.08\,\mathrm{W/(Gbit/s)}$
BS computational efficiency: L_{BS}	$75\,\mathrm{Gflops/W}$	$750\,\mathrm{Gflops/W}$
Power for backhaul traffic: P_{BT}	$0.25\,\mathrm{W/(Gbit/s)}$	$0.025\,\mathrm{W/(Gbit/s)}$

- The first set is inspired from [KKC+11, KG11, TSZ12, YM13].
- The second setup assumes the hardware's PC is reduced by a factor of two whereas the computational efficiencies (which benefit from Moore's law) is increased by a factor of ten.

Extremely hardware-specific: may take substantially different values — Matlab code (available online) enables simple testing of other values.

Energy Efficiency and Throughput Tradeoff

Problem Statement

The EE of cell j is thus computed as

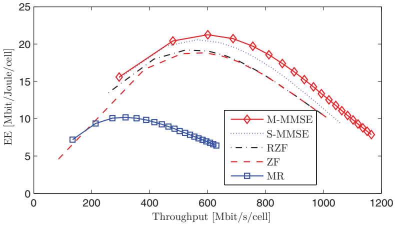
$$\mathsf{EE}_j = \frac{\mathsf{TR}_j}{\mathsf{ETP}_j + \mathsf{CP}_j}$$

with

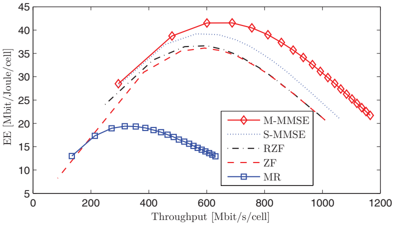
$$\mathsf{ETP}_j = \underbrace{\frac{\tau_p}{\tau_c} \sum_{k=1}^{K_j} \frac{1}{\mu_{\mathrm{UE},jk}} p_{jk}}_{\text{ETP for UL pilots}} + \underbrace{\frac{\tau_u}{\tau_c} \sum_{k=1}^{K_j} \frac{1}{\mu_{\mathrm{UE},jk}} p_{jk}}_{\text{ETP in the UL}} + \underbrace{\frac{1}{\mu_{\mathrm{BS},j}} \frac{\tau_d}{\tau_c} \sum_{k=1}^{K_j} \rho_{jk}}_{\text{ETP in the DL}}$$

Set $\mu_{\mathrm{UE},jk} = 0.4$ and $\mu_{\mathrm{BS},j} = 0.5$. Deliberately higher than 25% (i.e., higher than in contemporary PA): the low power levels per antenna (in the mW range) of Massive MIMO allow to use more efficient PA.

Running Example with $K = 10$, $M \in \{10, 20, \ldots, 200\}$



(c) First set of CP values



(d) Second set of CP values

- All schemes allow to jointly increase the EE and throughput
- The curves are quite smooth around the EE maximum — a variety of throughput (or BS antennas) values provide nearly maximum EE
- M-MMSE provides highest EE for any throughput value: its higher computational complexity pays off both in terms of SE and EE
- With the second set of values: EE is roughly doubled (CP coefficients are halved), but trends are the same
- The optimal EE is achieved for $M \in \{30, 40\}$: far from what it is envisioned for Massive MIMO, but $M/K \in \{3, 4\}$ is as expected

Power Allocation

Utility Function

How to measure network performance?

- There are $\sum_{l=1}^{L} K_l$ UEs, each with UL SE and DL SE
- Combining/precoding and transmit power allocation affect SE
- For given precoding, the DL SEs have a common structure:

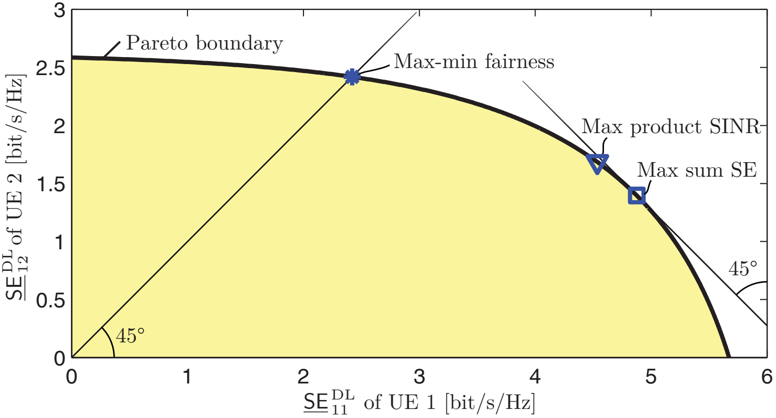
$$\underline{\mathsf{SE}}_{jk}^{\mathrm{DL}} = \frac{\tau_d}{\tau_c} \log_2 \left(1 + \frac{\rho_{jk} a_{jk}}{\sum\limits_{l=1}^{L} \sum\limits_{i=1}^{K_l} \rho_{li} b_{lijk} + \sigma_{\mathrm{DL}}^2} \right) \quad \text{for UE } k \text{ in cell } j$$

$$a_{jk} = |\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2 \quad b_{lijk} = \begin{cases} \mathbb{E}\{|\mathbf{w}_{li}^{\mathrm{H}} \mathbf{h}_{jk}^{l}|^2\} & (l,i) \neq (j,k) \\ \mathbb{E}\{|\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{l}|^2\} - |\mathbb{E}\{\mathbf{w}_{jk}^{\mathrm{H}} \mathbf{h}_{jk}^{j}\}|^2 & (l,i) = (j,k) \end{cases}$$

Utility function: Maps all SEs into a single performance metric

$$U(\mathsf{SE}_{11}, \ldots, \mathsf{SE}_{LK_L}) = \begin{cases} \sum_{j=1}^{L} \sum_{k=1}^{K_j} \mathsf{SE}_{jk} & \text{Max sum SE} \\ \min_{j,k} \mathsf{SE}_{jk} & \text{Max-min fairness} \\ \prod_{j=1}^{L} \prod_{k=1}^{K_j} \mathsf{SINR}_{jk} & \text{Max product SINR} \end{cases}$$

Example: SE Region and Operating Points



SE region with all $(\underline{\mathsf{SE}}_{11}^{\mathrm{DL}}, \underline{\mathsf{SE}}_{12}^{\mathrm{DL}})$ achieved by different power allocations

Pareto boundary contains all resource-efficient operating points

The operating points maximizing the three utility functions are indicated

Basic Optimization Theory

Optimization problem on standard form:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_n(\mathbf{x}) \leq 0 \quad n = 1, \ldots, N \end{aligned}$$

- Optimization variable $\mathbf{x} = [x_1 \, x_2 \, \ldots \, x_V]^{\mathrm{T}} \in \mathbb{R}^V$
- Utility function $f_0 : \mathbb{R}^V \to \mathbb{R}$
- Constraint functions $f_n : \mathbb{R}^V \to \mathbb{R}$, $n = 1, \ldots, N$

Solvable to global optimality with standard techniques (CVX, Yalmip) if

- **Linear program**: f_0 and $f_1, \ldots, f_N$ are linear or affine functions
- **Geometric program:** $-f_0$ and $f_1 - 1, \ldots, f_N - 1$ are posynomials[12]
- **Convex program:** $-f_0$ and $f_1, \ldots, f_N$ are convex functions

[12] f_n is posynomial if $f_n(\mathbf{x}) = \sum_{b=1}^{B} c_b x_1^{e_{1,b}} x_2^{e_{2,b}} \cdots x_V^{e_{V,b}}$ for some positive integer B, constants $c_b > 0$, and exponents $e_{1,b}, \ldots, e_{V,b} \in \mathbb{R}$ for $b = 1, \ldots, B$

Downlink Power Allocation

Power optimization problem:

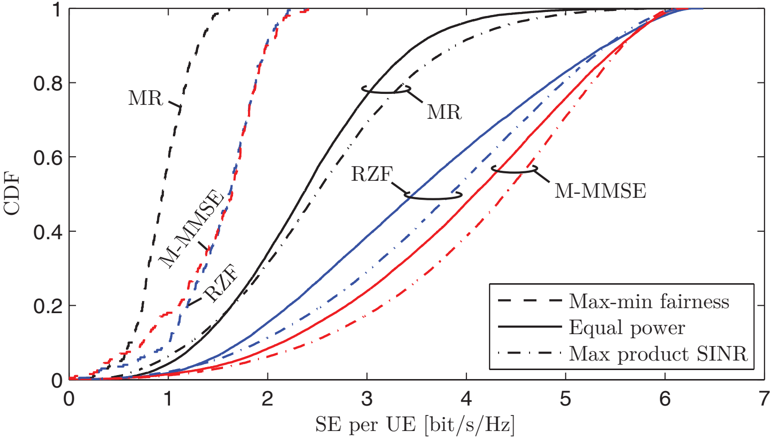
$$\begin{aligned} \underset{\rho_{11}\geq 0,\ldots,\rho_{LK_L}\geq 0}{\text{maximize}} \quad & U(\underline{\mathsf{SE}}_{11}^{\mathrm{DL}},\ldots,\underline{\mathsf{SE}}_{LK_L}^{\mathrm{DL}}) \\ \text{subject to} \quad & \sum_{k=1}^{K_j} \rho_{jk} \leq P_{\max}^{\mathrm{DL}}, \quad j=1,\ldots,L \end{aligned}$$

- Maximum total transmit power $P_{\max}^{\mathrm{DL}} \geq 0$ per BS
- Fixed precoding and UL transmit powers

Similar to classic single-antenna power allocation problems [CHLT08, WCLa$^+$12, BJ13]

- Max sum SE: Non-convex program, hard to solve
- Max-min fairness: Quasi-linear program, easy to solve
- Max product SINR: Geometric program, easy to solve

Running Example: Downlink with Power Optimization



CDF of DL SE per UE for the running example with $M = 100$, $K = 10$, $f = 2$, and Gaussian local scattering model with ASD $\sigma_\varphi = 10°$.

Max product SINR provides high rates and fairness

Uplink Power Control

Uplink transmit power optimization is complicated since it affects

- Quality of channel estimates
- Combining vectors
- Power of data symbols

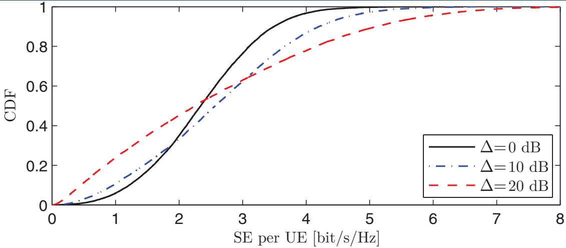
Heuristic power control

- Each UE has a maximum transmit power $P_{\max}^{\mathrm{UL}} > 0$
- Near-far effect: Reduce received power differences between UEs
- Maximum received power ratio $\Delta \geq 0\,$dB

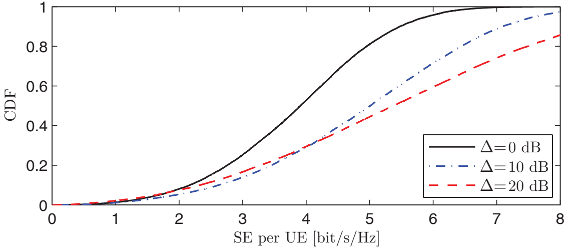
$$p_{jk} = \begin{cases} P_{\max}^{\mathrm{UL}} & \Delta > \frac{\beta_{jk}^{j}}{\beta_{j,\min}^{j}} \\ P_{\max}^{\mathrm{UL}} \Delta \frac{\beta_{j,\min}^{j}}{\beta_{jk}^{j}} & \Delta \leq \frac{\beta_{jk}^{j}}{\beta_{j,\min}^{j}} \end{cases}$$

with $\beta_{j,\min}^{j} = \min_{i=1,\dots,K_j} \beta_{ji}^{j}$

Running Example: Uplink with Power Control



MR combining



M-MMSE combining

CDF of UL SE per UE

$M = 100$, $K = 10$

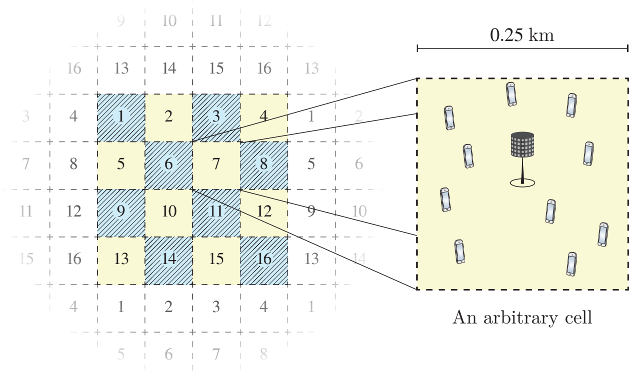
Gaussian local scattering model with $\sigma_{\varphi} = 10°$.

Small Δ improves SE of weakest UEs

Largest effect on MR

Case Study

Case Study: Scenario



Analyze practical baseline performance with

- 3GPP 3D UMi NLoS channel model[13]
- Optimized power allocation
- Least-square channel estimation (without channel statistics)
- MR or RZF processing

[13] Using QuaDRiGa implementation by Fraunhofer Heinrich Hertz Institute

Array and Transmission Configurations

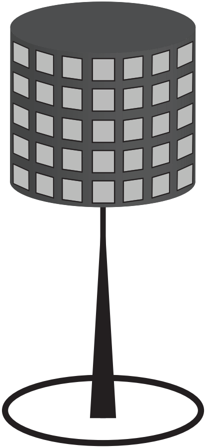
Maximum transmit power

- Uplink: 20 dBm per UE
- Downlink: 30 dBm per BS

Cylindrical array configurations
(“horizontal $\times$ vertical $\times$ polarization”):

1. $10 \times 5 \times 2$ $(M = 100)$
2. $20 \times 5 \times 1$ $(M = 100)$
3. $20 \times 5 \times 2$ $(M = 200)$

BS height 25 m, UE height 1.5 m

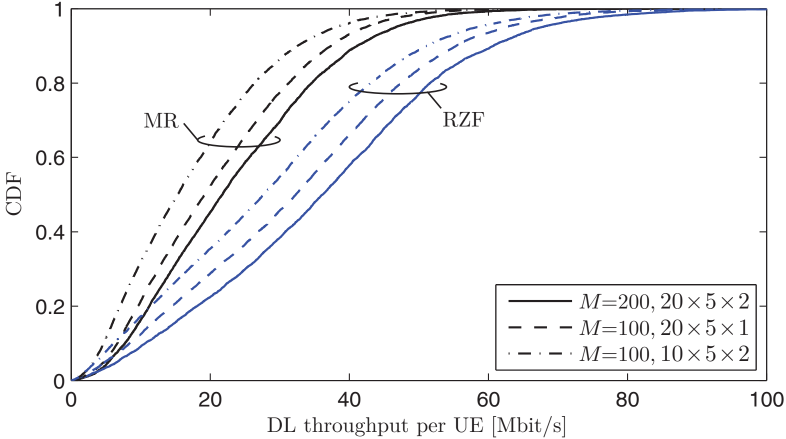


1) and 2) have same number of RF chains
2) and 3) have same physical size

Network Parameters

Parameter	Value
UE dropping	$K = 10$ UEs in 250 m × 250 m area around each BS, with 35 m minimum distance
Carrier frequency	2 GHz
Bandwidth	$B = 20$ MHz
Receiver noise power	-94 dBm
Number of subcarriers	2000
Subcarrier bandwidth	10 kHz
Cyclic prefix overhead	5%
Frame dimensions	$B_c = 50$ kHz, $T_c = 4$ ms
Subcarriers per frame	5
Useful samples per frame	$\tau_c = B_c T_c / 1.05 \approx 190$
Pilot reuse factor	$f = 2$
Number of UL pilot sequences	$\tau_p = 30$

Downlink: Max Product SINR Power Allocation

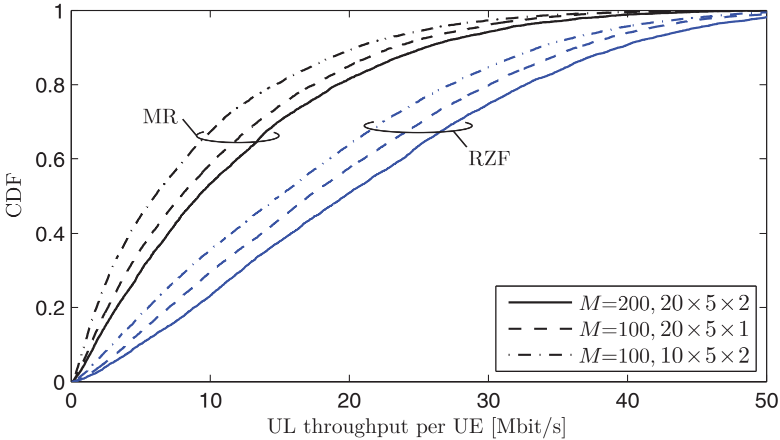


CDF of downlink throughput per UE in the case study

Fixed physical size: Use dual-polarization to double number of antennas

Fixed number of RF chains: Use larger uni-polarized array

Uplink: Heuristic power control $\Delta = 20\,\text{dB}$



CDF of uplink throughput per UE in the case study

Similar observations as in downlink

Case Study: Key Points

Average sum throughput over 20 MHz channel

- Downlink: 358 Mbit/s (area throughput: 5.7 Gbit/s/km^2)
- Uplink: 209 Mbit/s (area throughput: 3.3 Gbit/s/km^2)
- Difference due to twice as many downlink data samples per frame
- Tradeoff between high average throughput and user fairness

LTE in similar setup:

- Downlink area throughput: 263 Mbit/s/km^2
- Uplink area throughput: 115 Mbit/s/km^2
- Massive MIMO setup delivers 20–30 times higher throughput
- Gain from multiplexing and coherent precoding/combining

Open Problems

Goldmine of Open Problems

Massive MIMO is a mature research field, no low-hanging fruits!

Make Massive MIMO work in FDD mode

- Long-standing challenge. Is it practically feasible to exploit sparsity?

Channel measurements, channel modeling, data traffic modeling

- Required for system level simulations

Cross-layer system design

- Protocols for random access and system information broadcast
- Spatial resource allocation
- Power control balancing sum SE and fairness
- Estimation of spatial correlation properties under mobility

Further Open Problems

Information theory advances

- Tighter lower bounds on ergodic capacity, without channel hardening
- Non-trivial upper bounds on capacity

New deployment characteristics

- Multi-antenna users, distributed arrays, cell-free (network MIMO)

What will be the successor of Massive MIMO? Can we increase spectral efficiency with $10\times$ over Massive MIMO?

[AMVW91] S. Anderson, M. Millnert, M. Viberg, and B. Wahlberg, "An adaptive array for mobile communication systems," *IEEE Trans. Veh. Technol.*, vol. 40, no. 1, pp. 230–236, 1991.

[BHS18] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, 2018.

[BJ13] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[CHLT08] M. Chiang, P. Hande, T. Lan, and C. Tan, "Power control in wireless cellular networks," *Foundations and Trends in Networking*, vol. 2, no. 4, pp. 355–580, 2008.

[GERT11] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE VTC Fall*, 2011.

[GJ11] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *Proc. IEEE SPAWC*, 2011.

[HCPR12] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving "massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.

[HHWtB12] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *Proc. IEEE ISWCS*, 2012.

[JAMV11] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[KG11] R. Kumar and J. Gurugubelli, "How green the LTE technology can be?" in *Proc. Wireless VITAE*, 2011.

[KKC+11] D. Kang, D. Kim, Y. Cho, J. Kim, B. Park, C. Zhao, and B. Kim, "1.6 - 2.1 GHz broadband Doherty power amplifiers for LTE handset applications," in *Proc. IEEE MTT-S*, June 2011, pp. 1–4.

[Mar10] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[Qua12] Qualcomm, "Rising to meet the 1000x mobile data challenge," Qualcomm Incorporated, Tech. Rep., 2012.

[SBEM90] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 39, no. 1, pp. 56–67, 1990.

[TSZ12] S. Tombaz, K. W. Sung, and J. Zander, "Impact of densification on energy efficiency in wireless access networks," in *Proc. IEEE GLOBECOM Workshop*, Dec 2012, pp. 57–62.

[WCLa+12] P. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, "Weighted sum-rate maximization in wireless networks: A review," *Foundations and Trends in Networking*, vol. 6, no. 1-2, pp. 1–163, 2012.

[Win87] J. H. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Trans. Commun.*, vol. 35, no. 11, pp. 1222–1230, 1987.

[YGFL13] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[YM13] H. Yang and T. L. Marzetta, "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," in *Proc. IEEE Online GreenComm*, 2013, pp. 27–32.