

# Massive MIMO and HetNets: Benefits and Challenges

Jakob Hoydis

Bell Laboratories, Alcatel-Lucent, Stuttgart, Germany  
jakob.hoydis@alcatel-lucent.com

Newcom# Summer School  
Interference Management for Tomorrow's Wireless Networks  
Eurecom, Sophia-Antipolis, France  
May 29, 2013

## Some figures, industry trends, and the data explosion

Today,

- 66 % sleep with smart phone (USA)
- 84 % choose Internet over partner or car (Germany)
- 67 % would cut anything but mobile broadband (UK)

By 2017, there will be<sup>1</sup>

- 13 × more mobile data traffic than in 2012
- 10,000,000,000 connected devices
- 2/3 of the total traffic generated by mobile video streaming and communications

---

<sup>1</sup>Source: Cisco, Yankee

# Challenges for the next years

How can we

- provide the necessary area spectral efficiency?
  - ▶ Without new spectral resources ( $\leq 6$  GHz) or technological breakthroughs.
- avoid that the backhaul becomes the bottleneck?
  - ▶ While increasing the flexibility and cost of access point deployment.
- keep the energy consumption in the infrastructure and the terminals low?
  - ▶ Mobility is not anymore limited by coverage but rather by the battery duration.

## Possible solutions

Today, **network densification** is the only answer to the capacity crunch:

- Small cells: Network capacity scales linearly with the cell density.
- Massive MIMO: Interference can be almost entirely eliminated.

Fortunately, both technologies can also significantly reduce the radiated power. Increasing capacity while simultaneously reducing the energy consumption is possible!

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

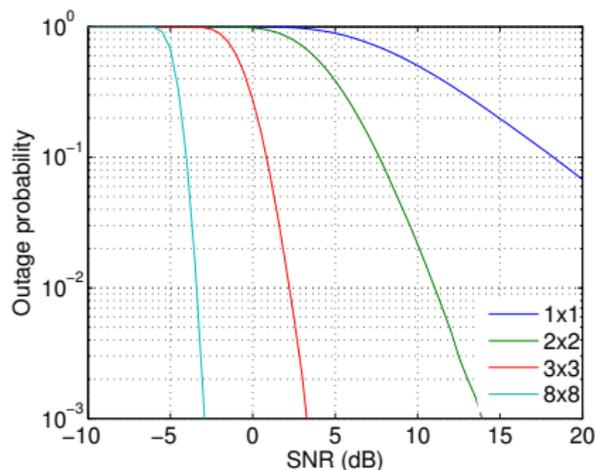
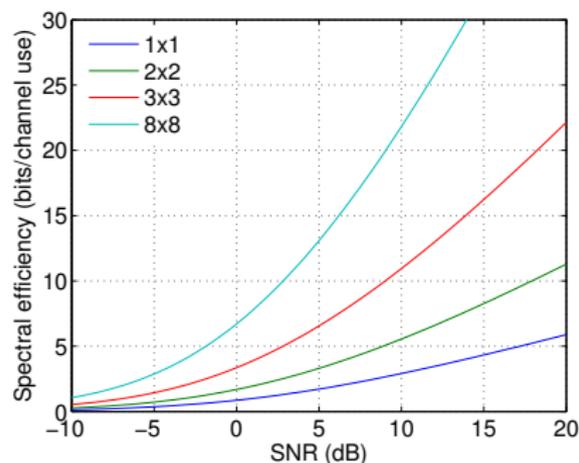
## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

# The benefits of MIMO



- Spatial multiplexing:  $C \sim \min(N_{\text{tx}}, N_{\text{rx}}) \log_2(\text{SNR})$  as  $\text{SNR} \rightarrow \infty$
- Diversity:  $P_{\text{out}} \sim \text{SNR}^{-N_{\text{tx}}N_{\text{rx}}}$
- $N_{\text{rx}}$ -fold receive and  $\frac{N_{\text{tx}}}{\min(N_{\text{rx}}, N_{\text{tx}})}$ -fold transmit beamforming gain
- Spatial division multiplexing (SDMA) or Multi-user (MU) MIMO



Source: [www.fotopedia.com/users/zlXpkbZmHZw](http://www.fotopedia.com/users/zlXpkbZmHZw)

**Massive MIMO is nothing new; but it takes MIMO to an entirely new level:**

*A base station (BS) with hundreds or even thousands of antennas simultaneously serves tens of user equipments (UEs) on the same time-frequency resource.*

# Outline

## 1 Massive MIMO

- **Benefits**
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

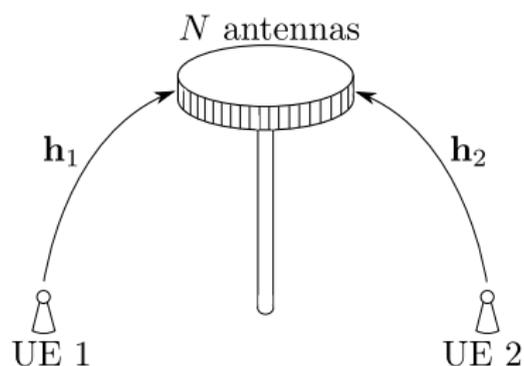
## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

## Massive MIMO: Uplink benefits



$$\mathbf{y} = \mathbf{h}_1 x_1 + \mathbf{h}_2 x_2 + \mathbf{n}$$

### Assumptions

- $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{C}^{N \times 1}$  have i.i.d. entries with zero mean and unit variance
- $\mathbf{h}_1, \mathbf{h}_2$  perfectly known at the base station (BS)
- $\mathbb{E}[|x_1|^2] = \mathbb{E}[|x_2|^2] = \text{SNR}$
- $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$

## Massive MIMO: Uplink benefits

The BS applies a simple matched filter to detect the symbol of UE 1:

$$\frac{1}{N} \mathbf{h}_1^H \mathbf{y} = \underbrace{x_1 \frac{1}{N} \sum_{i=1}^N |h_{1i}|^2}_{\text{useful signal}} + \underbrace{x_2 \frac{1}{N} \sum_{i=1}^N h_{1i}^* h_{2i}}_{\text{interference}} + \underbrace{\frac{1}{N} \sum_{i=1}^N h_{1i}^* n_i}_{\text{noise}}$$

By the strong law of large numbers:

$$\frac{1}{N} \sum_{i=1}^N h_{1i}^* h_{2i} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbb{E}[h_{11}^* h_{21}] = 0 \quad (\text{interference vanishes})$$

$$\frac{1}{N} \sum_{i=1}^N h_{1i}^* n_i \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbb{E}[h_{11}^* n_1] = 0 \quad (\text{noise vanishes})$$

Thus,

$$\frac{1}{N} \mathbf{h}_1^H \mathbf{y} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} x_1 \mathbb{E}[|h_{11}|^2] = x_1 \quad (\text{SNR can be made arbitrarily small})$$

## How fast can the SNR be scaled down with $N$ ?

The received SINR of UE 1 can be written as:

$$\text{SINR}_1^{\text{MF}} = \frac{\frac{1}{N} \mathbf{h}_1^H \mathbf{h}_1}{\frac{1}{N} \left| \frac{\mathbf{h}_1^H \mathbf{h}_2}{\|\mathbf{h}_1\|} \right|^2 + \frac{1}{N \text{SNR}}}$$

$$\text{Average signal power:} \quad \mathbb{E} \left[ \frac{1}{N} \mathbf{h}_1^H \mathbf{h}_1 \right] = 1$$

$$\text{Average interference power:} \quad \mathbb{E} \left[ \frac{1}{N} \left| \frac{\mathbf{h}_1^H \mathbf{h}_2}{\|\mathbf{h}_1\|} \right|^2 \right] = \frac{1}{N}$$

$$\text{Average noise power:} \quad \frac{1}{N \text{SNR}}$$

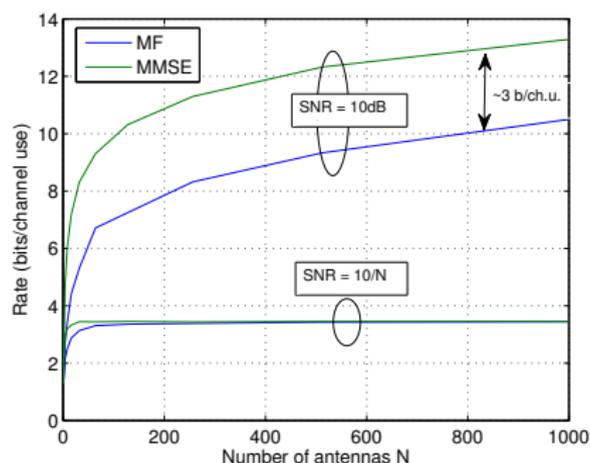
The SNR can be made inversely proportional to  $N$  without performance loss.

### Remark

The matched filter is optimal:

- 1 for any  $N$ , if  $\text{SNR} \rightarrow 0$ ,
- 2 for any SNR, if  $N \rightarrow \infty$ .

## Is the matched filter really optimal?



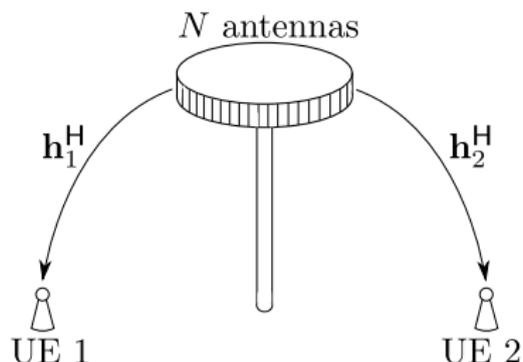
- With a MMSE detector the following ergodic rate for UE 1 is achievable

$$R_1^{\text{MMSE}} = \mathbb{E} \left[ \log_2 \left( \mathbf{1} + \mathbf{h}_1^H \left( \mathbf{h}_2 \mathbf{h}_2^H + \frac{1}{\text{SNR}} \mathbf{I}_N \right)^{-1} \mathbf{h}_1 \right) \right]$$

- How much do we gain over the matched filter  $R_1^{\text{MF}} = \mathbb{E} [\log_2(1 + \text{SINR}_1^{\text{MF}})]$ ?
- For  $\text{SNR} = \frac{P}{N}$ , one can show that  $\lim_{N \rightarrow \infty} R_1^{\text{MF}} = \lim_{N \rightarrow \infty} R_1^{\text{MMSE}} = \log_2(1 + P)$ .
- For  $\text{SNR} = P$ , we have  $\lim_{N \rightarrow \infty} R_1^{\text{MMSE}} - R_1^{\text{MF}} > 0$ .

The MF can be far from optimal [1], but can be implemented in a distributed manner.

## Massive MIMO: Downlink benefits



All of the previous conclusions hold similarly for the downlink [2]:

- Noise and interference vanish as  $N \rightarrow \infty$ .
- The transmit power per UE can be scaled down as  $1/N$  without performance loss.
- Maximum ratio transmission (MRT) is optimal if the power is scaled down.
- However, regularized zero-forcing can achieve the same performance with a significantly reduced number of antennas [1].

# Outline

## 1 Massive MIMO

- Benefits
- **Favorable propagation conditions**
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

## On favorable propagation conditions

Whenever  $\frac{1}{N} \mathbf{h}_1^H \mathbf{h}_2 \rightarrow 0$  as  $N \rightarrow \infty$ , we speak about *favorable propagation conditions* [3].

### Definition

To measure the “orthogonality” between two  $N$ -dimensional channel vectors, we define the *average correlation factor* as

$$\delta_N = \mathbb{E} \left[ \frac{|\mathbf{h}_1^H \mathbf{h}_2|^2}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2} \right] \in [0, 1].$$

How does  $\delta_N$  behave

- for i.i.d. fading channels?
- when there are line-of-sight components?
- with antenna correlation?
- channels obtained from ray-tracing?
- for measured channels?

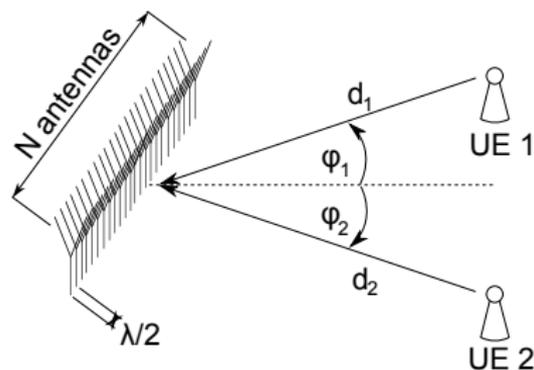
## Channel orthogonality for the i.i.d. model

Let  $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{C}^{N \times 1}$  have i.i.d. elements with zero mean, then

$$\delta_N^{\text{i.i.d.}} = \mathbb{E} \left[ \frac{|\mathbf{h}_1^H \mathbf{h}_2|^2}{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2} \right] = \frac{1}{N}.$$

Is this the best we can hope for?

## Channel orthogonality for LOS links



$$\mathbf{h}_i = \underbrace{a_i}_{\text{attenuation}} \underbrace{e^{-i2\pi \frac{d_i}{\lambda}}}_{\text{phase offset}} \underbrace{\mathbf{e}(\varphi_i)}_{\text{array response vector}}$$

- $\lambda$  is the wavelength
- uniform linear array:

$$\mathbf{e}(\varphi) = \left[ 1, e^{-i\pi \cos(\varphi)}, \dots, e^{-i(N-1)\pi \cos(\varphi)} \right]^T$$

One can show that (e.g., [4, (7.35)])

$$\delta_N^{\text{LOS}} = \left| \frac{\sin \left( N \frac{\pi}{2} [\cos(\varphi_2) - \cos(\varphi_1)] \right)}{N \sin \left( \frac{\pi}{2} [\cos(\varphi_2) - \cos(\varphi_1)] \right)} \right|^2 = \mathcal{O} \left( \frac{1}{N^2} \right), \quad \forall \varphi_1 \neq \varphi_2.$$

### Remark

For sufficiently large angular separation, the correlation factor decays much faster than for i.i.d. channels. Thus, LOS can be desirable for massive MIMO.

## Channel orthogonality with antenna correlation

Let

$$\mathbf{h}_1 = \mathbf{R}_1^{\frac{1}{2}} \mathbf{w}_1, \quad \mathbf{h}_2 = \mathbf{R}_2^{\frac{1}{2}} \mathbf{w}_2$$

- $\mathbf{w}_1, \mathbf{w}_2$  have i.i.d. elements, distributed as  $\mathcal{CN}(0, 1)$
- $\mathbf{R}_1, \mathbf{R}_2 \in \mathbb{C}^{N \times N}$  are deterministic Hermitian nonnegative-definite

One can show that

$$\delta_N^{\text{cor}} \approx \frac{\text{tr} \mathbf{R}_1 \mathbf{R}_2}{\text{tr} \mathbf{R}_1 \text{tr} \mathbf{R}_2}$$

Examples:

- $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{U} \mathbf{U}^H$ , where  $\mathbf{U} \in \mathbb{C}^{N \times L}$  has orthonormal columns. Then,  $\delta_N^{\text{cor}} = \frac{1}{L} > \delta_N^{\text{iid}}$ .
- Let  $\mathbf{R}_1, \mathbf{R}_2$  such that  $\text{tr} \mathbf{R}_1 \mathbf{R}_2 = 0$ . Then,  $\delta_N^{\text{cor}} = 0 \ll \delta_N^{\text{iid}}$ .

### Remark

Antenna correlation can either increase or decrease the channel orthogonality. This can be, for example exploited by smart user scheduling algorithms.

## Two useful results of random matrix theory

Lemma ([5, Lemma B.26], [6, Lemma 14.2])

Let  $\mathbf{A} \in \mathbb{C}^{N \times N}$  and  $\mathbf{x} = [x_1 \dots x_N]^T \in \mathbb{C}^N$  be a random vector of i.i.d. entries, independent of  $\mathbf{A}$ . Assume  $\mathbb{E}[x_i] = 0$ ,  $\mathbb{E}[|x_i|^2] = 1$ ,  $\mathbb{E}[|x_i|^8] < \infty$ , and  $\limsup_N \|\mathbf{A}\| < \infty$ , almost surely. Then,

$$\frac{1}{N} \mathbf{x}^H \mathbf{A} \mathbf{x} - \frac{1}{N} \text{tr} \mathbf{A} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 0.$$

Lemma ([6, Lemma 3.7])

Let  $\mathbf{y}$  be another independent random vector with the same distribution as  $\mathbf{x}$ . Then,

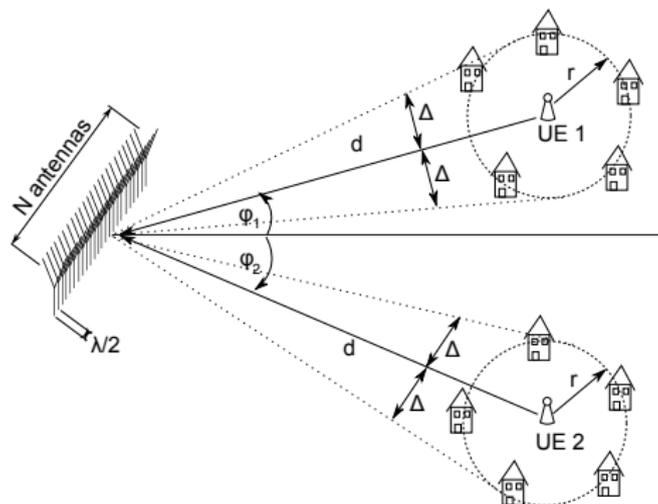
$$\frac{1}{N} \mathbf{x}^H \mathbf{A} \mathbf{y} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 0.$$

Remark

For  $\mathbf{A} = \mathbf{I}_N$ , these results are simple consequences of the strong law of large numbers:

$$\frac{1}{N} \mathbf{x}^H \mathbf{I}_N \mathbf{x} = \frac{1}{N} \sum_{i=1}^N |x_i|^2 \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbb{E}[|x_i|^2] = 1, \quad \frac{1}{N} \mathbf{x}^H \mathbf{I}_N \mathbf{y} = \frac{1}{N} \sum_{i=1}^N x_i^* y_i \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbb{E}[x_i^* y_i] = 0.$$

## A simple correlation model



- “One-ring” model (see, e.g., [7]): elevated BS, UE surrounded by local scatterers

- $$[\mathbf{R}_\ell]_{k,j} = \frac{1}{2\Delta} \int_{\varphi_\ell - \Delta}^{\varphi_\ell + \Delta} e^{-i\pi(k-j)\sin(\alpha)} d\alpha, \quad 1 \leq k, j, \leq N$$

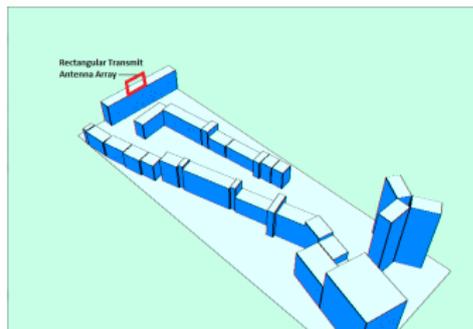
- It was recently shown [8] that  $\text{tr} \mathbf{R}_1 \mathbf{R}_2 \xrightarrow{N \rightarrow \infty} 0$  if

$$[\varphi_1 - \Delta, \varphi_1 + \Delta] \cap [\varphi_2 - \Delta, \varphi_2 + \Delta] = \emptyset.$$

- It also holds that  $\text{rank}(\mathbf{R}_\ell) \approx N |\cos(\varphi_\ell) \sin(\Delta)|$ ,  $\Delta \approx \tan^{-1}(r/d)$   
 Example:  $\varphi_\ell = 0^\circ$ ,  $r = 50$  m,  $d = 250$  m  $\Rightarrow \text{rank}(\mathbf{R}_\ell) \approx \frac{N}{5}$ .

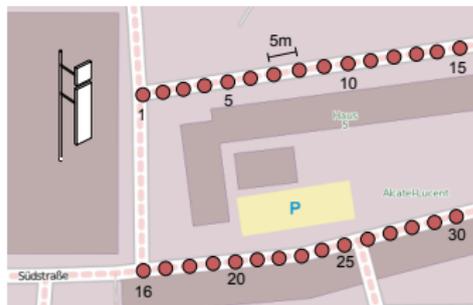
# Channel orthogonality: Ray tracing and measurements

## Ray tracing:



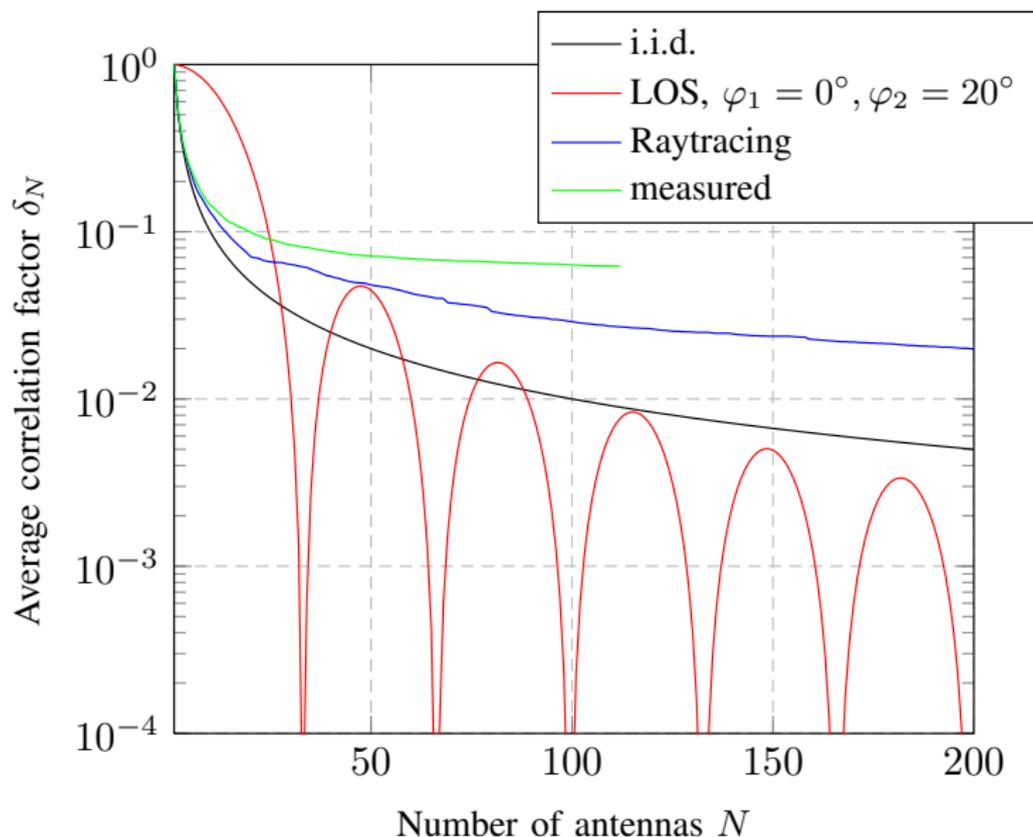
- Channel impulse responses created by ray tracing (WISE Tool [9])
- Very detailed 3D model of the ALU campus
- Average correlation factor computed over random position-pairs

## Channel measurements:

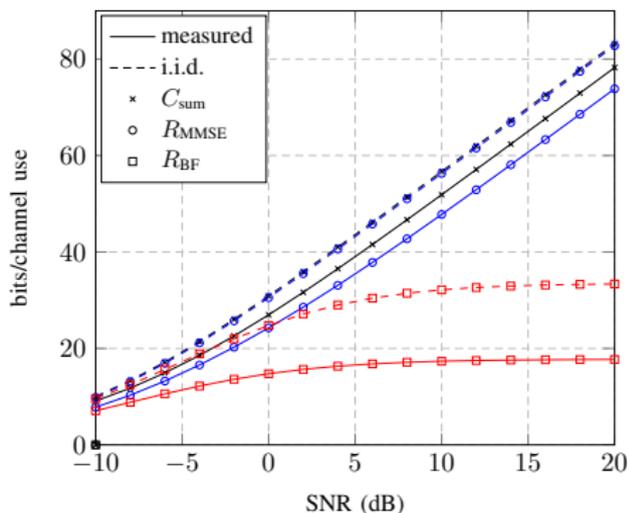
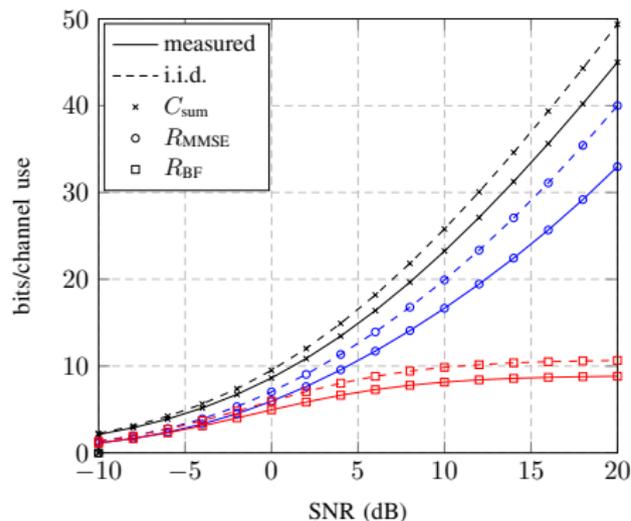


- Channel impulse responses taken from a measurement campaign on the ALU campus [10]
- Virtual antenna array with up to 112 antennas
- Average correlation factor computed over random pairs of measurement positions

## Channel orthogonality: Theory versus reality



## Impact on achievable rates



Achievable rates with different precoding schemes for a broadcast channel with  $K = 8$  UEs over i.i.d. and measured channels for  $N = 10$  (l.) and  $N = 112$  (r.) BS-antennas.

- 4 dB performance loss for the measured channels (MMSE precoding,  $N = 112$ )
- MRT (or beamforming (BF)) highly suboptimal for large  $N$  and SNR

## Lessons learned I

- 1 With very large antenna arrays, noise and interference vanish and the SNR can be made inversely proportional to the number of antennas  $N$  without performance loss. This holds for the uplink and downlink.
- 2 The above conclusions hold with perfect CSI at the BS and under favorable propagation characteristics, i.e., the channel vectors between to different UEs become orthogonal as  $N$  grows.
- 3 The channel orthogonality factors scales as  $1/N$  for the i.i.d. model and as  $1/N^2$  for pure LOS links.
- 4 Antenna correlation can have positive or negative effects depending on how UEs are scheduled. For a the one-ring model, the larger the angular separation between two UEs, the more orthogonal the subspaces spanned by the correlation matrices.
- 5 Ray tracing and channel measurements show that we can expect favorable propagation conditions in practice. However, the orthogonality scaling is slower than predicted by theory.

What happens if the channel must be estimated?  
What is the impact of hardware imperfections?

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- **Channel estimation and pilot contamination**
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

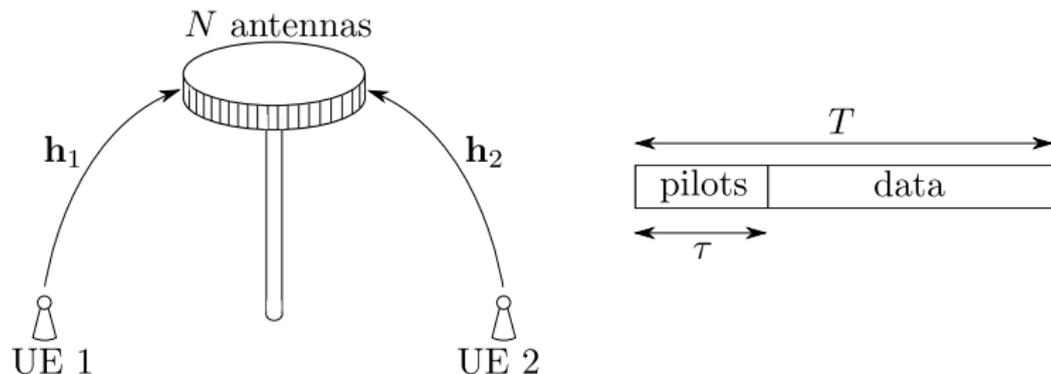
## Channel estimation and the role of TDD

- Channel knowledge at the BS is a must for precoding/beamforming and coherent detection.
- The channel coherence time fundamentally limits the number of BS-antennas and/or UEs:
  - ▶ FDD: Downlink training + feedback as well as uplink training necessary
    - Pilot overhead proportional to the number of antennas (unless certain channel conditions hold [8])
  - ▶ TDD: Only uplink training needed
    - Pilot overhead proportional to the number of UEs
    - We can add antennas at the BSs for “free” [11]

### Remark

TDD relies on reciprocity of the uplink and downlink channels. While reciprocity holds without doubt for the physical propagation channel, it does not without calibration for the transceiver RF chains. However, end-to-end reciprocity can be established by different internal and external calibration mechanisms, as has been demonstrated in practice (see, e.g., [12]).

## Uplink channel training



- The UEs transmit mutually orthogonal pilot sequences  $\phi_i \in \mathbb{C}^{\tau \times 1}$  of length  $\tau$ :

$$\mathbf{Y} = (\mathbf{h}_1 \mathbf{h}_2) \begin{pmatrix} \phi_1^\top \\ \phi_2^\top \end{pmatrix} + \mathbf{N}$$

where  $\mathbf{Y}, \mathbf{N} \in \mathbb{C}^{N \times \tau}$  and  $\text{vec}(\mathbf{N}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N\tau})$ .

- The BS correlates  $\mathbf{Y}$  with  $\frac{\phi_i^*}{\|\phi_i\|^2}$ :

$$\mathbf{Y} \frac{\phi_i^*}{\|\phi_i\|^2} = \mathbf{h}_i + \tilde{\mathbf{n}}_i$$

where  $\tilde{\mathbf{n}}_i \sim \mathcal{CN}(\mathbf{0}, \|\phi_i\|^{-2} \mathbf{I}_N)$ .

## Uplink channel training (cont.)

- Let  $\mathbf{h}_1, \mathbf{h}_2$  be independent and denote  $\mathbf{R}_i \triangleq \mathbb{E} [\mathbf{h}_i \mathbf{h}_i^H]$  for  $i = 1, 2$ .
- The BS calculates the linear MMSE estimate of  $\mathbf{h}_i$  based on  $\mathbf{Y} \frac{\phi_i^*}{\|\phi_i\|^2}$ :

$$\hat{\mathbf{h}}_i = \mathbf{R}_i \left( \mathbf{R}_i + \frac{1}{\|\phi_i\|^2} \mathbf{I}_N \right)^{-1} \mathbf{Y} \frac{\phi_i^*}{\|\phi_i\|^2}$$

- We can now decompose

$$\mathbf{h}_i = \hat{\mathbf{h}}_i + \tilde{\mathbf{h}}_i$$

where  $\mathbb{E} [\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^H] = \mathbf{R}_i \left( \mathbf{R}_i + \frac{1}{\|\phi_i\|^2} \mathbf{I}_N \right)^{-1} \mathbf{R}_i \triangleq \Phi_i$  and  $\mathbb{E} [\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H] = \mathbf{R}_i - \Phi_i$ .

- Due to the orthogonality principle of the MMSE estimator,  $\mathbb{E} [\hat{\mathbf{h}}_i^H \tilde{\mathbf{h}}_i] = 0$ . If  $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i)$ , then  $\hat{\mathbf{h}}_i \sim \mathcal{CN}(\mathbf{0}, \Phi_i)$  and  $\tilde{\mathbf{h}}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i - \Phi_i)$ .

### Assumptions for asymptotic considerations (as $N \rightarrow \infty$ ):

- $\text{tr} \mathbf{R}_i = c_i N$  for some  $0 < c_i \leq 1$  (linear energy growth)
- $\liminf_N \frac{1}{N} \text{rank}(\mathbf{R}_i) > 0$  (infinite degrees of freedom)
- $\limsup_N \|\mathbf{R}_i\| < \infty$  (finite energy per degree of freedom)

## Achievable rates with imperfect CSI

- During  $T - \tau$  channel uses, the UEs transmit their data:

$$\mathbf{y}(t) = \sum_{i=1}^2 \mathbf{h}_i x_i(t) + n(t) = \sum_{i=1}^2 (\hat{\mathbf{h}}_i + \tilde{\mathbf{h}}_i) x_i(t) + n(t)$$

- The BS estimates  $x_i(t)$  after the projection of  $\mathbf{y}(t)$  on  $\hat{\mathbf{h}}_i$ . An achievable rate for UE 1 is given as [13]

$$R_1 = \left(1 - \frac{\tau}{T}\right) \mathbb{E} \left[ \log \left( 1 + \frac{\|\hat{\mathbf{h}}_1\|^4}{\hat{\mathbf{h}}_1^H \left( \mathbf{h}_2 \mathbf{h}_2^H + \mathbb{E} [\tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^H | \hat{\mathbf{h}}_1] + \frac{1}{\text{SNR}} \mathbf{I}_N \right) \hat{\mathbf{h}}_1} \right) \right]$$

- For large  $N$ , we can find the following approximation:

$$R_1 \approx \left(1 - \frac{\tau}{T}\right) \log \left( 1 + \frac{\left(\frac{1}{N} \text{tr } \Phi_1\right)^2}{\underbrace{\frac{1}{N^2} \text{tr } \Phi_1 \mathbf{R}_2}_{\text{interference}} + \underbrace{\frac{1}{N^2} \text{tr } \Phi_1 (\mathbf{R}_1 - \Phi_1)}_{\text{imperfect CSI}} + \underbrace{\frac{1}{N \text{SNR}} \frac{1}{N} \text{tr } \Phi_1}_{\text{noise}}} \right)$$

## How does imperfect CSI affect the power scaling?

- Let  $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}_N$ .
- Then, one can show that:

$$R_1 \approx \left(1 - \frac{\tau}{T}\right) \log \left( 1 + \frac{1}{\underbrace{\frac{1}{N}}_{\text{interference}} + \underbrace{\frac{1}{\|\phi_1\|^2 N} \left(2 + \frac{1}{\text{SNR}}\right)}_{\text{imperfect CSI}} + \underbrace{\frac{1}{N \text{SNR}}}_{\text{noise}}} \right)$$

- Let  $\|\phi_1\|^2 \sim N^{-a}$ ,  $\text{SNR} \sim N^{-b}$
- Only if  $a + b \leq 1$ , we have  $\liminf_N R_1 > 0$ .

### Remark

With imperfect CSI, the rate at which the transmit power can be reduced with the number of antennas is smaller than for the case of perfect CSI. For equal pilot and data power, i.e.,  $a = b = \frac{1}{2}$ , the power can be made only proportional to  $1/\sqrt{N}$  in contrast to  $1/N$  for perfect CSI.

## Pilot contamination

- The number of orthogonal pilot sequences is limited by the channel coherence time. Thus, the pilot sequences must be reused in neighboring cells.
- Assume that both UEs transmit the same pilot sequence  $\phi_1 \in \mathbb{C}^{\tau \times 1}$ :

$$\mathbf{Y} = (\mathbf{h}_1 \mathbf{h}_2) \begin{pmatrix} \phi_1^T \\ \phi_1^T \end{pmatrix} + \mathbf{N}$$

- The BS correlates  $\mathbf{Y}$  with  $\frac{\phi_1^*}{\|\phi_1\|^2}$ :

$$\mathbf{Y} \frac{\phi_1^*}{\|\phi_1\|^2} = \mathbf{h}_1 + \mathbf{h}_2 + \tilde{\mathbf{n}}_1$$

- ...and calculates the linear MMSE estimate

$$\hat{\mathbf{h}}_1 = \mathbf{R}_1 \underbrace{\left( \mathbf{R}_1 + \mathbf{R}_2 + \frac{1}{\|\phi_1\|^2} \mathbf{I}_N \right)^{-1}}_{\triangleq \mathbf{s}_1} \mathbf{Y} \frac{\phi_1^*}{\|\phi_1\|^2}$$

- $\mathbb{E} \left[ \hat{\mathbf{h}}_1 \hat{\mathbf{h}}_1^H \right] \triangleq \Phi_1 = \mathbf{R}_1 \left( \mathbf{R}_1 + \mathbf{R}_2 + \frac{1}{\|\phi_1\|^2} \mathbf{I}_N \right)^{-1} \mathbf{R}_1$  and  $\mathbb{E} \left[ \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^H \right] = \mathbf{R}_1 - \Phi_1$ .

### Remark

The channel estimate  $\hat{\mathbf{h}}_1$  is correlated with  $\mathbf{h}_2$ . This effect is called pilot contamination.

## Achievable rate with pilot contamination

- Similar to the case without pilot contamination, an achievable rate is given by

$$R_1 = \left(1 - \frac{\tau}{T}\right) \mathbb{E} \left[ \log \left( 1 + \frac{\|\hat{\mathbf{h}}_1\|^4}{\hat{\mathbf{h}}_1^H \left( \mathbf{h}_2 \mathbf{h}_2^H + \mathbb{E} \left[ \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^H \right] + \frac{1}{\text{SNR}} \mathbf{I}_N \right) \hat{\mathbf{h}}_1} \right) \right]$$

$$\stackrel{(N \text{ large})}{\approx} \left(1 - \frac{\tau}{T}\right) \log(1 + \gamma_1)$$

where

$$\gamma_1 = \frac{\left(\frac{1}{N} \text{tr} \Phi_1\right)^2}{\underbrace{\frac{1}{N^2} \text{tr} \left( \mathbf{R}_1 + \frac{1}{\|\phi_1\|^2} \mathbf{I}_N \right) \mathbf{S}_1^H \mathbf{R}_2 \mathbf{S}_1}_{\text{interference}} + \underbrace{\frac{1}{N^2} \text{tr} \Phi_1 (\mathbf{R}_1 - \Phi_1)}_{\text{imperfect CSI}} + \underbrace{\frac{1}{N \text{SNR}} \frac{1}{N} \text{tr} \Phi_1}_{\text{noise}} + \underbrace{\left| \frac{1}{N} \text{tr} \mathbf{S}_1 \mathbf{R}_2 \right|^2}_{\text{pilot cont.}}}$$

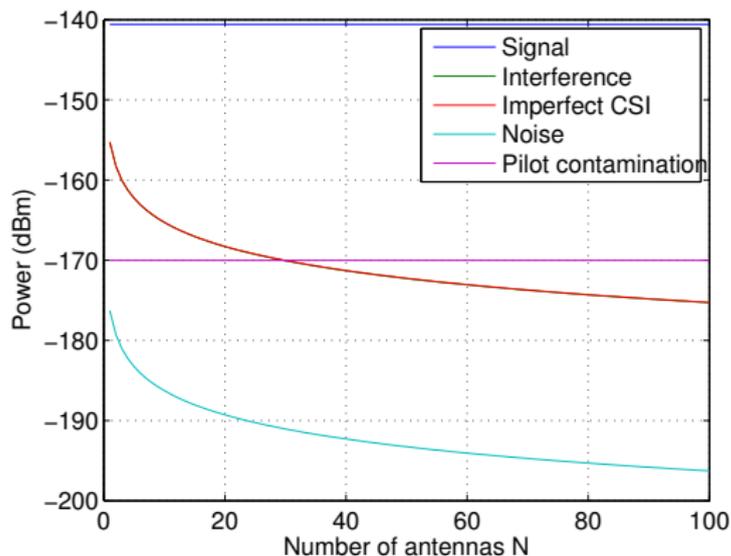
- As  $N \rightarrow \infty$ ,

$$\lim_N \gamma_1 = \lim_N \frac{\left(\frac{1}{N} \text{tr} \Phi_1\right)^2}{\left| \frac{1}{N} \text{tr} \mathbf{R}_1 \left( \mathbf{R}_1 + \mathbf{R}_2 + \frac{1}{\|\phi_1\|^2} \mathbf{I}_N \right)^{-1} \mathbf{R}_2 \right|^2} \stackrel{(\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}_N)}{=} 1$$

### Remark

As  $N \rightarrow \infty$ , pilot contamination can become a limiting factor.

## When does pilot contamination become dominating?



Powers of different parts of the received signal at the BS for  $\mathbf{R}_i = d_i^{-3.6} \mathbf{I}_N$  with  $d_1 = 200$  m,  $d_2 = 500$  m, and  $\|\phi\|^2 = \text{SNR} = 121$  dB.

## What can we do against pilot contamination?

- **Correlation helps [14]:** Let  $\mathbf{R}_1 = \mathbf{U}_1 \mathbf{D}_1 \mathbf{U}_1^H$  and  $\mathbf{R}_2 = \mathbf{U}_2 \mathbf{D}_2 \mathbf{U}_2^H$  such that  $\mathbf{U}_1^H \mathbf{U}_2 = \mathbf{0}_N$ . Then,  $R_1 \rightarrow \infty$  as  $N$  grows. *Idea:* Assign pilot sequences based on second-order channel statistics in order to reduce pilot contamination.
- **Blind or EVD-based channel estimation [15, 16]:** Let  $\mathbf{R}_i = \beta_i \mathbf{I}_N$  for  $i = 1, 2$  and assume  $\beta_1 > \beta_2$ . The BS observes the data transmissions during  $T$  time slots:

$$\mathbf{Y} = (\mathbf{h}_1 \mathbf{h}_2) \begin{pmatrix} x_1(1) & \dots & x_1(T) \\ x_2(1) & \dots & x_2(T) \end{pmatrix} + \mathbf{N}$$

where  $x_i(t)$ ,  $\forall i, t$ , are i.i.d with zero mean and unit variance and  $\mathbf{N}_{i,j} \sim \mathcal{CN}(0, 1)$ . Let  $\mathbf{Y} = [\mathbf{u}_1 \dots \mathbf{u}_N] \mathbf{D} \mathbf{V}^H$ . Then, as  $T, N \rightarrow \infty$ ,  $N/T \rightarrow c$  and  $\beta_1 > \sqrt{c}$ ,

$$\frac{1}{\sqrt{N}} \mathbf{u}_1^H \mathbf{h}_2 \xrightarrow[N, T \rightarrow \infty]{\text{a.s.}} 0, \quad \left| \frac{1}{\sqrt{N}} \mathbf{u}_1^H \mathbf{h}_1 \right|^2 \xrightarrow[N, T \rightarrow \infty]{\text{a.s.}} \beta_1.$$

Thus, the BS can detect  $x_1(t)$  either blindly based on  $\mathbf{u}_1^H \mathbf{Y}$  or estimate the effective channel  $\mathbf{u}_1^H \mathbf{h}_1$  from pilots without pilot contamination.

- **Other techniques:** Pilot contamination precoding (multi-cell precoding based on channel statistics and user data sharing) [17], Time-shifted pilots (UEs in different cells transmit the pilots at different times) [18], many more...

Pilot contamination is not a fundamental limitation and can be efficiently mitigated!

## What about FDD?

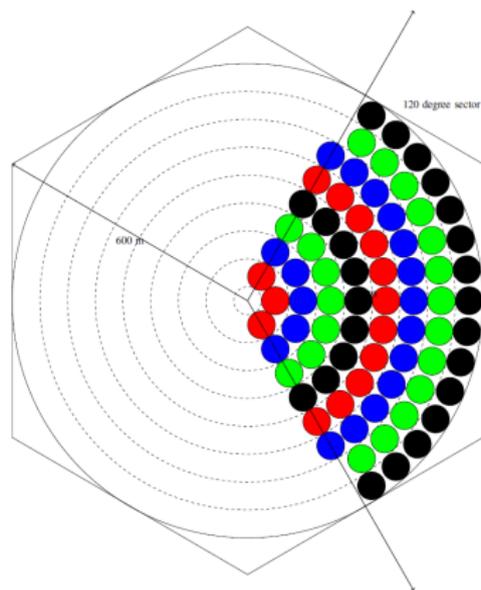
- Channel estimation for the uplink is essentially the same as for TDD.
- Estimating the downlink channel at the BS requires downlink training and feedback from the UEs which is proportional to  $N$ .
- Antenna correlation can be exploited to reduce this overhead!
- Let the received signal at a UE be

$$y = \mathbf{h}^H \mathbf{w}x + n$$

where  $\mathbf{w} \in \mathbb{C}^{N \times 1}$  is a precoding vector,  $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ ,  $\text{rank}(\mathbf{R}) = L$ , with SVD  $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^H$  and  $\mathbf{U} \in \mathbb{C}^{N \times L}$ .

- Assuming that  $\mathbf{R}$  is known, the BS can apply an “outer”-precoder  $\mathbf{U}$  which reduces the  $N$ -dimensional channel  $\mathbf{h}$  to the  $L$ -dimensional effective channel  $\mathbf{U}^H \mathbf{h}$ , i.e.,  $\mathbf{w} = \mathbf{U}\tilde{\mathbf{w}}$ ,  $\tilde{\mathbf{w}} \in \mathbb{C}^{L \times 1}$ . The UE only estimates and feeds back the coefficients of the effective channel. This overhead is proportional to  $L$ .
- This idea can be extended to multiple UEs [8]: Group users with similar covariance matrices together, use MU-MIMO on the effective channel to separate the UEs of a group, simultaneously schedule groups with almost orthogonal covariance matrices. The number of UEs per group is limited by the rank of the covariance matrix.

## Joint Space-Division and Multiplexing (JSDM)



An idea from [8]: UEs in different circles of the same color have covariance matrices which span (nearly) orthogonal sub-spaces. They are simultaneously served by a combination of a deterministic pre-beamforming matrix and MU-MIMO schemes based on the reduced-dimensional effective channel. The BS is elevated from the ground and has a 2-dimensional array which makes it possible to separate UE-clusters in the azimuth and elevation angular domain.

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- **Hardware impairments**
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

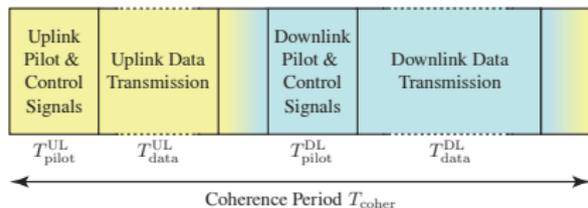
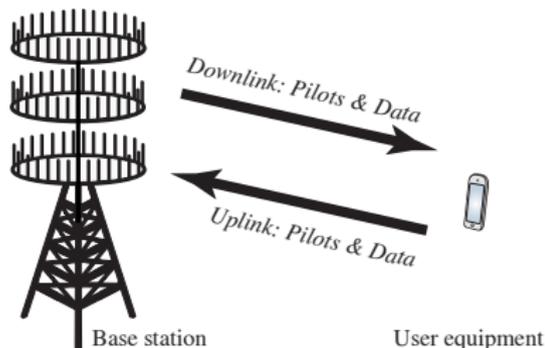
## Hardware impairments

- Any transceiver suffers from hardware impairments which
  - ▶ create a mismatch between the intended and emitted signal
  - ▶ distort the received signal
- The distortions depend in general on the transmitted or received power.
- Sources of impairments are:
  - ▶ oscillator phase noise
  - ▶ amplifier non-linearity (especially for OFDM with high PAPR)
  - ▶ IQ imbalance due to imperfections in the quadrature mixer
  - ▶ quantization noise
- Hardware impairments are known to limit the performance in the high-SNR regime but much less is known for the large- $N$  regime [19].
- For massive MIMO cheap, low-power, and low-cost transceivers are desirable.

What are the impacts of hardware imperfections?  
Can the hardware quality be reduced by increasing  $N$ ?

*All of the results in this part are taken from [20, 21].*

# Uplink system and channel model



$$\text{Uplink : } \mathbf{y}^{\text{BS}} = \mathbf{h} \left( x^{\text{UE}} + \eta_t^{\text{UE}} \right) + \boldsymbol{\eta}_r^{\text{BS}} + \mathbf{n}^{\text{BS}}$$

where  $\mathbb{E} \left[ |x^{\text{UE}}|^2 \right] = \rho^{\text{UE}}$ ,  $\mathbf{n}^{\text{BS}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{S})$ , and  $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ .

## Additive distortion

- UE transmit distortion:  $\eta_t^{\text{UE}} \sim \mathcal{CN}(0, \kappa_t^{\text{UE}} \rho^{\text{UE}})$
- BS receive distortion:  $\boldsymbol{\eta}_r^{\text{BS}} \sim \mathcal{CN}(\mathbf{0}, \kappa_r^{\text{BS}} \rho^{\text{UE}} \text{diag}(|h_1|^2, \dots, |h_N|^2))$

The parameters  $\kappa_t^{\text{UE}}$  and  $\kappa_r^{\text{BS}}$  are usually in the range  $[0, 0.03]$ . The smaller these values, the more accurate and expensive is the corresponding transceiver hardware component.

## Channel estimation with hardware imperfections

- The BSs computes the linear MMSE estimate of  $\mathbf{h}$  based on a single uplink pilot  $x^{\text{UE}} = d$ ,  $|d|^2 = p^{\text{UE}}$  :

$$\hat{\mathbf{h}} = d^* \mathbf{R} \mathbf{Z}^{-1} \mathbf{y}^{\text{BS}}$$

where  $\mathbf{R}_{ii}$  is the  $i$ th diagonal element of  $\mathbf{R}$  and

$$\mathbf{Z} = \mathbb{E} \left[ \mathbf{y}^{\text{BS}} (\mathbf{y}^{\text{BS}})^{\text{H}} \right] = p^{\text{UE}} \left( 1 + \kappa_t^{\text{UE}} \right) \mathbf{R} + p^{\text{UE}} \kappa_r^{\text{BS}} \text{diag}(\mathbf{R}_{11}, \dots, \mathbf{R}_{NN}) + \mathbf{S}.$$

- We can decompose the channel as  $\mathbf{h} = \hat{\mathbf{h}} + \tilde{\mathbf{h}}$ , where

$$\mathbb{E} \left[ \hat{\mathbf{h}} \hat{\mathbf{h}}^{\text{H}} \right] = \mathbf{R} - \mathbf{C}$$

$$\mathbb{E} \left[ \tilde{\mathbf{h}} \tilde{\mathbf{h}}^{\text{H}} \right] = \mathbf{C}$$

$$\mathbf{C} = \mathbf{R} - p^{\text{UE}} \mathbf{R} \mathbf{Z}^{-1} \mathbf{R}.$$

### Remarks

- Due to the distortion noise, the MMSE estimator is very complicated to derive and not identical to the LMMSE estimator.
- $\hat{\mathbf{h}}$  and  $\tilde{\mathbf{h}}$  are neither independent nor jointly complex Gaussian, but only uncorrelated with zero mean.

## Channel estimation with hardware imperfections: Insights

### Impact of pilot power

For  $\mathbf{R} = \mathbf{S} = \mathbf{I}_N$ , we have

$$\mathbf{C} = \left( 1 - \frac{1}{1 + \kappa_t^{\text{UE}} + \kappa_r^{\text{BS}} + \frac{1}{\rho^{\text{UE}}}} \right) \mathbf{I}_N, \quad \lim_{\rho^{\text{UE}} \rightarrow \infty} \mathbf{C} = \left( 1 - \frac{1}{1 + \kappa_t^{\text{UE}} + \kappa_r^{\text{BS}}} \right) \mathbf{I}_N.$$

Thus, even with infinite pilot power, perfect estimation accuracy cannot be achieved and the performance is limited by the sum of the hardware impairments at the UE and BS.

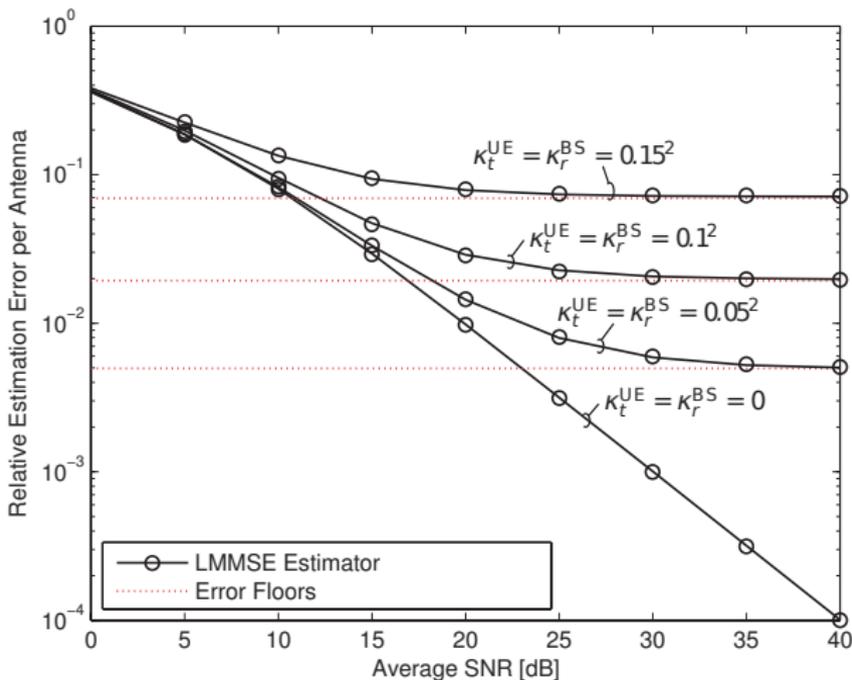
### Impact of pilot length

If the UE sends  $B$  pilot tones, the estimation accuracy can be improved by averaging  $B$  separate LMMSE estimates

$$\hat{\mathbf{h}} = \frac{1}{B} \sum_{i=1}^B \hat{\mathbf{h}}_i = \mathbf{h} - \frac{1}{B} \sum_{i=1}^B \tilde{\mathbf{h}}_i.$$

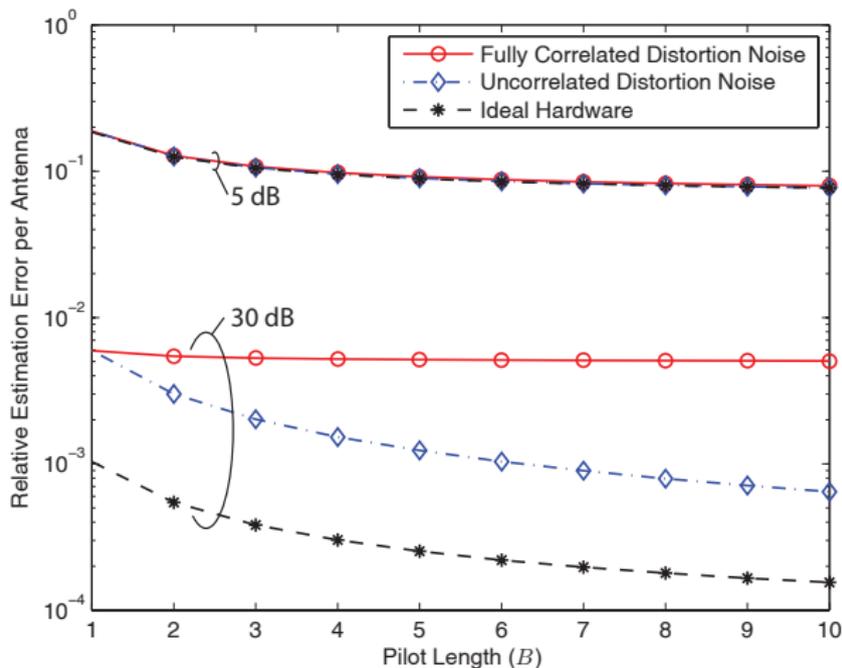
If the errors  $\tilde{\mathbf{h}}_i$  are uncorrelated, they average out as  $B$  increases. However, the distortion might be correlated and increasing  $B$  decreases the time for data transmission.

## Channel estimation with hardware imperfections: Numerical results



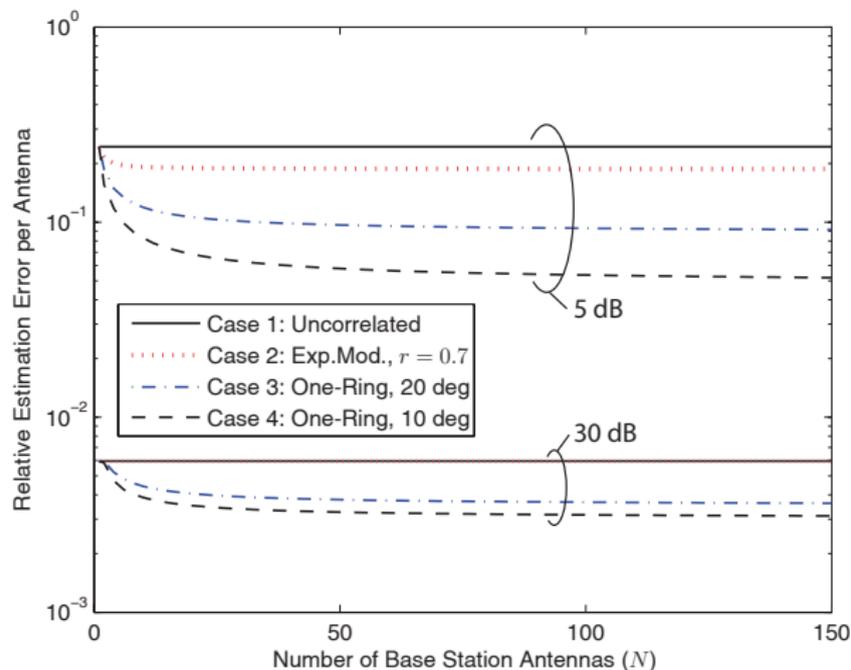
Relative estimation error  $\frac{\text{tr} \mathbf{C}}{\text{tr} \mathbf{R}}$  for  $N = 50$  over  $\text{SNR} = \rho^{\text{UE}} \frac{\text{tr} \mathbf{R}}{\text{tr} \mathbf{S}}$ . The matrix  $\mathbf{R}$  is generated by the exponential model [22] with correlation coefficient  $\rho = 0.7$  and  $\mathbf{S} = \mathbf{I}_N$ .

# Channel estimation with hardware imperfections: Numerical results



Relative estimation error  $\frac{\text{tr} \mathbf{C}}{\text{tr} \mathbf{R}}$  over the pilot length  $B$  for for  $N = 50$  and  $\kappa_t^{\text{UE}} = \kappa_r^{\text{BS}} = 0.05^2$ .

## Channel estimation with hardware imperfections: Numerical results

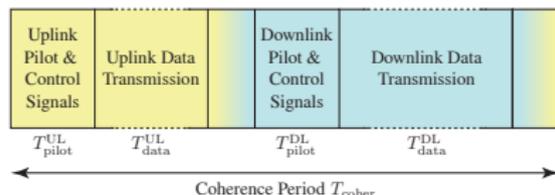
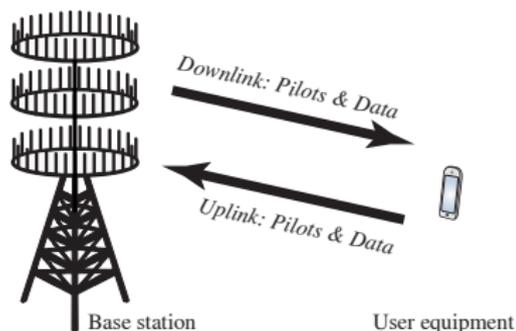


Relative estimation error  $\frac{\text{tr} \mathbf{C}}{\text{tr} \mathbf{R}}$  over  $N$  for  $\kappa_t^{\text{UE}} = \kappa_r^{\text{BS}} = 0.05^2$ .

## Channel estimation with hardware imperfections: Discussion

- Perfect channel estimation with hardware imperfections is impossible.
- The estimation accuracy is fundamentally limited by
  - ▶ Distortions for high SNR
  - ▶ Coherence time and correlation of distortion noise for long pilot sequences
- Hardware qualities of the UE and the BS are equally important.
- The channel model largely impacts the estimation accuracy: Correlated channels are easier to estimate [23].
- For large  $N$ , accurate estimates of  $\mathbf{R}$  and  $\mathbf{S}$  are difficult to obtain. The impact on the estimation performance is unclear.
- In practice [24], one would expect additive and multiplicative distortions. The latter are more difficult to analyze and can make things only worse.

## Downlink channel model with hardware imperfections



$$\text{Downlink : } y^{\text{UE}} = \mathbf{h}^{\text{H}} \left( \mathbf{w}x^{\text{BS}} + \boldsymbol{\eta}_t^{\text{BS}} \right) + \eta_r^{\text{UE}} + n^{\text{UE}}$$

where  $\mathbb{E} \left[ |x^{\text{BS}}|^2 \right] = p^{\text{UE}}$ ,  $\|\mathbf{w}\| = 1$ , and  $n^{\text{UE}} \sim \mathcal{CN}(0, \sigma_{\text{UE}}^2)$ .

### Additive distortion

- BS transmit distortion:  $\boldsymbol{\eta}_t^{\text{BS}} \sim \mathcal{CN}(\mathbf{0}, \kappa_t^{\text{BS}} p^{\text{BS}} \text{diag}(|w_1|^2, \dots, |w_N|^2))$
- UE receive distortion:  $\eta_r^{\text{UE}} \sim \mathcal{CN}(0, \kappa_r^{\text{UE}} p^{\text{BS}} |\mathbf{h}^{\text{H}} \mathbf{w}|)$

### Remark

Despite channel reciprocity, the UE needs to estimate the effective SISO channel  $\mathbf{h}^{\text{H}} \mathbf{w}$ .

## Capacity upper bounds

Assuming perfect CSI at the BS and UE and Gaussian signaling, one can show that the ergodic UL and DL capacities are bounded from above by

$$\begin{aligned} C^{\text{UL}} &\leq \frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} \mathbb{E} \left[ \log_2 \left( 1 + \frac{\mathbf{h}^H \left( \kappa_r^{\text{BS}} \mathbf{D}_{\mathbf{h}} + \frac{1}{\rho^{\text{UE}}} \mathbf{S} \right)^{-1} \mathbf{h}}{1 + \kappa_t^{\text{UE}} \mathbf{h}^H \left( \kappa_r^{\text{BS}} \mathbf{D}_{\mathbf{h}} + \frac{1}{\rho^{\text{UE}}} \mathbf{S} \right)^{-1} \mathbf{h}} \right) \right] \\ &\leq C_{\text{upper}}^{\text{UL}} \triangleq \frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} \log_2 \left( 1 + \frac{G^{\text{UL}}}{1 + \kappa_t^{\text{UE}} G^{\text{UL}}} \right) \end{aligned}$$

$$\begin{aligned} C^{\text{DL}} &\leq \frac{T_{\text{data}}^{\text{DL}}}{T_{\text{coher}}} \mathbb{E} \left[ \log_2 \left( 1 + \frac{\mathbf{h}^H \left( \kappa_t^{\text{BS}} \mathbf{D}_{\mathbf{h}} + \frac{\sigma_{\text{UE}}^2}{\rho^{\text{BS}}} \mathbf{I} \right)^{-1} \mathbf{h}}{1 + \kappa_r^{\text{UE}} \mathbf{h}^H \left( \kappa_t^{\text{BS}} \mathbf{D}_{\mathbf{h}} + \frac{\sigma_{\text{UE}}^2}{\rho^{\text{BS}}} \mathbf{I} \right)^{-1} \mathbf{h}} \right) \right] \\ &\leq C_{\text{upper}}^{\text{DL}} \triangleq \frac{T_{\text{data}}^{\text{DL}}}{T_{\text{coher}}} \log_2 \left( 1 + \frac{G^{\text{DL}}}{1 + \kappa_r^{\text{UE}} G^{\text{DL}}} \right) \end{aligned}$$

where  $G^{\text{UL}} = \mathbb{E} \left[ \mathbf{h}^H \left( \kappa_r^{\text{BS}} \mathbf{D}_{\mathbf{h}} + \frac{1}{\rho^{\text{UE}}} \mathbf{S} \right)^{-1} \mathbf{h} \right]$ ,  $G^{\text{DL}} = \mathbb{E} \left[ \mathbf{h}^H \left( \kappa_t^{\text{BS}} \mathbf{D}_{\mathbf{h}} + \frac{\sigma_{\text{UE}}^2}{\rho^{\text{BS}}} \mathbf{I} \right)^{-1} \mathbf{h} \right]$ , and  $\mathbf{D}_{\mathbf{h}} = \text{diag}(|h_1|^2, \dots, |h_N|^2)$ .

## Asymptotic behavior of the upper bounds

### High-SNR regime

$$\lim_{\rho^{\text{UE}} \rightarrow \infty} C_{\text{upper}}^{\text{UL}} = \frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} \log_2 \left( 1 + \frac{N}{\kappa_r^{\text{BS}} + N\kappa_t^{\text{UE}}} \right)$$

$$\lim_{\rho^{\text{BS}} \rightarrow \infty} C_{\text{upper}}^{\text{DL}} = \frac{T_{\text{data}}^{\text{DL}}}{T_{\text{coher}}} \log_2 \left( 1 + \frac{N}{\kappa_t^{\text{BS}} + N\kappa_r^{\text{UE}}} \right)$$

### Large- $N$ regime

$$\lim_{N \rightarrow \infty} C_{\text{upper}}^{\text{UL}} = \frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} \log_2 \left( 1 + \frac{1}{\kappa_t^{\text{UE}}} \right), \quad \lim_{N \rightarrow \infty} C_{\text{upper}}^{\text{DL}} = \frac{T_{\text{data}}^{\text{DL}}}{T_{\text{coher}}} \log_2 \left( 1 + \frac{1}{\kappa_r^{\text{UE}}} \right)$$

### Observations

- The UL/DL capacities have finite ceilings which depend on the hardware quality.
- The UE impairments are  $N$  times more influential than the BS impairments.
- As  $N \rightarrow \infty$ , only the UE hardware limits the performance; the BS distortion averages out.

## Capacity lower bounds

One can show that the ergodic UL/DL capacities are bounded from below by ([25, 13])

$$C^{\text{UL}} \geq C_{\text{lower}}^{\text{UL}} = \frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} \log_2 \left( 1 + \text{SINR}_{\text{lower}}^{\text{UL}} \right)$$
$$C^{\text{DL}} \geq C_{\text{lower}}^{\text{DL}} = \frac{T_{\text{data}}^{\text{DL}}}{T_{\text{coher}}} \log_2 \left( 1 + \text{SINR}_{\text{lower}}^{\text{DL}} \right)$$

where  $\mathbf{v}^{\text{UL}}, \mathbf{v}^{\text{DL}}$  are functions of  $\hat{\mathbf{h}}$  and

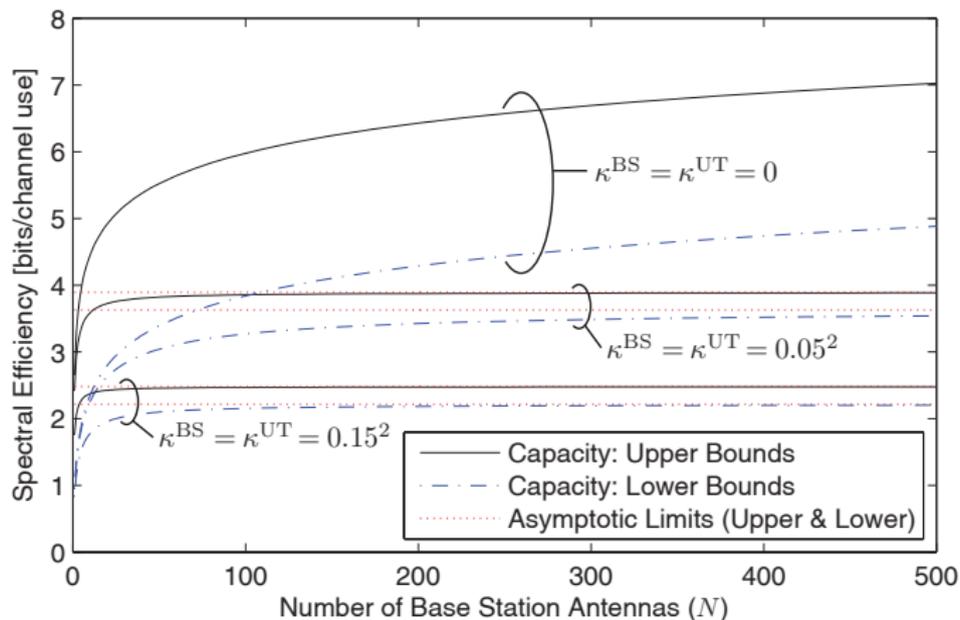
$$\text{SINR}_{\text{lower}}^{\text{UL}} = \frac{|\mathbb{E}[\mathbf{h}^H \mathbf{v}^{\text{UL}}]|^2}{(1 + \kappa_t^{\text{UE}}) \mathbb{E}[|\mathbf{h}^H \mathbf{v}^{\text{UL}}|^2] - |\mathbb{E}[\mathbf{h}^H \mathbf{v}^{\text{UL}}]|^2 + \kappa_r^{\text{BS}} \sum_{i=1}^N \mathbb{E}[|h_i|^2 |v_i^{\text{UL}}|^2] + \frac{\mathbb{E}[(\mathbf{v}^{\text{UL}})^H \mathbf{S} \mathbf{v}^{\text{UL}}]}{\rho^{\text{UE}}}}$$

$$\text{SINR}_{\text{lower}}^{\text{DL}} = \frac{|\mathbb{E}[\mathbf{h}^H \mathbf{v}^{\text{DL}}]|^2}{(1 + \kappa_t^{\text{UE}}) \mathbb{E}[|\mathbf{h}^H \mathbf{v}^{\text{DL}}|^2] - |\mathbb{E}[\mathbf{h}^H \mathbf{v}^{\text{DL}}]|^2 + \kappa_t^{\text{BS}} \sum_{i=1}^N \mathbb{E}[|h_i|^2 |v_i^{\text{DL}}|^2] + \frac{\sigma_{\text{UE}}^2}{\rho^{\text{BS}}}}$$

### Remarks

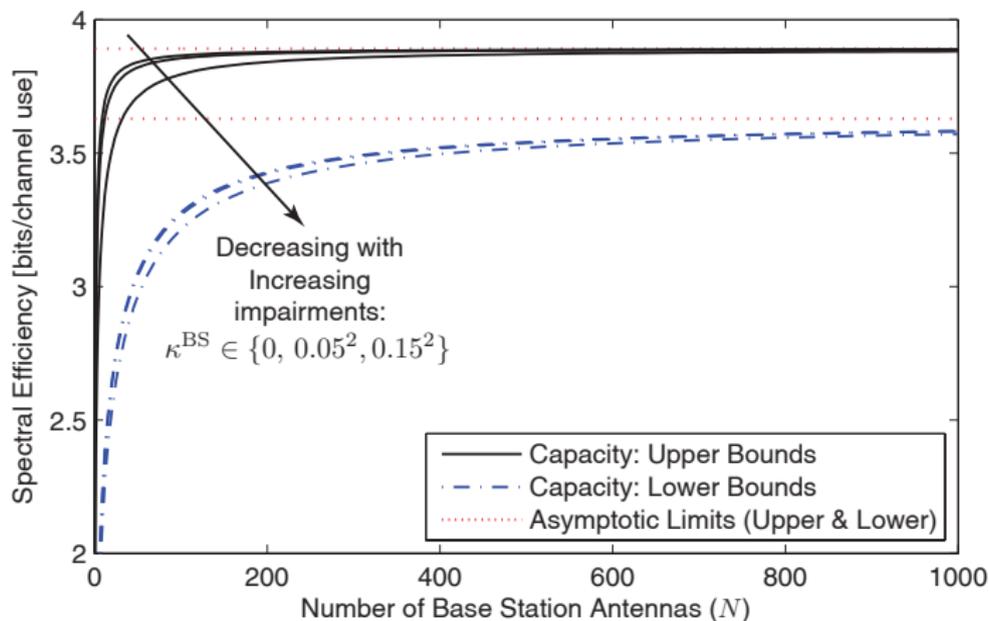
- The bounds hold for any  $\mathbf{v}^{\text{UL}}, \mathbf{v}^{\text{DL}}$ ; especially for  $\mathbf{v}^{\text{UL}} = \mathbf{v}^{\text{DL}} = \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}$ .
- $C_{\text{lower}}^{\text{UL}}$  depends only on  $\kappa_t^{\text{UE}}, \kappa_r^{\text{BS}}$  while  $C_{\text{lower}}^{\text{DL}}$  depends on all  $\kappa$ -parameters.
- The impact of  $\kappa_t^{\text{BS}}$  vanishes as  $N \rightarrow \infty$ .
- One can show that, if  $\kappa_t^{\text{UE}} = 0$ ,  $\lim_{N \rightarrow \infty} C_{\text{upper}}^{\text{DL}} - C_{\text{lower}}^{\text{DL}} = 0$ . This does not happen for  $\kappa_r^{\text{BS}} = 0$ . Thus, the capacity limit is mainly determined by the impairments at the UE.

## Capacity bounds: Numerical results



Lower and upper bounds on the capacity for  $\mathbf{R} = \mathbf{S} = \mathbf{I}$

## Capacity bounds: Numerical results



Lower and upper bounds on the capacity for  $\mathbf{R} = \mathbf{S} = \mathbf{I}$  and  $\kappa_t^{\text{UE}} = \kappa_r^{\text{UE}} = 0.05^2$ . The impact of the hardware impairments at the BS vanishes asymptotically.

## Scaling down the power or hardware quality

- **Uplink:** Let  $p^{\text{UE}} \sim \frac{1}{N^t}$ ,  $0 < t < \frac{1}{2}$  or  $\kappa_r^{\text{BS}} \sim N^t$ ,  $0 < t < \frac{1}{4}$ . Then,

$$\lim_{N \rightarrow \infty} C^{\text{UL}} \geq \log_2 \left( 1 + \frac{1}{2\kappa_t^{\text{UE}} + (\kappa_t^{\text{UE}})^2} \right)$$

- **Downlink:** Let  $p^{\text{BS}} \sim \frac{1}{N^{t_{\text{BS}}}}$ ,  $p^{\text{UE}} \sim \frac{1}{N^{t_{\text{UE}}}}$ ,  $t_{\text{BS}} + t_{\text{UE}} < 1$ ,  $t_{\text{BS}} \geq 0$ ,  $0 < t_{\text{UE}} < \frac{1}{2}$  or  $\kappa_r^{\text{BS}} \sim N^{t_1}$ ,  $\kappa_t^{\text{BS}} \sim N^{t_2}$ ,  $0 < t_1 + t_2 < \frac{1}{2}$ . Then,

$$\lim_{N \rightarrow \infty} C^{\text{DL}} \geq \log_2 \left( 1 + \frac{1}{\kappa_r^{\text{UE}} + \kappa_t^{\text{UE}} + \kappa_r^{\text{UE}} \kappa_t^{\text{UE}}} \right)$$

### Remark

The transmit powers can be roughly reduced as  $\frac{1}{\sqrt{N}}$  or the hardware quality of the BS transceiver can be reduced as  $\frac{1}{\sqrt{\sqrt{N}}}$  while still a non-zero capacity is achieved.

## Lessons learned II

- With TDD, the pilot overhead is independent of  $N$ .
- The effects of imperfect CSI vanish as  $N \rightarrow \infty$ .
- The transmit power can be made proportional to  $1/\sqrt{N}$  when channel estimation is taken into account (in contrast to  $1/N$  with perfect CSI). This holds even with hardware impairments.
- Pilot contamination can be a performance bottleneck, but it is not a fundamental limitation. It can be mitigated by blind channel estimation schemes, scheduling, or precoding.
- Hardware imperfections seem to be a fundamental limitation.
- The quality of the UE-transceiver is more important than that of the BS-transceiver.
- The quality of the BS-transceivers can be decreased as  $N$  grows (roughly proportional to  $N^{-\frac{1}{4}}$ ).

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- **Research topics**

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

## Some interesting areas for future research

- Channel modeling: Measurements, ray-tracing, new statistical models [26, 10]
- New deployment models: Distributed massive MIMO, antennas in building facades or windows [27], massive MIMO in HetNets [28, 29, 30]
- New applications for massive MIMO: Wireless backhaul [31], sensor networks
- Estimation of covariance matrices: Impact for channel estimation, precoding/detection
- Hardware impairments: Fundamental limits and ways to mitigate them [20]
- Cost and impacts of TDD reciprocity calibration
- Combination of massive MIMO and stochastic geometry [32, 33]
- Total energy efficiency of massive MIMO systems [34, 28, 20, 21]

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

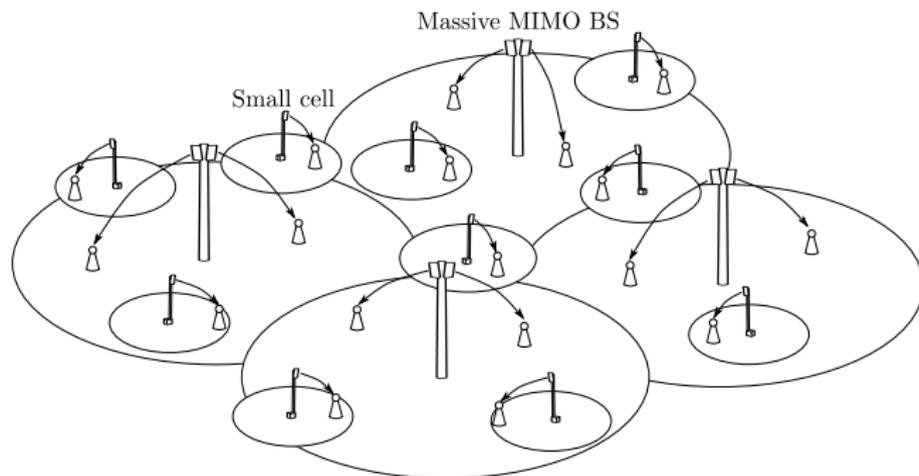
- Motivation and advantages
- Power and “antenna” minimization problem

## Massive MIMO versus Small Cells

- From a coverage as well as area spectral efficiency point of view, it is preferable to distribute the available antennas as much as possible [35].
- However, with small cells deployed below the roof tops, it is difficult to
  - ▶ ensure coverage
  - ▶ support highly mobile UEs (due to frequent handovers)
- But, massive MIMO is *particularly* suited to
  - ▶ provide coverage of large areas
  - ▶ support highly mobile UEs (TDD reduces the turnaround time)

Can we integrate the complementary benefits of both?

## A two-tier network architecture



- Massive MIMO BSs overlaid with many small cells (SCs)
- BSs ensure coverage and serve highly mobile UEs
- SCs drive the capacity (hot spots, indoor, etc.)

There is a large number of *excess antennas* in the network, which should be exploited.

## The role of TDD

A network-wide synchronized TDD protocol has the following advantages:

- The downlink channels can be estimated from uplink pilots.
  - Enable massive MIMO
- Channel reciprocity holds for the desired *and* interfering channels.
  - Knowledge about the interfering channels can be acquired for free at every device (BS, SC, UE).

Idea: “*Sacrifice*” excess antennas to reduce intra- and inter-tier interference.

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

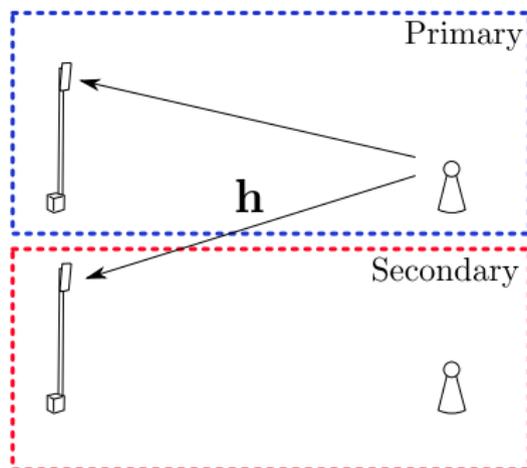
## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- **An idea from cognitive radio**
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

## A simple idea from cognitive radio [36]



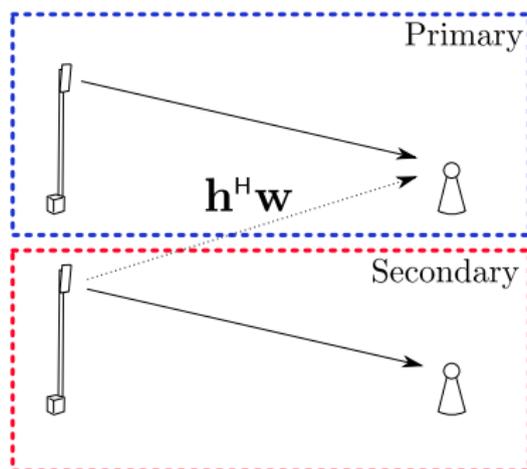
- 1 The secondary BS listens to the transmission from the primary UE:

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

- 2 ...and computes the covariance matrix of the receive signal:

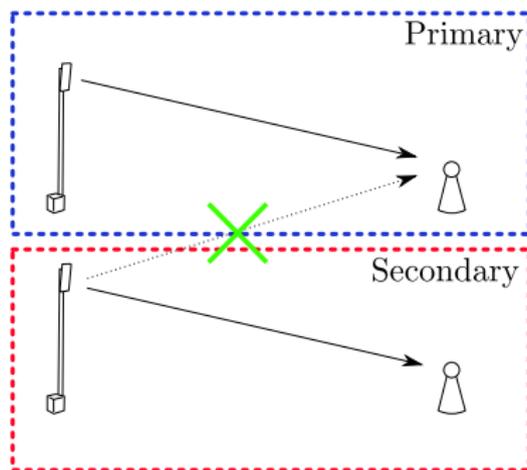
$$\mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{h}\mathbf{h}^H + \text{SNR}^{-1}\mathbf{I}$$

## A simple idea from cognitive radio [36]



- ③ With the knowledge of the SNR, the secondary BS designs a precoder  $\mathbf{w}$  which is orthogonal to the sub-space spanned by  $\mathbf{h}\mathbf{h}^H$ .

## A simple idea from cognitive radio [36]



- ③ With the knowledge of the SNR, the secondary BS designs a precoder  $\mathbf{w}$  which is orthogonal to the sub-space spanned by  $\mathbf{h}\mathbf{h}^H$ .
- ④ The interference to the primary UE can be entirely eliminated without explicit knowledge of  $\mathbf{h}$ .

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

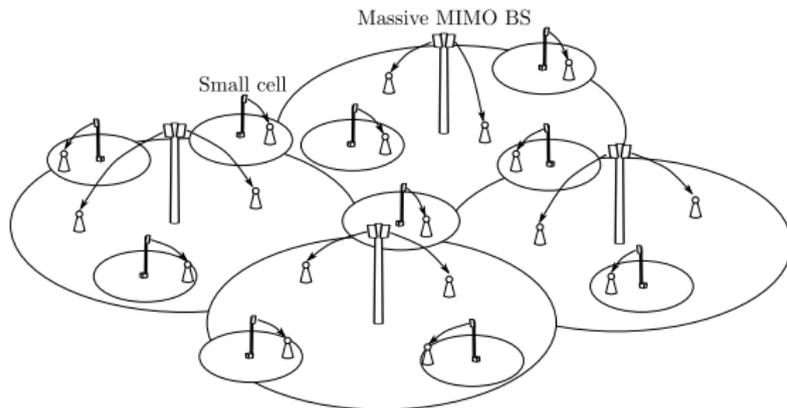
## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- **Translate this idea to HetNets**

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

## Translate this idea to HetNets

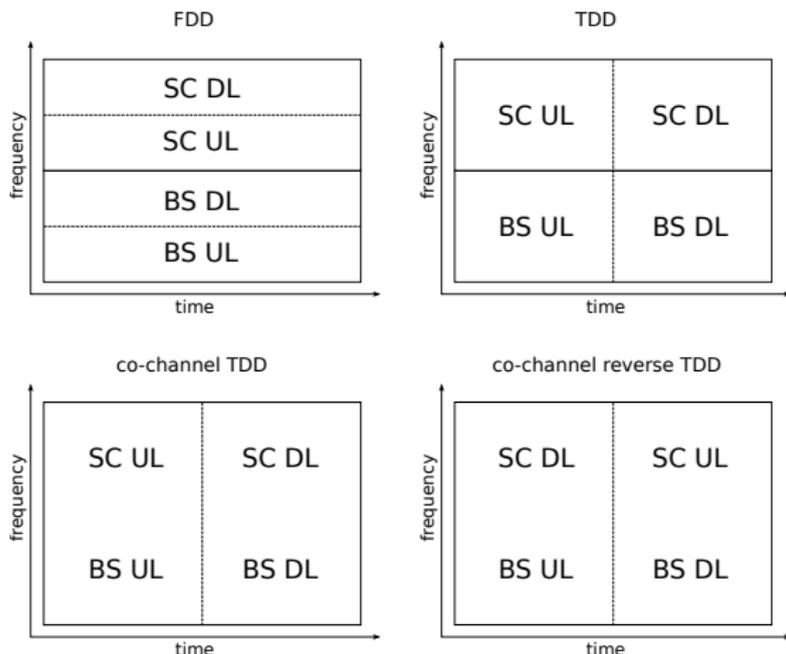


Every device estimates its received interference covariance matrix and precodes (partially) orthogonal to the dominating interference subspace.

### Advantages

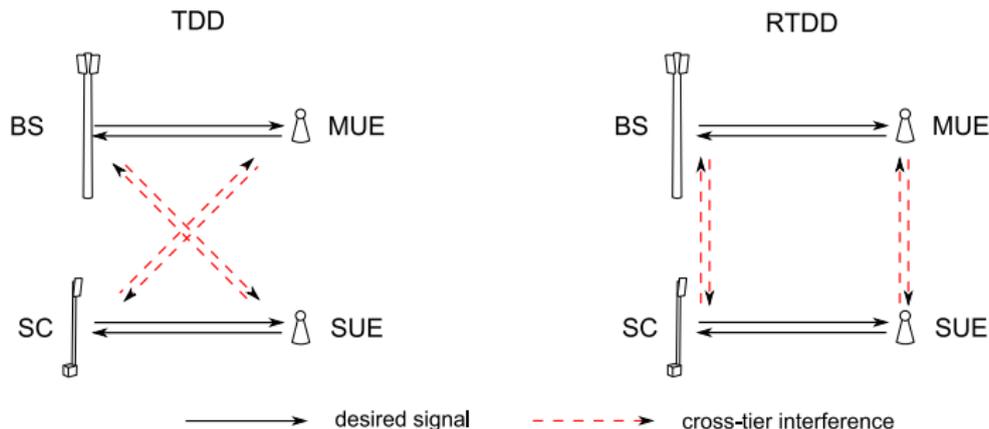
- Interference to the directions from which a device received most interference is reduced.
- No feedback or data exchange between the devices is needed.
- Every device relies on locally available information alone.
- The scheme is fully distributed and scalable.

## Baseline scenarios: FDD, TDD, RTDD, and co-channel deployment



- FDD: Channel reciprocity does not hold
- TDD: Only intra-tier interference can be reduced
- co-channel TDD: Inter and intra-tier interference can be reduced
- reverse (R) TDD: Order of UL/DL is reversed in one of the tiers

## TDD versus reverse TDD (RTDD)



- Order of UL/DL periods decides which devices interfere with each other:

- ▶ TDD

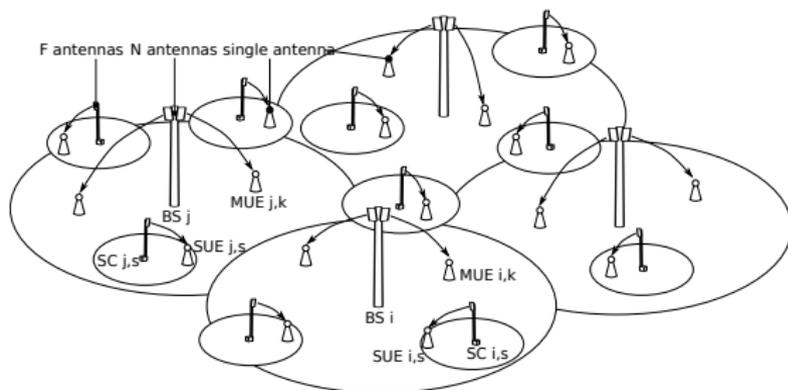
- ★ BS and small cell UEs (SUEs)
- ★ SCs and macro UEs (MUEs)
- ★ Intra-tier interference (SC-SUE, BS-MUE)

- ▶ RTDD

- ★ BS and SCs
- ★ MUEs and SUEs
- ★ Intra-tier interference (SC-SUE, BS-MUE)

- The BS-SC channels change very slowly (quasi-static). Thus, covariance estimation becomes easier for RTDD.

## Co-channel TDD: Uplink signaling



- $B$  BSs with  $N$  antennas
- $K$  single-ant. MUEs per BS
- $S$  SCs per BS with  $F$  antennas
- 1 single-ant. SUE per SC

Received signals at the  $i$ th BS and  $j$ th SC in cell  $i$ :

$$\mathbf{y}_i^{\text{BS}} = \sum_{b=1}^B \left( \sum_{k=1}^K \sqrt{P_{\text{MUE}}} \mathbf{h}_{ibk}^{\text{BS-MUE}} x_{bk}^{\text{MUE}} + \sum_{s=1}^S \sqrt{P_{\text{SUE}}} \mathbf{h}_{ibs}^{\text{BS-SUE}} x_{bs}^{\text{SUE}} \right) + \mathbf{n}_i^{\text{BS}}$$

$$y_{ij}^{\text{SC}} = \sum_{b=1}^B \left( \sum_{k=1}^K \sqrt{P_{\text{MUE}}} \mathbf{h}_{ijk}^{\text{SC-MUE}} x_{bk}^{\text{MUE}} + \sum_{s=1}^S \sqrt{P_{\text{SUE}}} \mathbf{h}_{ijbs}^{\text{SC-SUE}} x_{bs}^{\text{SUE}} \right) + n_{ij}^{\text{SC}}$$

## Co-channel TDD: Uplink rates

- BSs and SCs have perfect knowledge of the direct channels and the individual receive covariance matrices:

$$\mathbf{Q}_i^{\text{BS}} = \mathbb{E} \left[ \mathbf{y}_i^{\text{BS}} (\mathbf{y}_i^{\text{BS}})^H \right], \quad \mathbf{Q}_{ij}^{\text{SC}} = \mathbb{E} \left[ \mathbf{y}_{ij}^{\text{SC}} (\mathbf{y}_{ij}^{\text{SC}})^H \right]$$

- Achievable rates of the MUE  $k$  and SUE  $s$  in cell  $i$  with MMSE detection:

$$R_{ik}^{\text{UL,MUE}} = \frac{T_{\text{UL}}}{T} \log_2 \left( 1 + P_{\text{MUE}} (\mathbf{h}_{iik}^{\text{BS-MUE}})^H \left( \mathbf{Q}_i^{\text{BS}} - P_{\text{MUE}} \mathbf{h}_{iik}^{\text{BS-MUE}} (\mathbf{h}_{iik}^{\text{BS-MUE}})^H \right)^{-1} \mathbf{h}_{iik}^{\text{BS-MUE}} \right)$$

$$R_{is}^{\text{UL,SUE}} = \frac{T_{\text{UL}}}{T} \log_2 \left( 1 + P_{\text{SUE}} (\mathbf{h}_{isis}^{\text{SC-SUE}})^H \left( \mathbf{Q}_{is}^{\text{SC}} - P_{\text{SUE}} \mathbf{h}_{isis}^{\text{SC-SUE}} (\mathbf{h}_{isis}^{\text{SC-SUE}})^H \right)^{-1} \mathbf{h}_{isis}^{\text{SC-SUE}} \right)$$

where  $T$  is the channel coherence time and  $T_{\text{UL}}$  is the duration of the uplink cycle.

### On the CSI assumption

In practice, the direct channels must be estimated by uplink pilots. For a sufficiently long coherence time, the covariance matrices can be estimated by simple time averages. It is implicitly assumed that the transmit powers and noise variances are perfectly known.

## Co-channel TDD: Downlink signaling

Received signals at the  $j$ th MUE and  $j$ th SUE in cell  $i$ :

$$y_{ij}^{\text{MUE}} = \sum_{b=1}^B \left( \sum_{k=1}^K \sqrt{\frac{P_{\text{BS}}}{K}} (\mathbf{h}_{bij}^{\text{BS-MUE}})^H \mathbf{w}_{bk}^{\text{BS}} x_{bk}^{\text{BS}} + \sum_{s=1}^S \sqrt{P_{\text{SC}}} (h_{bsij}^{\text{SC-MUE}})^H \mathbf{w}_{bs}^{\text{SC}} x_{bs}^{\text{SC}} \right) + n_{ij}^{\text{MUE}}$$
$$y_{ij}^{\text{SUE}} = \sum_{b=1}^B \left( \sum_{k=1}^K \sqrt{\frac{P_{\text{BS}}}{K}} (\mathbf{h}_{bij}^{\text{BS-SUE}})^H \mathbf{w}_{bk}^{\text{BS}} x_{bk}^{\text{BS}} + \sum_{s=1}^S \sqrt{P_{\text{SC}}} (h_{bsij}^{\text{SC-SUE}})^H \mathbf{w}_{bs}^{\text{SC}} x_{bs}^{\text{SC}} \right) + n_{ij}^{\text{SUE}}$$

The BSs and SCs apply linear beamforming to serve their UEs:

$$\mathbf{w}_{bk}^{\text{BS}} = \kappa_{bk}^{\text{BS}} \left( (1 - \alpha) P_{\text{MUE}} \sum_j \mathbf{h}_{bbj}^{\text{BS-MUE}} (\mathbf{h}_{bbj}^{\text{BS-MUE}})^H + \alpha \mathbf{Q}_b^{\text{BS}} (1 - \alpha) N_o \mathbf{I}_N \right)^{-1} \mathbf{h}_{bbk}^{\text{BS-MUE}}$$
$$\mathbf{w}_{bs}^{\text{SC}} = \kappa_{bs}^{\text{SC}} \left( (1 - \beta) P_{\text{SUE}} \mathbf{h}_{bsbs}^{\text{SC-SUE}} (\mathbf{h}_{bsbs}^{\text{SC-SUE}})^H + \beta \mathbf{Q}_{bs}^{\text{SC}} + (1 - \beta) N_o \mathbf{I}_F \right)^{-1} \mathbf{h}_{bsbs}^{\text{SC-SUE}}$$

where  $\alpha, \beta$  are regularization parameters and  $\kappa_{bk}^{\text{BS}}, \kappa_{bs}^{\text{SC}}$  normalize the vector norms to one.

### About the regularization parameters

For  $\alpha, \beta = 0$ , the BSs and SCs transmit as if they were in an isolated cell, i.e., MMSE precoding (BSs) and maximum-ratio transmissions. By increasing  $\alpha, \beta$ , the precoding vectors become increasingly orthogonal to the interference subspace.

## Co-channel TDD: Downlink rates

Achievable downlink rates of the MUE  $k$  and SUE  $s$  in cell  $i$ :

$$R_{ik}^{\text{DL,MUE}} = \left(1 - \frac{T_{\text{UL}}}{T}\right) \log_2 \left(1 + \text{SINR}_{ik}^{\text{DL,MUE}}\right)$$

$$R_{is}^{\text{DL,SUE}} = \left(1 - \frac{T_{\text{UL}}}{T}\right) \log_2 \left(1 + \text{SINR}_{is}^{\text{DL,SUE}}\right)$$

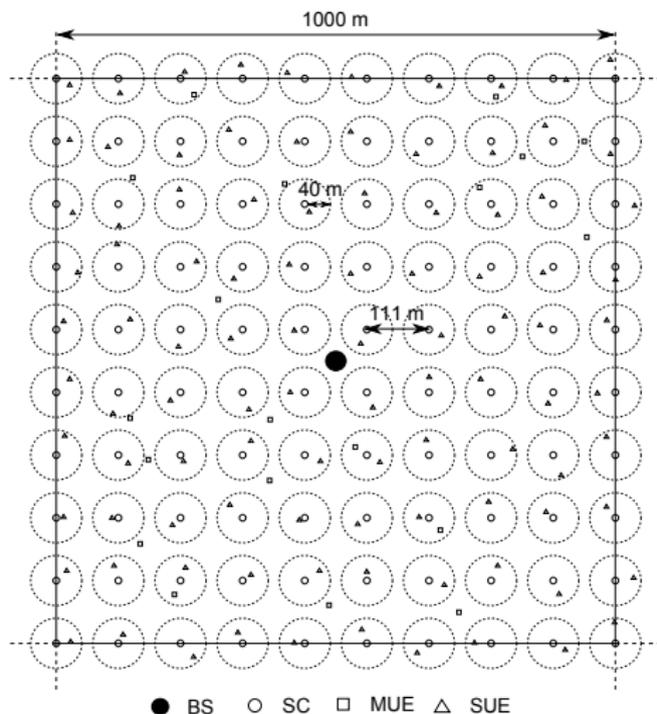
where

$$\text{SINR}_{ik}^{\text{DL,MUE}} = \frac{\frac{P_{\text{BS}}}{K} |(\mathbf{h}_{iik}^{\text{BS-MUE}})^H \mathbf{w}_{ik}^{\text{BS}}|^2}{N_o + \frac{P_{\text{BS}}}{K} \sum_{(b,j) \neq (i,k)} |(\mathbf{h}_{bik}^{\text{BS-MUE}})^H \mathbf{w}_{bj}^{\text{BS}}|^2 + P_{\text{SC}} \sum_{b,s} |(\mathbf{h}_{bsik}^{\text{SC-MUE}})^H \mathbf{w}_{bs}^{\text{SC}}|^2}$$
$$\text{SINR}_{is}^{\text{DL,SUE}} = \frac{P_{\text{SC}} |(\mathbf{h}_{isis}^{\text{SC-SUE}})^H \mathbf{w}_{is}^{\text{SC}}|^2}{N_o + \frac{P_{\text{BS}}}{K} \sum_{(b,j)} |(\mathbf{h}_{bis}^{\text{BS-SUE}})^H \mathbf{w}_{bj}^{\text{BS}}|^2 + P_{\text{SC}} \sum_{(b,s) \neq (i,s)} |(\mathbf{h}_{bjis}^{\text{SC-SUE}})^H \mathbf{w}_{bj}^{\text{SC}}|^2}$$

### Other duplexing schemes

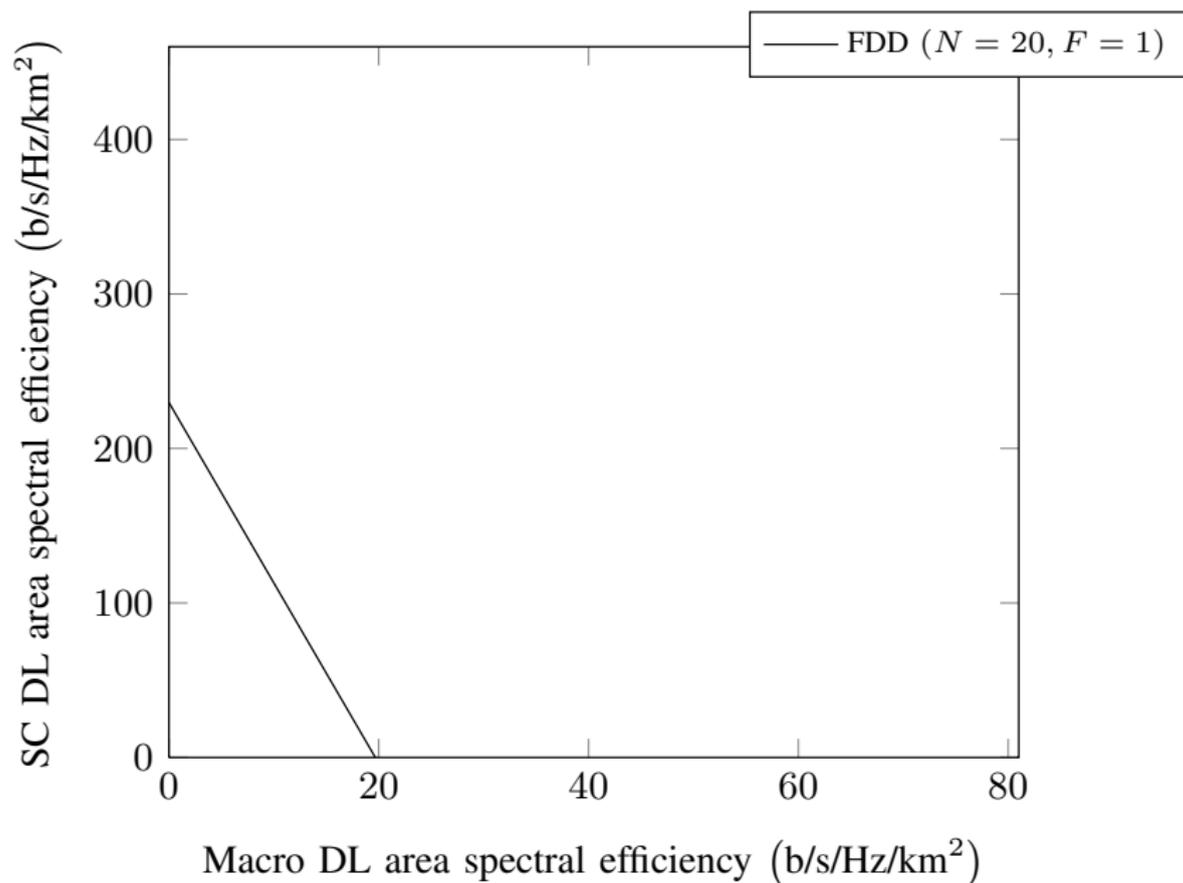
The other duplexing schemes work similarly, by adapting the covariance matrices and interference terms. In FDD, covariance knowledge cannot be exploited as channel reciprocity does not hold. Without co-channel deployment, there is only intra-tier interference.

## Numerical results

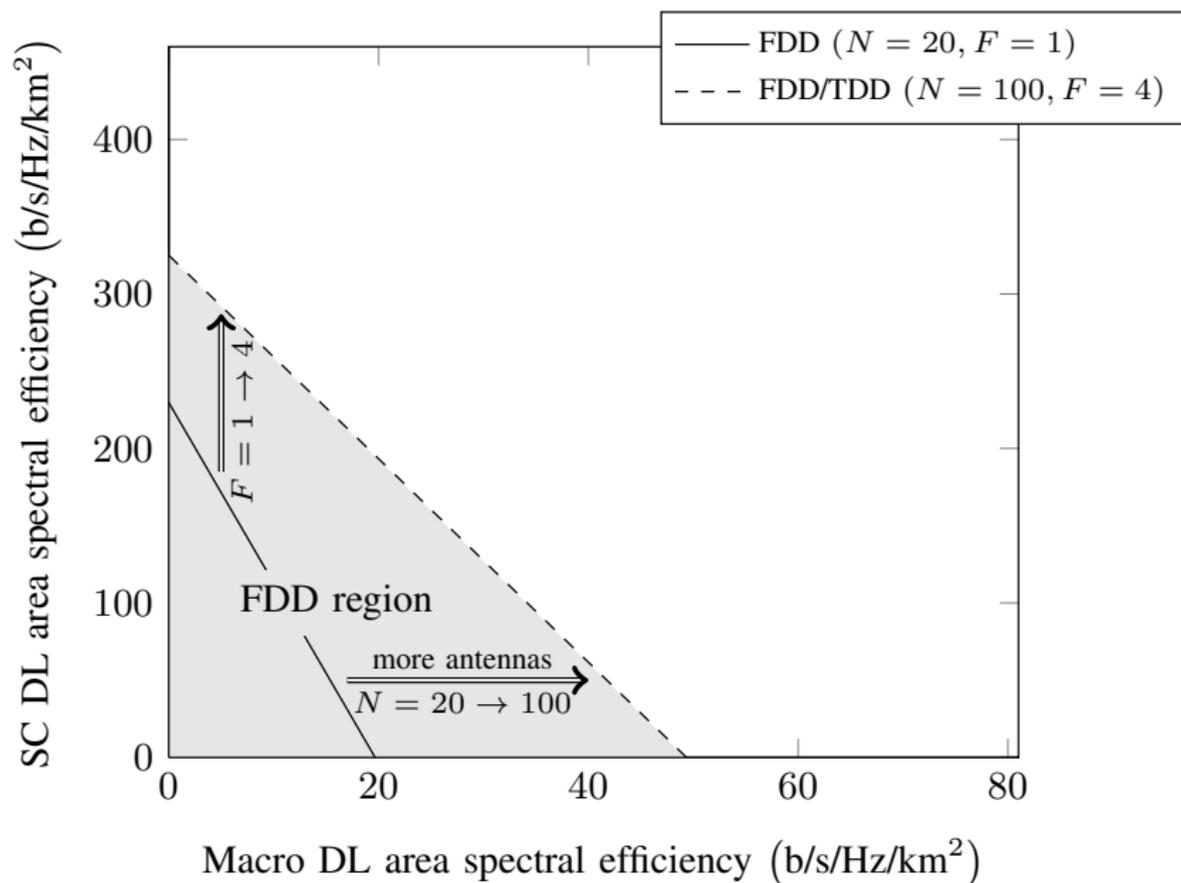


- $3 \times 3$  grid of BSs with wrap around
- $S = 81$  SCs per cells on a regular grid
- $K = 20$  MUEs randomly distributed in each cell
- 1 SUE per SC randomly distributed on a disc around each SC
- channel model with path loss, shadowing and fast fading, N/LOS links from [37]
- TX powers: 46 dBm (BS), 24 dBm (SC), 23 dBm (MUE/SUE)
- 20 MHz bandwidth, 2 GHz center frequency
- no user scheduling, power control
- averages over channel realizations and UE locations
- $T_{UL} = 0.5T$

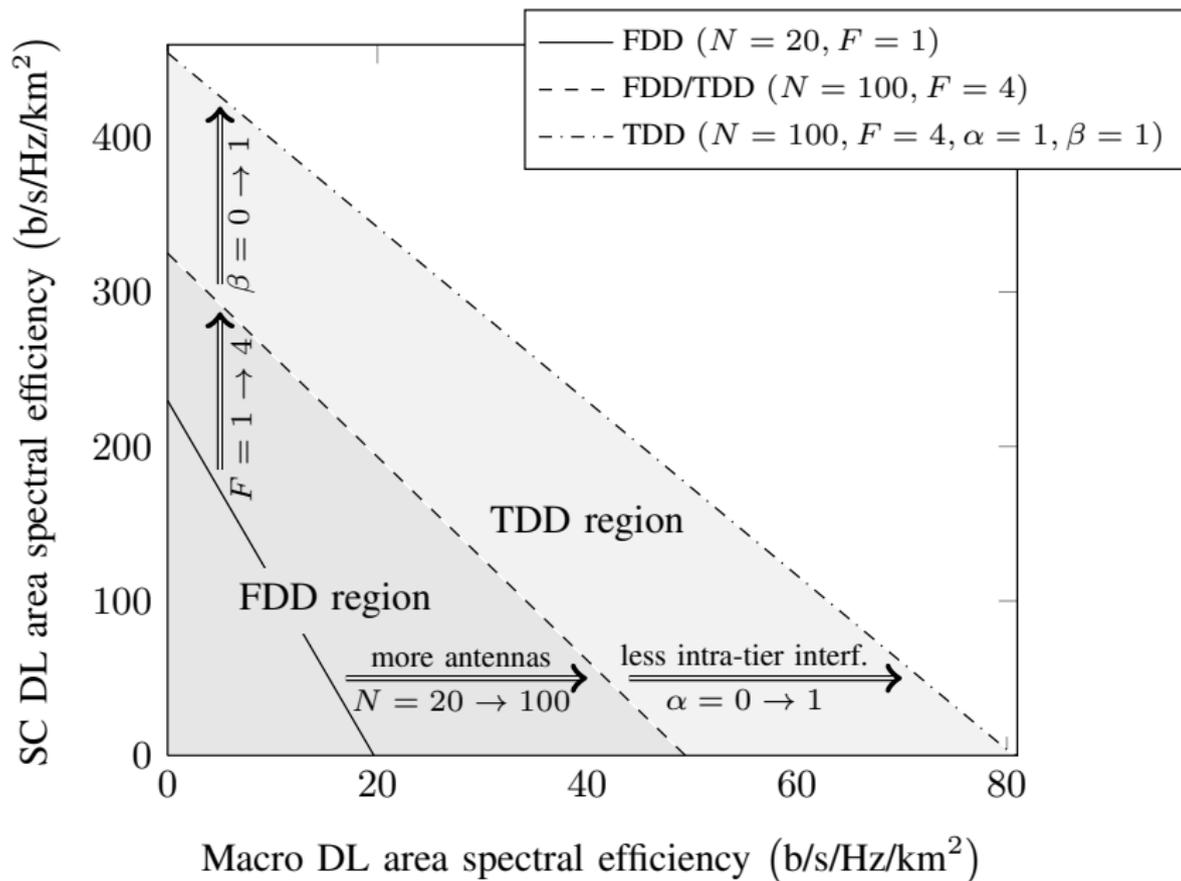
## Downlink rate regions



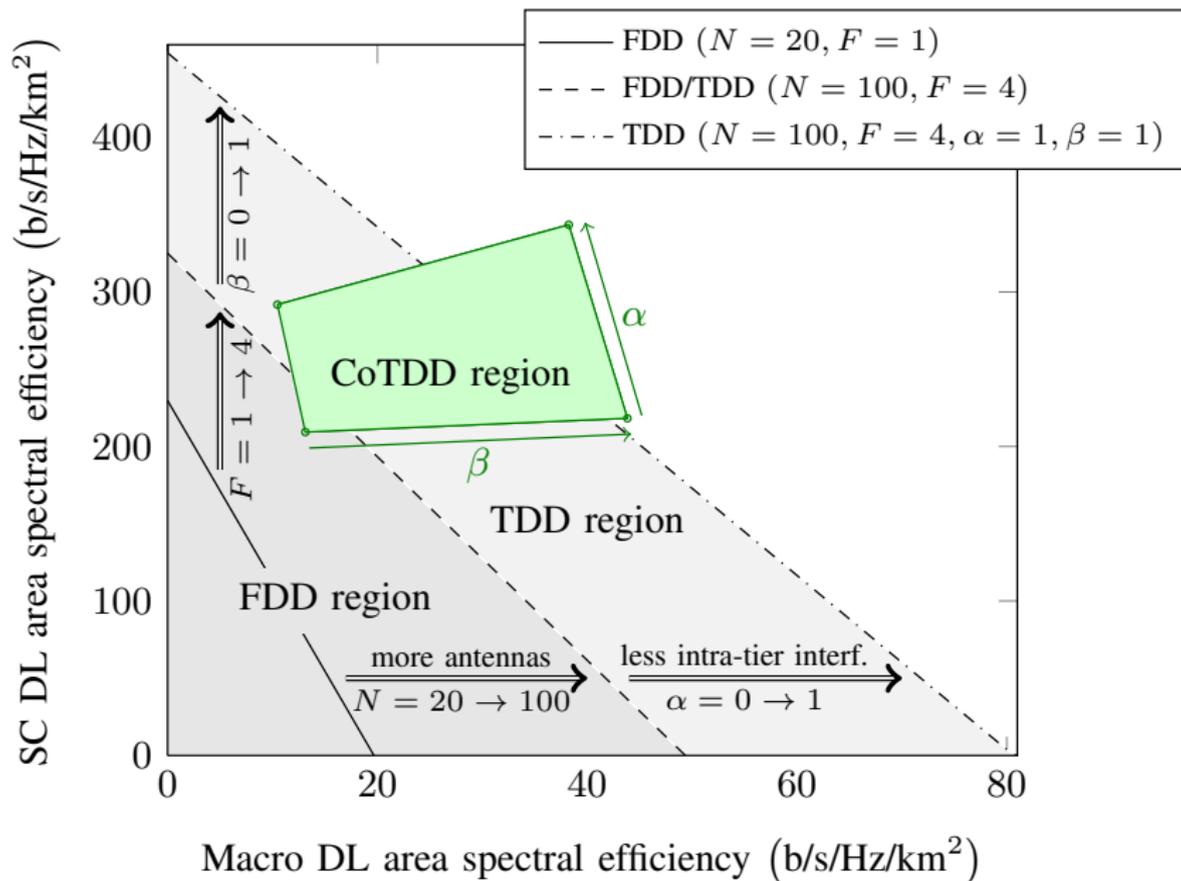
## Downlink rate regions



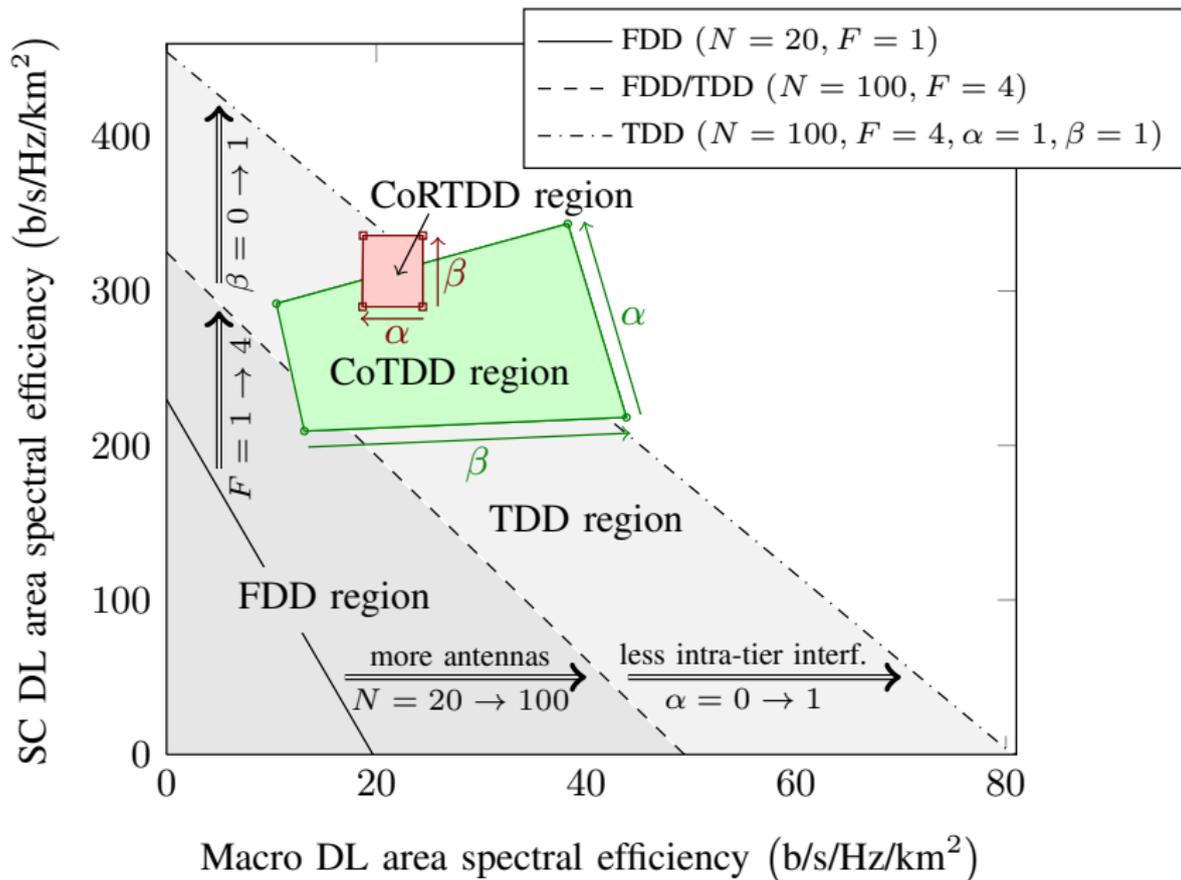
## Downlink rate regions



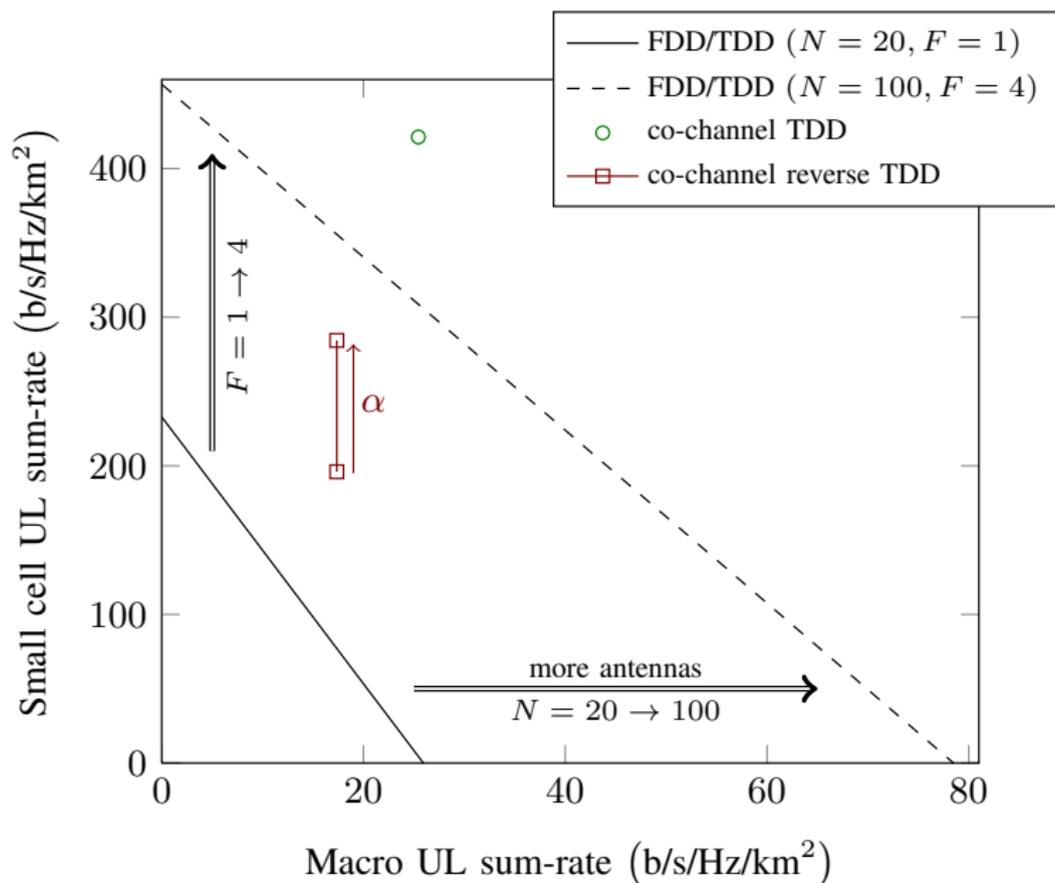
## Downlink rate regions



## Downlink rate regions



## Uplink rate regions



## Observations

- Increasing the number of antennas at each device leads to tremendous performance improvements for all duplexing schemes ( $N = 20 \rightarrow 100$ ,  $F = 1 \rightarrow 4$ , FDD):

+200 % BS UL, +150 % BS DL, +100 % SC UL, +50 % SC DL

- TDD channel reciprocity allows for intra-tier interference reduction ( $\alpha, \beta : 0 \rightarrow 1$ ):

+50 % BS DL, +30 % SC DL

- Even a few “excess” antennas at the SCs leads to significant gains.
- With RTDD, the SCs cannot reduce interference towards the BS (since  $F \ll N$ ).
- With the proposed precoding scheme, a co-channel deployment of BSs and SCs leads to the highest area spectral efficiency ( $\alpha = \beta = 1$ , 20 MHz bandwidth):

	UL	DL
Area throughput	7.63 Gb/s/km <sup>2</sup>	8.93 Gb/s/km <sup>2</sup>
Rate per MUE	38.2 Mb/s	25.4 Mb/s
Rate per SUE	84.8 Mb/s	104 Mb/s

- As the scheme is fully distributed and requires not data exchange between the devices, the rates can be simply increased by adding more antennas to the BSs/SCs or increasing the SC-density.

- Channel reciprocity requires:
  - ▶ Hardware calibration [38, 39, 40, 12]
  - ▶ Scheduling of UEs on the same resource blocks in subsequent UL/DL cycles.
- The network-wide TDD protocol requires tight synchronization of all devices:
  - ▶ GPS (outdoor)
  - ▶ NTP/PTP (indoor)
  - ▶ BS reference signals
- Channel estimation will suffer from pilot contamination.
- Covariance matrix estimation becomes difficult for large  $N, F$  [41, 42].
- We have considered a worst case model with fixed cell association, no power control or scheduling. Location-dependent user scheduling and interference-temperature power control could further enhance the performance [43].
- Band switching duplexing could be used to reduce duplexing delays [44].

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

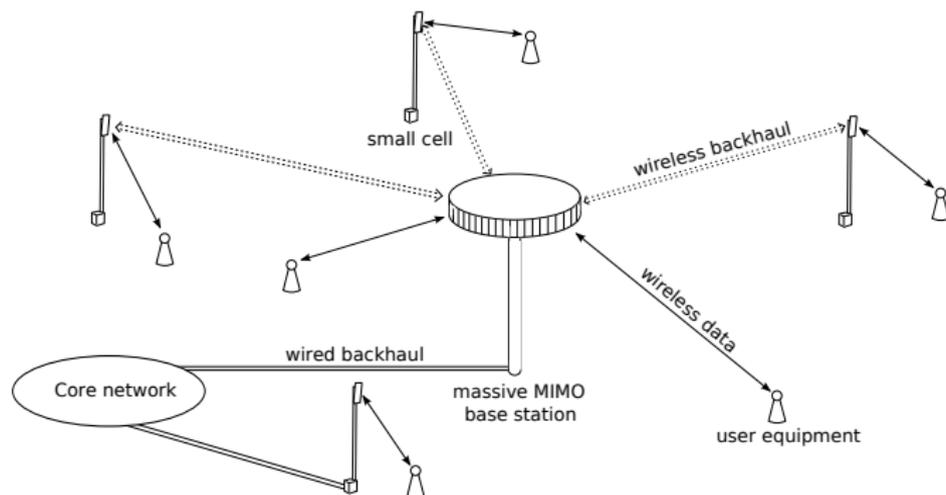
## 3 Massive MIMO for wireless backhaul

- **Motivation and advantages**
- Power and “antenna” minimization problem

# Massive MIMO for wireless backhaul

- The unrestrained SC-deployment “where needed” rather than “where possible” requires a high-capacity and easily accessible backhaul network.
- Already for most WiFi deployments, the backhaul capacity (10–100 Mbit/s) and not the air interface (54–600 Mbit/s) is the bottleneck.

## Massive MIMO for wireless backhaul



- The unrestrained SC-deployment “where needed” rather than “where possible” requires a high-capacity and easily accessible backhaul network.
- Already for most WiFi deployments, the backhaul capacity (10–100 Mbit/s) and not the air interface (54–600 Mbit/s) is the bottleneck.

Why not provide wireless backhaul with massive MIMO [31]?

## Massive MIMO wireless backhaul: Advantages

- No standardization or backward-compatibility required
- BS-SC channels change very slowly over time:
  - ▶ Complex transmission/detection schemes (e.g., CoMP) can be easily implemented
  - ▶ FDD might be possible due to reduced CSI overhead
- Provide backhaul where needed:
  - ▶ Adapt backhaul capacity to the load
  - ▶ Statistical multiplexing opportunity to avoid over-provisioning of backhaul
- SCs require only a power connection to be operational
- Line-of-sight not necessary if operated at low frequencies

### Remark

Interestingly, the backhaul load is highest in lightly loaded cells where a single UE with a very good channel achieves the maximum possible data rate on the wireless link [45].

## Massive MIMO wireless backhaul: Is it feasible?

How many antennas are needed to satisfy the desired backhaul rates with a given transmit power budget?

### Assumptions:

- Every BS knows the channels to all SCs.
- The BSs can exchange some control information.
- Full user data sharing between the BSs is not possible.
- Single-antenna SCs, BSs with  $N$  antennas
- TDD with channel reciprocity

Check if the power minimization problem with target SINR constraints for the multi-cell multi-antenna wireless system is feasible [46].

# Outline

## 1 Massive MIMO

- Benefits
- Favorable propagation conditions
- Channel estimation and pilot contamination
- Hardware impairments
- Research topics

## 2 Massive MIMO and HetNets

- Small cells, a two-tier network architecture, and the role of TDD
- An idea from cognitive radio
- Translate this idea to HetNets

## 3 Massive MIMO for wireless backhaul

- Motivation and advantages
- Power and “antenna” minimization problem

# The multi-cell multi-antenna power minimization and beamforming problem

## P 1: Power minimization problem [46]

$$\begin{aligned} \min \quad & \sum_{bs} \mathbf{w}_{bs}^H \mathbf{w}_{bs} \\ \text{s.t.} \quad & \text{SINR}_{bs} \geq \gamma_{bs} \quad 1 \leq b \leq B, 1 \leq s \leq S \end{aligned}$$

- $\mathbf{w}_{bs}$  is the beamforming vector of BS  $b$  towards SC  $s$  in its cell
- $\gamma_{bs}$  is the target SINR of the backhaul link to SC  $s$  in cell  $b$
- $\text{SINR}_{bs}$  is the SINR of the backhaul link to SC  $s$  in cell  $b$ :

$$\text{SINR}_{bs} = \frac{|\mathbf{w}_{bs}^H \mathbf{h}_{bbs}|^2}{\sum_{l \neq s} |\mathbf{w}_{bl}^H \mathbf{h}_{bbs}|^2 + \sum_{m \neq b, l} |\mathbf{w}_{ml}^H \mathbf{h}_{mbs}|^2 + N_o}$$

where  $\mathbf{h}_{mbs} \in \mathbb{C}^{N \times 1}$  is the channel from BS  $m$  to SC  $s$  in cell  $b$ .

### Remark

The problem is always feasible if  $N \geq B \times S$  since zero-forcing can be applied.

## Solution to the power minimization problem

- 1 Find through a classic fixed-point algorithm the values  $\lambda_{bs}$  satisfying

$$\lambda_{bs} = \frac{1}{\left(1 + \frac{1}{\gamma_{bs}}\right) \mathbf{h}_{bbs}^H \boldsymbol{\Sigma}_b^{-1} \mathbf{h}_{bbs}} \quad \forall b, s$$

where  $\boldsymbol{\Sigma}_b = \mathbf{I}_N + \sum_{i,s} \lambda_{is} \mathbf{h}_{b,i,s} \mathbf{h}_{b,i,s}^H$ .

- 2 Compute the optimal “uplink” receive filters.

$$\hat{\mathbf{w}}_{bs} = \frac{1}{N_o} \boldsymbol{\Sigma}_b^{-1} \mathbf{h}_{bbs} \quad \forall b, s$$

- 3 Find the downlink beamformers by scaling of  $\hat{\mathbf{w}}_{bs}$ :

$$\mathbf{w}_{bs} = \sqrt{\delta_{bs}} \hat{\mathbf{w}}_{bs}$$

where  $\boldsymbol{\delta} = [\delta_{11}, \dots, \delta_{1S}, \delta_{21}, \dots, \delta_{BS}]^T$  is given by  $\boldsymbol{\delta} = \mathbf{F}^{-1} \mathbf{1}_{BS} N_o$  with

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}^{11} & \dots & \mathbf{F}^{1B} \\ \vdots & \ddots & \vdots \\ \mathbf{F}^{B1} & \dots & \mathbf{F}^{BB} \end{pmatrix}, \quad [\mathbf{F}^{im}]_{jn} = \begin{cases} \frac{1}{\gamma_{ij}} \left| \hat{\mathbf{w}}_{ij}^H \mathbf{h}_{ij} \right|^2 & , m = i, n = j \\ - \left| \hat{\mathbf{w}}_{in}^H \mathbf{h}_{ij} \right|^2 & , m = i, n \neq j \\ - \left| \hat{\mathbf{w}}_{mn}^H \mathbf{h}_{mij} \right|^2 & , m \neq i \end{cases}$$

### Remark

The computational complexity of this algorithm grows quickly with  $N$ ,  $B$  and the convergence speed of step 1 increases with the target SINRs  $\gamma_{bs}$ .

## “Antenna” minimization problem

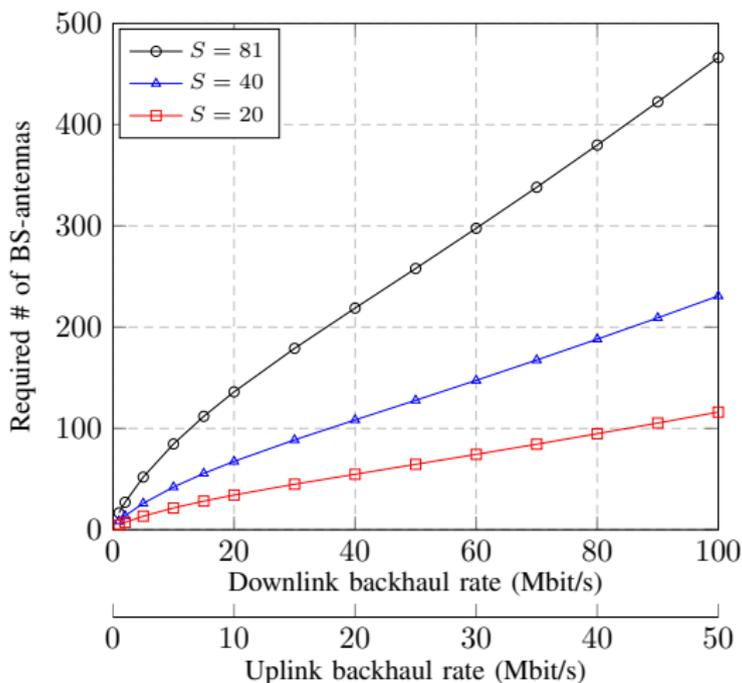
### P 2: Minimization of $N$

$$\begin{aligned} & \min N \\ & \text{s.t. } \frac{1}{B} \sum_{b,s} \mathbf{w}_{bs}^H \mathbf{w}_{bs} \leq P_{BS} \end{aligned}$$

where  $\mathbf{w}_{bs}$  are given by the solution of the power minimization problem P 1.

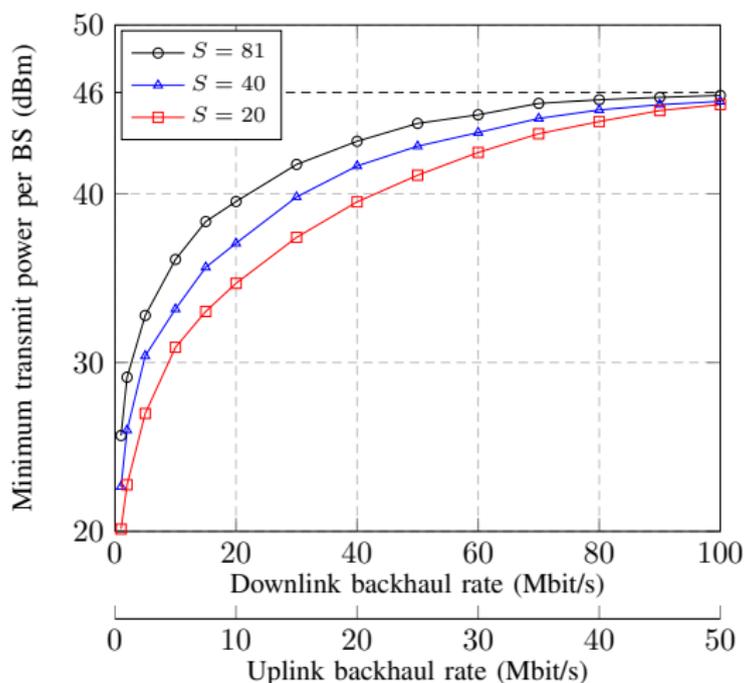
- Solution of P 1 becomes prohibitive for large  $N, S$ : Large system approximation [47]
  - ▶ computed based on the second-order statistics of the channels
  - ▶ complexity is independent of  $N$
  - ▶ solution is asymptotically optimal (for  $N, S \rightarrow \infty$ )
- By UL/DL duality, the same target DL SINRs  $\gamma_{bs}$  can be achieved in the UL with the receive filters  $\hat{\mathbf{w}}_{bs}$  at the BSs and the SC transmit powers  $\lambda_{bs}$ . However, the individual power budgets of the SCs are not respected in this case.
- Numerical results based on the  $3 \times 3$  macro-cell setup as described earlier. The downlink TDD cycle is twice as long as the uplink cycle.

## Massive MIMO backhaul: Numerical results



Average minimum number of required BS-antennas  $N$  to serve  $S \in \{20, 40, 81\}$  randomly chosen SCs with the same target backhaul rate and with 46 dBm maximum average transmit power per BS.

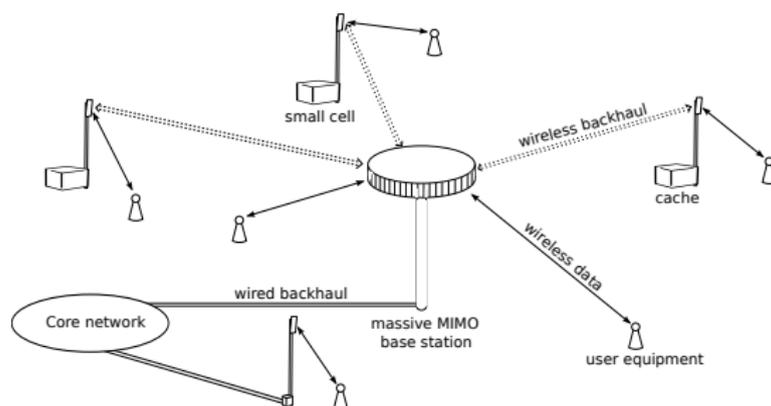
## Massive MIMO backhaul: Numerical results



Minimum average required transmit power per BS to serve  $S \in \{20, 40, 81\}$  randomly chosen SCs with the same target backhaul rate and the smallest possible number of BS-antennas  $N$ .

- Is there an advantage to serve UEs via SCs rather than directly by the BSs?
  - ▶ Long coherence time of the BS-SC channels enables complex precoding schemes.
  - ▶ Uplink transmit power can be drastically reduced.
  - ▶ SCs with wireless backhaul are essentially relays which can cover hidden areas.
- Low frequencies are suited for (NLOS) backhaul (small propagation losses) while higher frequencies are suited for SCs to reduce interference.
- Given enough BS-antennas, even in-band backhaul could be provided.

## Discussion



- Is there an advantage to serve UEs via SCs rather than directly by the BSs?
  - ▶ Long coherence time of the BS-SC channels enables complex precoding schemes.
  - ▶ Uplink transmit power can be drastically reduced.
  - ▶ SCs with wireless backhaul are essentially relays which can cover hidden areas.
- Low frequencies are suited for (NLOS) backhaul (small propagation losses) while higher frequencies are suited for SCs to reduce interference.
- Given enough BS-antennas, even in-band backhaul could be provided.
- SCs could preload popular files on integrated hard discs [48] based on learning and prediction algorithms to reduce the backhaul bandwidth.

## Lessons learned III

- Massive MIMO and SCs have distinct advantages which complement each other:
  - ▶ Massive MIMO for area coverage and mobility support
  - ▶ SCs for capacity and indoor coverage
- TDD and the resulting channel reciprocity allows every device to fully exploit its available degrees of freedom for intra-/inter-tier interference mitigation.
- A co-channel deployment of massive MIMO BSs and SCs can achieve a very attractive rate region.
- Massive MIMO BSs can provide wireless backhaul to a large number of SCs: With  $N = 466$  antennas per BS, it is possible to provide 81 SCs with 100 Mb/s (DL)/ 50 Mb/s (UL) backhaul.

Thank you!

# References I

- [1] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–46, Jan. 2013.
- [4] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.
- [5] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Series in Statistics, New York, NY, USA, 2009.
- [6] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.
- [7] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [8] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing," 2012, submitted. [Online]. Available: <http://arxiv.org/abs/1209.1402>
- [9] S. J. Fortune, D. M. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela, and M. H. Wright, "WISE design of indoor wireless systems: practical computation and optimization," *IEEE Computational Science Engineering*, vol. 2, no. 1, pp. 58–68, Spring 1995.
- [10] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Paris, France, Aug. 2012, pp. 811–815.
- [11] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, US, Nov. 2006, pp. 359–363.

## References II

- [12] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, Istanbul, Turkey, Aug. 2012, pp. 53–64.
- [13] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [14] H. Yin, D. Gesbert, M. Filippou, and L. Yingzhuang, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.
- [15] R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," *Proc. 17th Intl ITG Workshop on Smart Antennas (WSA)*, Mar. 2013.
- [16] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.
- [17] A. Ashikhmin and T. L. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 1137–1141.
- [18] F. Fernandes, A. Ashikhmin, and T. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.
- [19] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on uplink performance of large MU-MIMO systems," in *Proc. 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Urbana-Champaign, IL, USA, Oct. 2012, pp. 1190–1197.
- [20] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Hardware impairments in large-scale MISO systems: Energy efficiency, estimation, and capacity limits," in *Proc. International Conference on Signal Processing (DSP): Special Session on Signal Processing and Optimization for Green Energy and Green Communications*, Santorini, Greece, Jul. 2013. [Online]. Available: <http://arxiv.org/abs/1305.4651>
- [21] —, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," 2013, in preparation.

## References III

- [22] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.
- [23] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.
- [24] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep. 216, 1952.
- [25] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [26] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 Ghz," in *Proc. 6th European Conference on Antennas and Propagation (EuCAP)*, Prague, Czech Republic, Mar. 2012, pp. 433–437.
- [27] "Window of opportunity," [http://www.ericsson.com/thinkingahead/networked\\_society/window\\_of\\_opportunity](http://www.ericsson.com/thinkingahead/networked_society/window_of_opportunity), accessed: 2013-05-17.
- [28] E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *International Conference on Telecommunications (ICT)*, Casablanca, Morocco, May 2013.
- [29] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, Sep. 2013.
- [30] "Massive MIMO and small cells: How to densify heterogeneous networks," in *IEEE International Conference on Communications (ICC)*, Budapest, Hungary, Jun. 2013.
- [31] T. L. Marzetta and H. Yang, "Dedicated LSAS for metro-cell wireless backhaul - Part I: Downlink," Bell Laboratories, Alcatel-Lucent, Tech. Rep., Dec. 2012.
- [32] T. Bai and R. W. Heath, "Asymptotic coverage probability and rate in massive MIMO networks," 2013, submitted. [Online]. Available: <http://arxiv.org/abs/1305.2233>

## References IV

- [33] J. Hoydis, A. Müller, R. Couillet, and M. Debbah, "Analysis of multicell cooperation with random user locations via deterministic equivalents," in *IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt): Workshop on Spatial Stochastic Models for Wireless Networks (SPASWIN)*, Paderborn, Germany, May 2012, pp. 374–379.
- [34] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 4, no. 61, pp. 1436–1449, Apr. 2013.
- [35] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO hetnets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, 2013, submitted. [Online]. Available: <http://arxiv.org/abs/1301.5034>
- [36] R. Zhang, F. Gao, and Y.-C. Liang, "Cognitive beamforming made practical: Effective interference channel and learning-throughput tradeoff," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 706–718, Feb. 2010.
- [37] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network, Evolved Universal Terrestrial Radio Access (E-UTRA), Further Enhancements to LTE Time Division Duplex (TDD) for Downlink-Uplink (DL-UL) Interference Management and Traffic Adaptation (Release 11), 3GPP TR 36.828, v11.0.0," Jun. 2012. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/36828.htm>
- [38] J.-C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," in *IEEE Vehicular Technology Conference (VTC Fall)*, Los Angeles, CA, USA, Sep. 2004, pp. 4265–4269.
- [39] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *IEEE International Symposium on Signal Processing and Its Applications (ISSPA)*, Sydney, Australia, Aug. 2005, pp. 403–406.
- [40] J. Shi, Q. Luo, and M. You, "An efficient method for enhancing TDD over the air reciprocity calibration," in *IEEE Wireless Communication and Networking Conference (WCNC)*, Cancun, Mexico, Mar. 2011, pp. 339–344.
- [41] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, vol. 30, pp. 24–39, Jan. 2013.
- [42] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, p. 365411, Feb. 2004.

## References V

- [43] A. Adhikary and G. Caire, "On the coexistence of macrocell spatial multiplexing and cognitive femtocells," in *IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Jun. 2012, pp. 6830–6834.
- [44] P. Bosch and S. J. Mullender, "Band switching for coherent beam forming in full-duplex wireless communication," U.S. Patent 2005/0243 748 A1, 2005. [Online]. Available: <http://www.freepatentsonline.com/y2005/0243748.html>
- [45] Small Cell Forum, "Release One, Document 049.01.01, Backhaul technologies for small cells," Feb. 2013. [Online]. Available: [http://scf.io/en/documents/049\\_Backhaul\\_technologies\\_for\\_small\\_cells.php](http://scf.io/en/documents/049_Backhaul_technologies_for_small_cells.php)
- [46] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.
- [47] S. Lakshminarayana, J. Hoydis, M. Debbah, and M. Assaad, "Asymptotic analysis of distributed multi-cell beamforming," in *IEEE International Symposium in Personal Indoor and Mobile Radio Communications (PIMRC)*, Istanbul, Turkey, Sep. 2010, pp. 2105–2110.
- [48] N. Golrezaei, K. Shanmugam, A. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *IEEE International Conference on Computer Communications*, Orlando, Florida, USA, Mar. 2012, pp. 1107–1115.