The Path Towards 5G:
Massive MIMO

Dr. Emil Björnson

Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, Linköping, Sweden

LINKÖPING UNIVERSITY

Outline

- Introduction
- How to Achieve Higher Spectral Efficiency
- Basic Properties of Massive MIMO
- Massive MIMO Transmission Protocol
- 4 Myths and Misconceptions
- Research Trends and Open Questions
- Summary

INTRODUCTION

Incredible Success of Wireless Communications

Martin Cooper's law

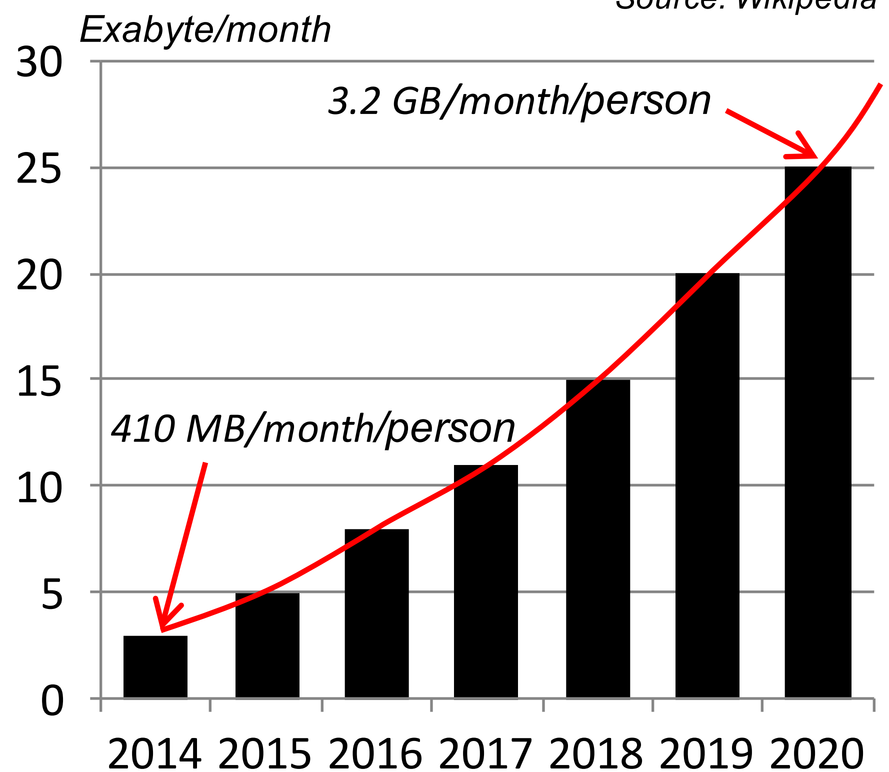
The number of voice/data connections has doubled every 2.5 years (+32% per year) since the beginning of wireless

Last 45 years: 1 Million Increase in Wireless Traffic

Two-way radio, FM/AM radio, satellites, cellular, WiFi, etc.

Source: Wikipedia

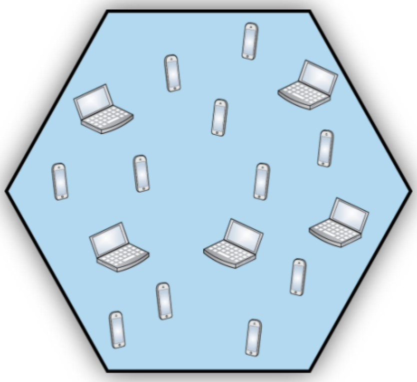
- Future Network Traffic Growth
  - 38% annual data traffic growth
  - Slightly faster than in the past!
  - Exponential increase
  - Extrapolation: 5x in 5 years
    25x in 10 years
    125x in 15 years



Source: Ericsson (November 2014)

Evolving Networks for Higher Traffic

- Increase Network Throughput [bit/s/km^2]
  - Consider a given area



- Simple Formula for Network Throughput:

$$\underbrace{\text{Throughput}}_{\text{bit/s/km}^2} = \underbrace{\text{Available spectrum}}_{\text{Hz}} \cdot \underbrace{\text{Cell density}}_{\text{Cell/km}^2} \cdot \underbrace{\text{Spectral efficiency}}_{\text{bit/s/Hz/Cell}}$$

- Ways to Achieve 1000x Improvement:

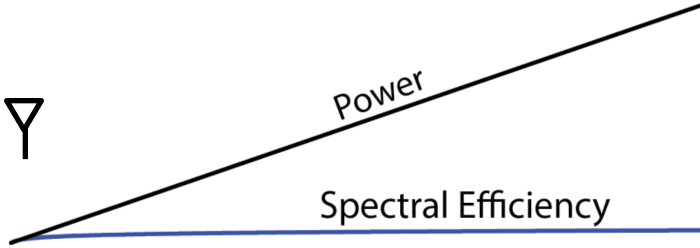
	More spectrum	Higher cell density	Higher spectral efficiency
Nokia (2011)	10x	10x	10x
SK Telecom (2012)	3x	56x	6x

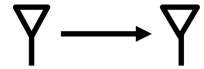
New regulations, cognitive radio, mmWave bands	*Ultra dense, heterogeneous deployments*	***Massive MIMO (Topic of this talk) ?x***

How to achieve

HIGHER SPECTRAL EFFICIENCY

Higher Spectral Efficiency

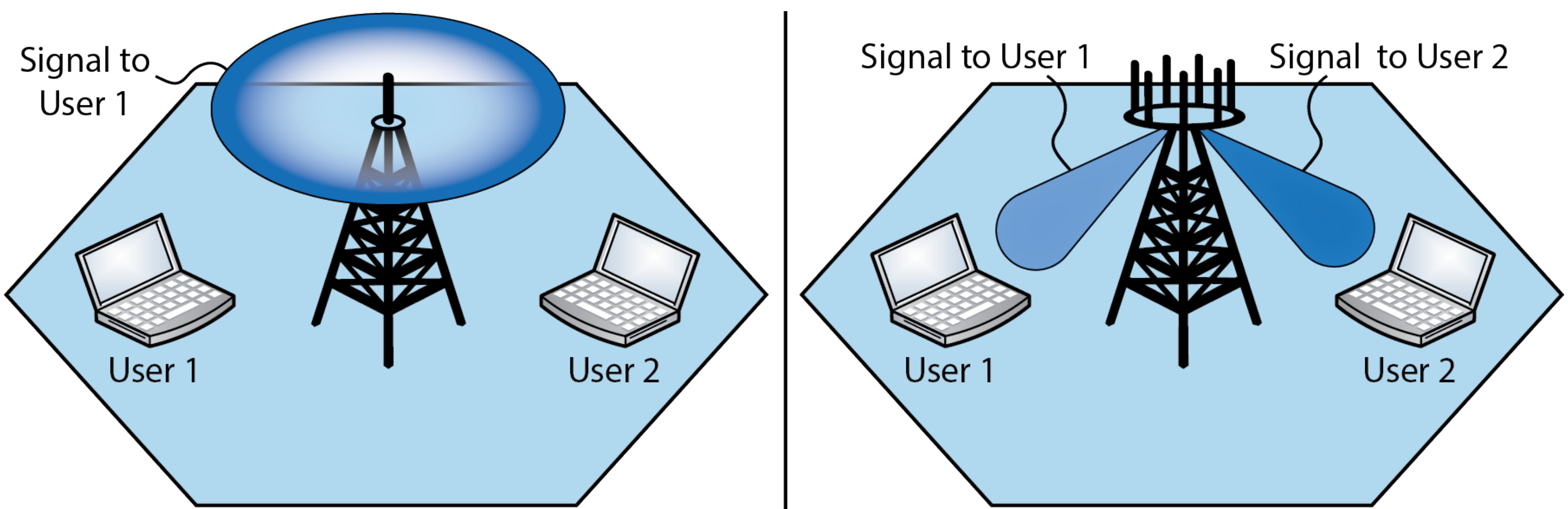


- Point-to-Point Spectral Efficiency: 
  - Governed by Shannon's capacity limit:

$$\log_2\left(1 + \frac{\text{Received Signal Power}}{\text{Interference Power} + \text{ Noise Power}}\right) \text{ [bit/s/Hz/user]}$$

  - Cannot do much: 4 bit/s/Hz → 8 bit/s/Hz requires 17 times more power!

- Many Parallel Transmissions: *Spatially focused to each desired user*

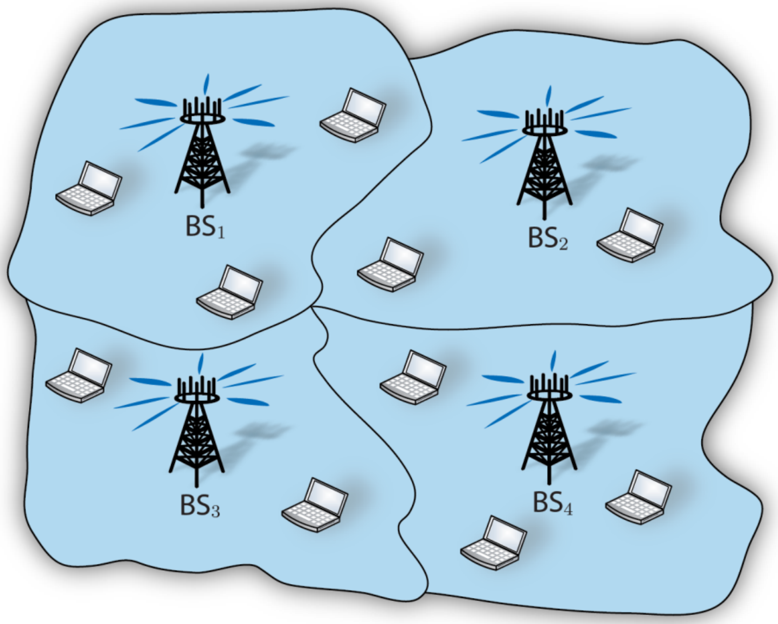


Single-Antenna Transmission

Multi-Antenna Transmission

Multi-User MIMO (Multiple-input Multiple-output)

- Multi-Cell Multi-User MIMO
  - Base stations (BSs) with M antennas
  - Parallel uplink/downlink for K users
  - Channel coherence block: S symbols



- Theory: Hardware is Limiting
  - Spectral efficiency roughly prop. to

$$\min\left(M, K, \frac{S}{2}\right)$$

  - 2x improvement = 2x antennas and users (since $S \in [100{,}10000]$)

- Practice: Interference is Limiting
  - Multi-user MIMO in LTE-A: Up to 8 antennas
  - Small gains since: Hard to learn users' channels
    Hard to coordinate BSs

End of the MIMO road?

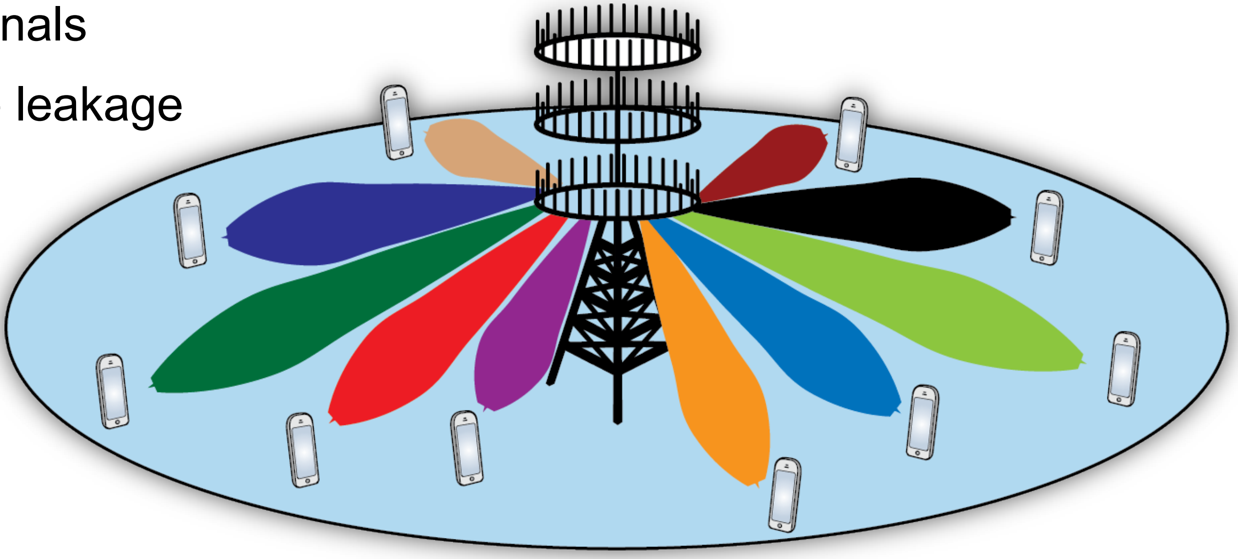
No reason to add more antennas/users?

Taking Multi-User MIMO to the Next Level

- Network Architecture: Massive MIMO
  - BS with many antennas; e.g., $M \approx 200$ antennas, $K \approx 40$ users
  - Key: Excessive number of antennas, $M \gg K$
  - Very directive signals
  - Little interference leakage

Spectral efficiency prop. to number of users!

$$\min\left(M, K, \frac{S}{2}\right) \approx K$$



- 2013 IEEE Marconi Prize Paper Award
  Thomas Marzetta, "*Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas*," IEEE Trans. Wireless Communications, 2010.
  - Analysis based on $M \to \infty$, but concept applicable at any M
  - Intended for conventional cellular frequencies (< 6 GHz)

What is the Key Difference from Today?

- Number of Antennas? **No, we already have many antennas!**
  - 3G/UMTS: 3 sectors x 20 element-arrays = 60 antennas
  - 4G/LTE-A: 4-MIMO x 60 = 240 antennas

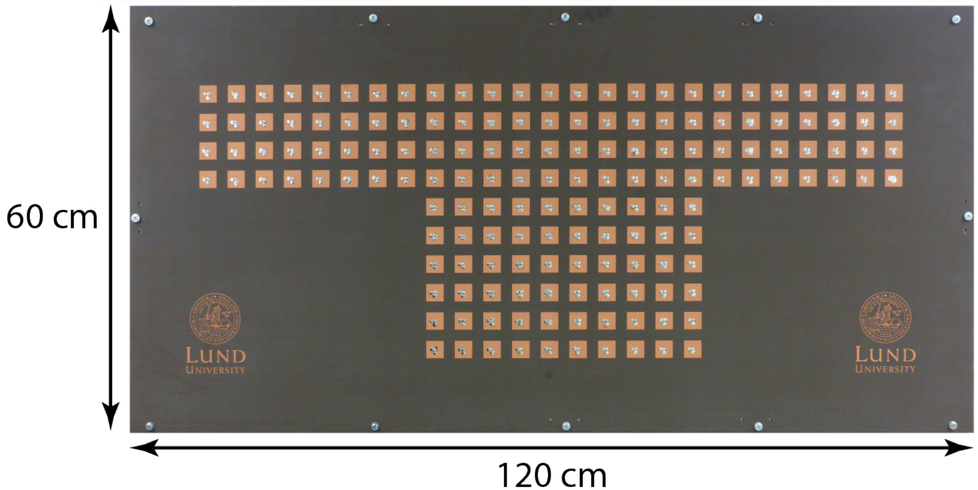
Typical vertical array:
10 antennas x 2 polarizations
Only 1-2 antenna ports

Massive MIMO Characteristics

Many small dipoles with transceiver chains

Spatial multiplexing of tens of users

Massive in numbers – not massive in size

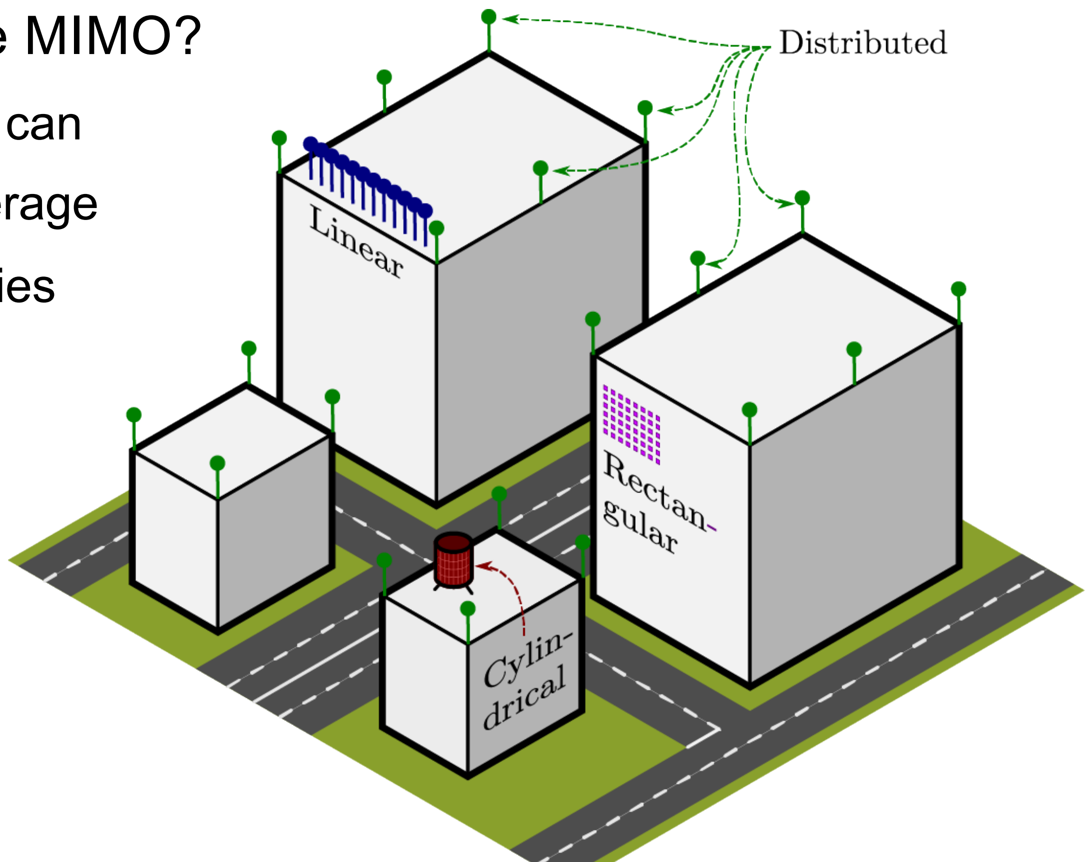


160 antenna elements, *LuMaMi testbed, Lund University*

3 sectors, 4 vertical arrays per sector
Image source: gigaom.com

Massive MIMO Deployment

- When to Deploy Massive MIMO?
  - The future will tell, but it can
  1. Improve wide-area coverage
  2. Handle high user densities



- Co-located Deployment
  - 1D, 2D, or 3D arrays

- Distributed Deployment
  - Remote radio heads

Benefits with Massive MIMO

Outdoor users: Handle mobility and provide coverage

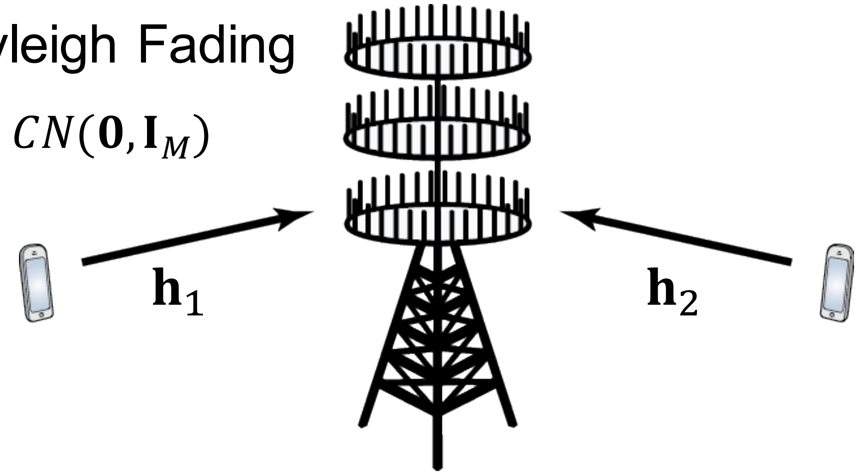
Indoor users: No need to put BSs inside buildings

Basic Properties of

MASSIVE MIMO

Asymptotic Channel Orthogonality

- Example: Uplink with Isotropic/Rayleigh Fading
  - Two users, i.i.d. channels: $\mathbf{h}_1, \mathbf{h}_2 \sim CN(\mathbf{0}, \mathbf{I}_M)$
  - Signals: s_1, s_2 with power P
  - Noise: $\mathbf{n} \sim CN(\mathbf{0}, \mathbf{I}_M)$
  - Received: $\mathbf{y} = \mathbf{h}_1 s_1 + \mathbf{h}_2 s_2 + \mathbf{n}$



- Linear Processing for User 1: $\tilde{y}_1 = \mathbf{w}_1^H \mathbf{y} = \boxed{\mathbf{w}_1^H \mathbf{h}_1} s_1 + \boxed{\mathbf{w}_1^H \mathbf{h}_2} s_2 + \boxed{\mathbf{w}_1^H \mathbf{n}}$
  - Maximum ratio filter: $\mathbf{w}_1 = \frac{1}{M} \mathbf{h}_1$
  - Signal remains: $\mathbf{w}_1^H \mathbf{h}_1 = \frac{1}{M} ||\mathbf{h}_1||^2 \xrightarrow{M \to \infty} \mathrm{E}\left[|h_{11}|^{\mathbf{2}}\right] = 1$
  - Interference vanishes: $\mathbf{w}_1^H \mathbf{h}_2 = \frac{1}{M} \mathbf{h}_1^H \mathbf{h}_2 \xrightarrow{M \to \infty} \mathrm{E}[h_{11}^H h_{21}] = 0$
  - Noise vanishes: $\mathbf{w}_1^H \mathbf{n} = \frac{1}{M} \mathbf{h}_1^H \mathbf{n} \xrightarrow{M \to \infty} \mathrm{E}[h_{11}^H n_1] = 0$

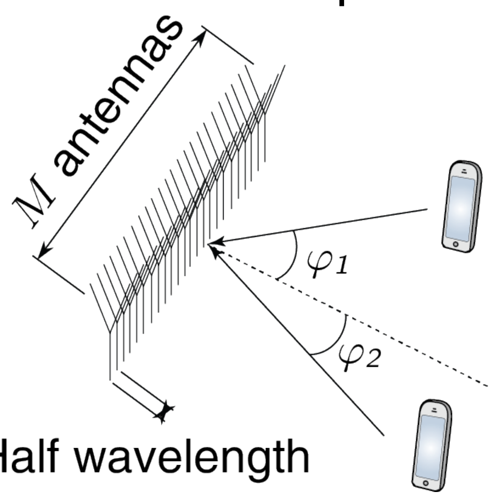
Asymptotically noise/interference-free communication: $\tilde{y}_1 \xrightarrow{M \to \infty} s_1$

Is this Result Limited to Isotropic Fading?

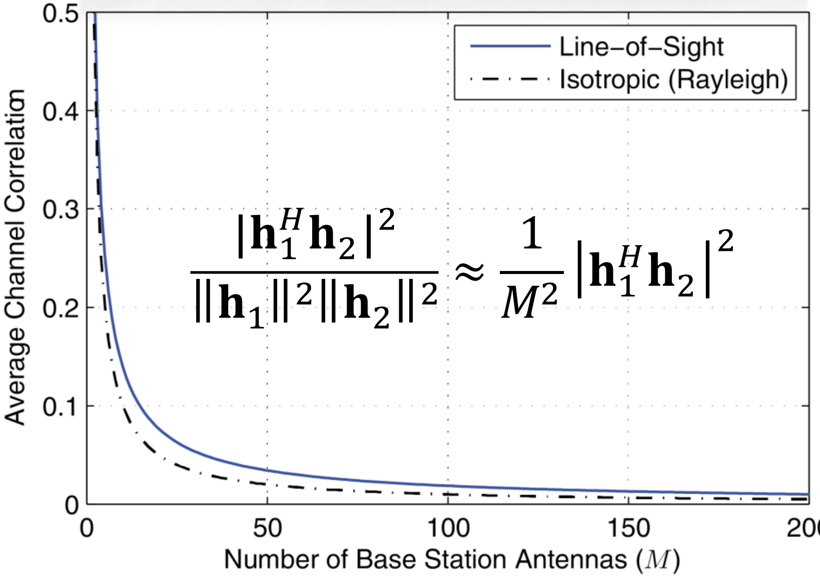
- Assumptions in i.i.d. Rayleigh Fading
  - No dominant directivity
  - Very many scattering objectives

Less true as $M \to \infty$

- Example: Line-of-Sight Propagation



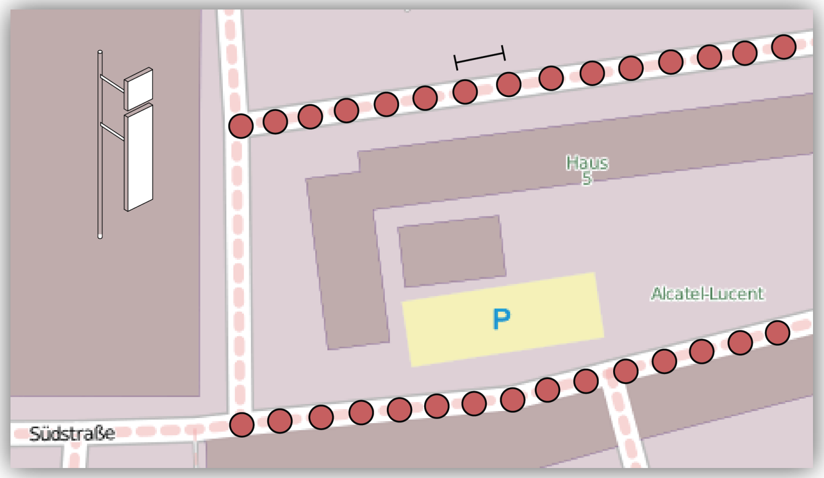
- Uniform linear array
- Random user angles
- M observations:
  - Stronger signal
  - Suppressed noise
- What is $\mathbf{h}_1^H \mathbf{h}_2 \to ?$



***Only difference**:*
How quickly interference is suppressed

How will Practical Channels Behave?

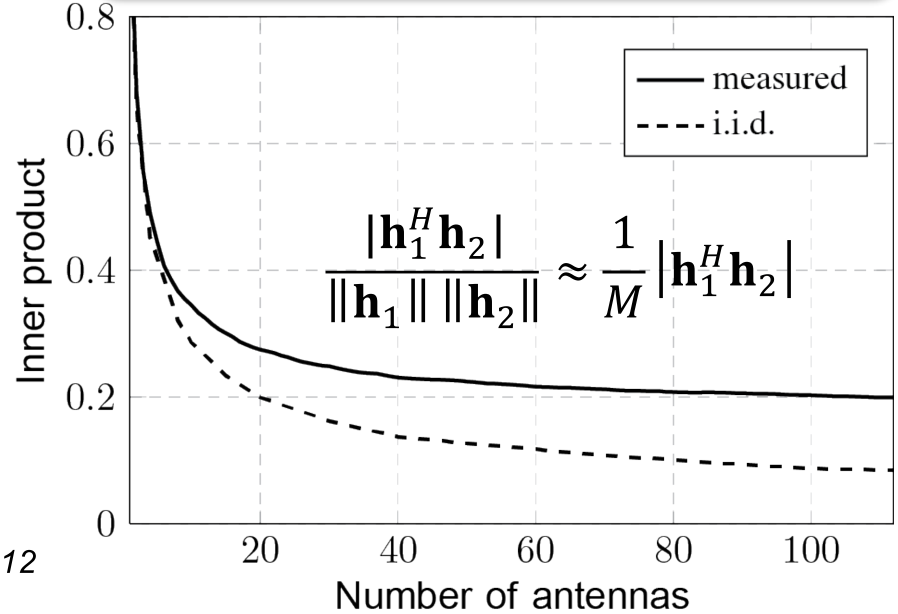
- Measurements show similar results



Source: J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel Measurements for Large Antenna Arrays," ISWCS 2012

Spectral Efficiency

Only 10-20% lower than i.i.d. fading



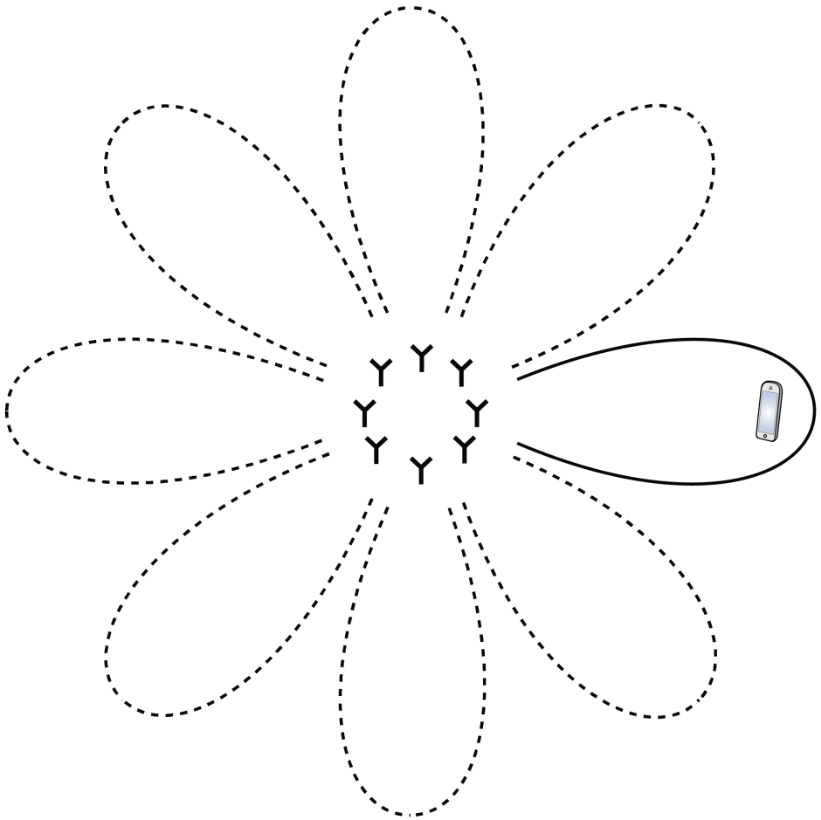
- Asymptotic Favorable Propagation: $\frac{1}{M}\mathbf{h}_1^H \mathbf{h}_2 \rightarrow 0$ as $M \rightarrow \infty$
  - Achieved in Rayleigh fading and line-of-sight – two extremes!
  - Same behavior expected and seen in practice

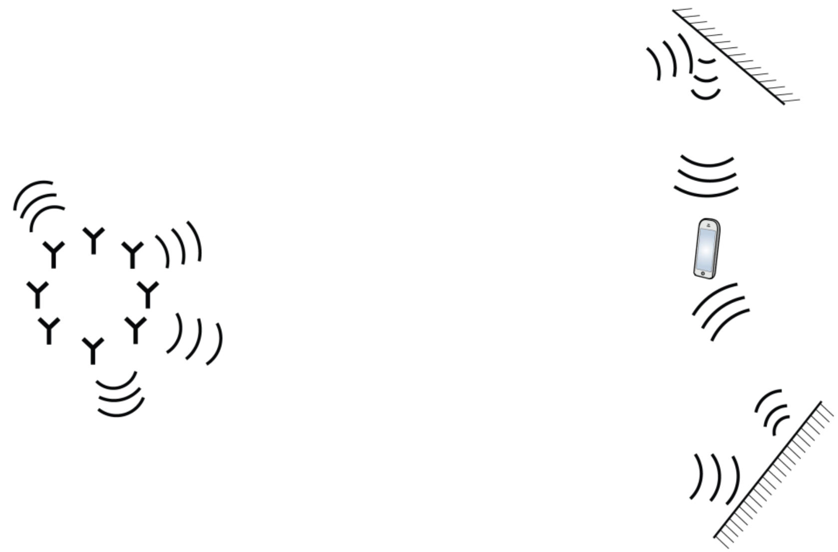
Massive MIMO

TRANSMISSION PROTOCOL

MIMO Precoding

Same principles for MIMO detection





Line-of-Sight

Channels determined by angles

1-2 parameters to estimate per user

Precoding = Beamforming

Non-Line-of-Sight

Rich multipath propagation

M parameters to estimate per user

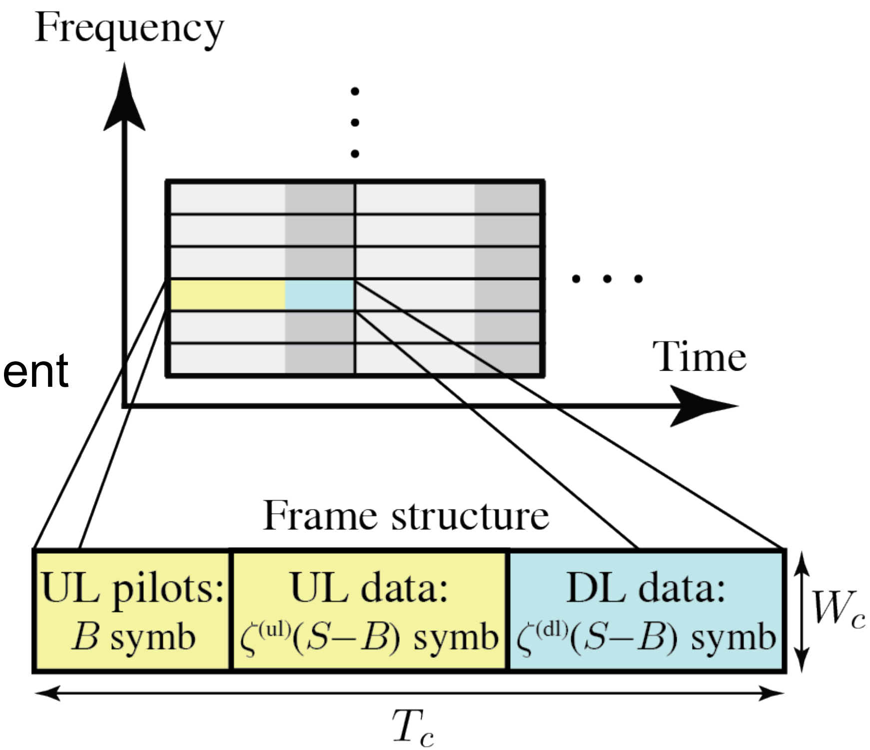
Precoding $\neq$ Beamforming

Easy: Codebooks can be used

Hard: Requires pilots!

Transmission Protocol

- Coherence Blocks
  - Fixed channel responses
  - Coherence time: T_c s
  - Coherence bandwidth: W_c Hz
  - Depends on mobility and environment
  - Block length: $S = T_c W_c$ symbols
  - Typically: $S \in [100{,}10000]$



- Time-Division Duplex (TDD)
  - Switch between downlink and uplink on all frequencies
  - B symbols/block for uplink pilots – to estimate channel responses ($B \geq K$)
  - $S - B$ symbols/block for uplink and/or downlink payload data

Linear versus Non-linear Processing

- Capacity-Achieving Non-linear Processing
  - Downlink: Dirty paper coding
  - Uplink: Successive interference cancellation

Do we need it in Massive MIMO?

Linear Processing

Bad when $M \approx K$

Good when $M/K > 2$

Relative low complexity

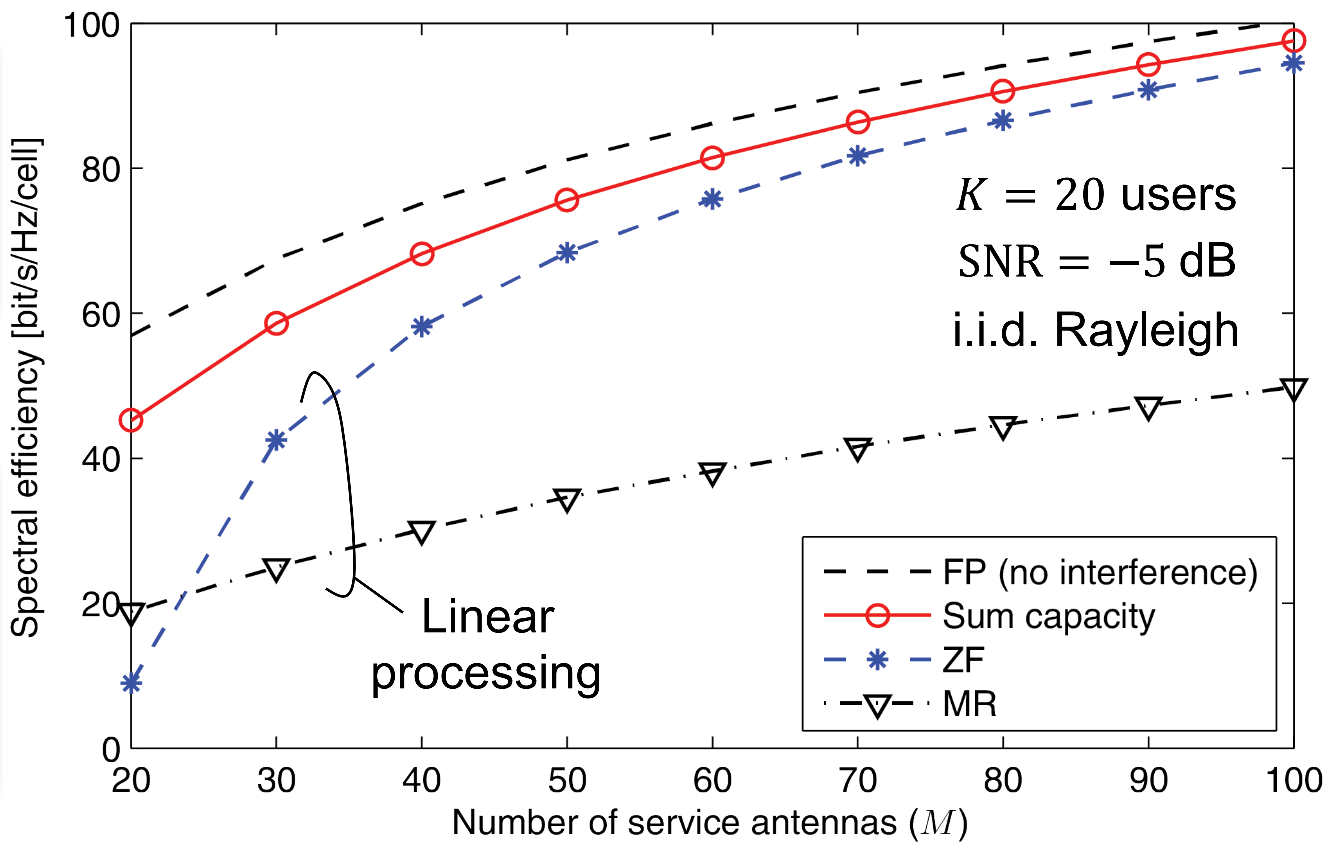
Massive MIMO

Uses linear processing:

Maximum ratio (MR)

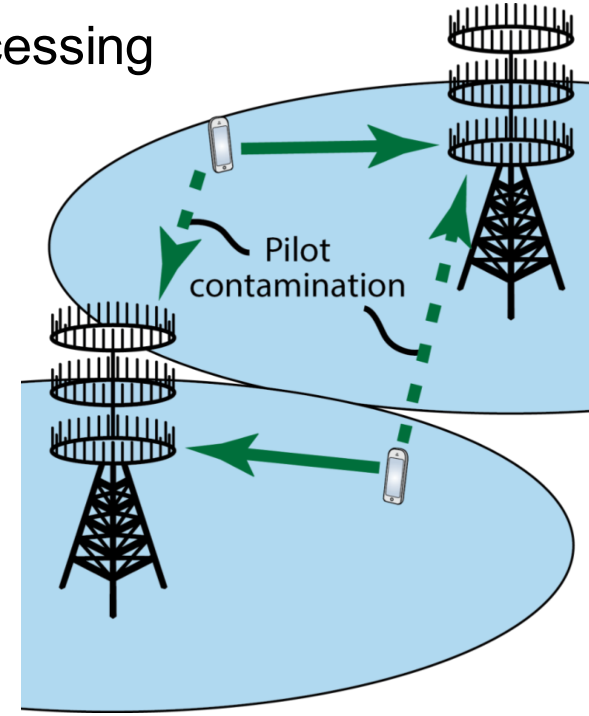
Zero-forcing (ZF)

Optimal: MMSE

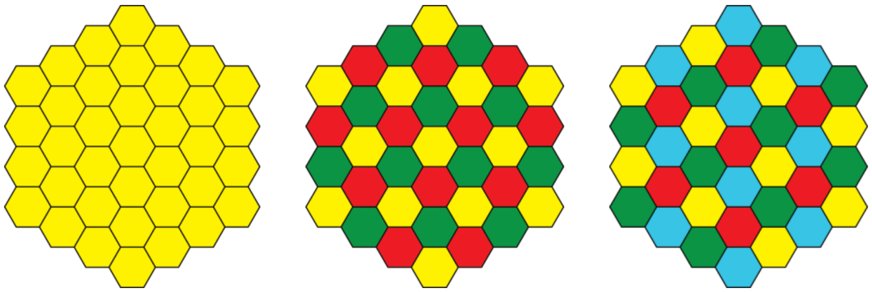


Channel Acquisition in Massive MIMO

- BS Needs Channel Responses for Linear Processing
  - Estimate using uplink pilot symbols
  - Only B pilot symbols available ($B \leq S$)
  - Must use same pilot symbols in different cells
  - BSs cannot tell some users apart



- Called: Pilot Contamination
  - Recall: Noise and interference vanish as $M \rightarrow \infty$
  - Not interference between users with same pilot!
- Solution: Select how often pilots are reused
  - Pilot reuse factor $\beta \geq 1$
  - Users per cell: $K = \frac{B}{\beta}$
  - Higher β → Fewer users per cell, but interferers further away

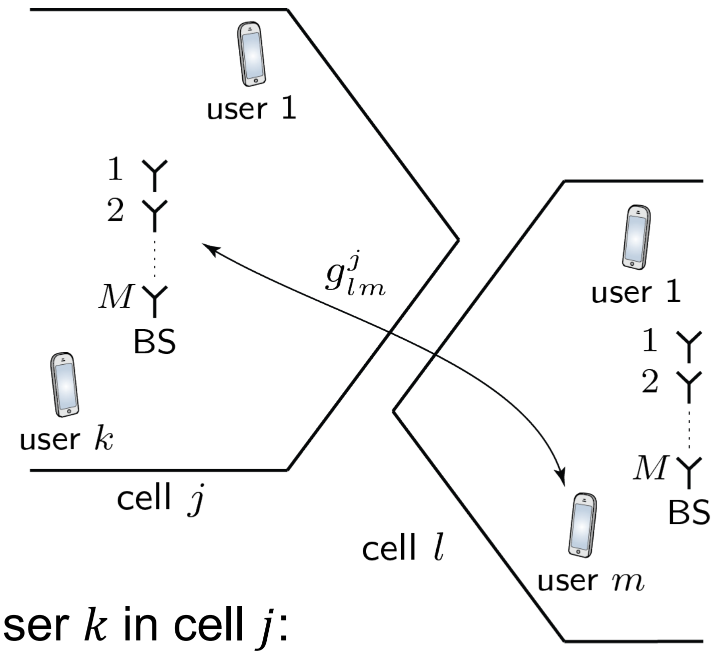


Pilot reuse $\beta = 1$ ***Pilot reuse*** $\beta = 3$ ***Pilot reuse*** $\beta = 4$

Intuitive Spectral Efficiency Expressions

- Simple Performance Expressions
  - No small-scale fading, only gain g^j_{lm}
  - Depends on transmit powers: ρ^u_{jk}, ρ^p_{jk}
  - Depends on estimation quality:

$$\gamma^j_{lk} = \frac{\rho^p_{lk} g^j_{lk}}{\sum_{l' \in \mathcal{P}_l} \rho^p_{l'k} g^j_{l'k} + 1}$$



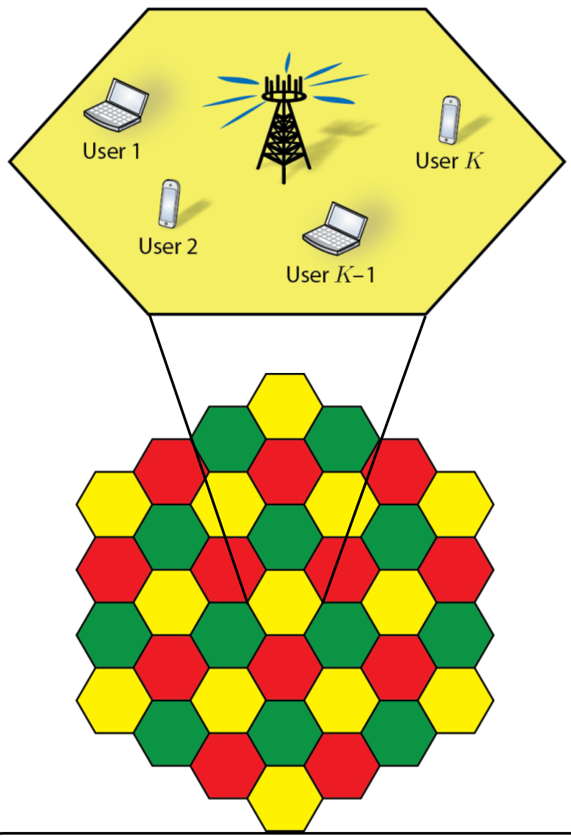
  - Ergodic spectral efficiency with MR for user k in cell j:

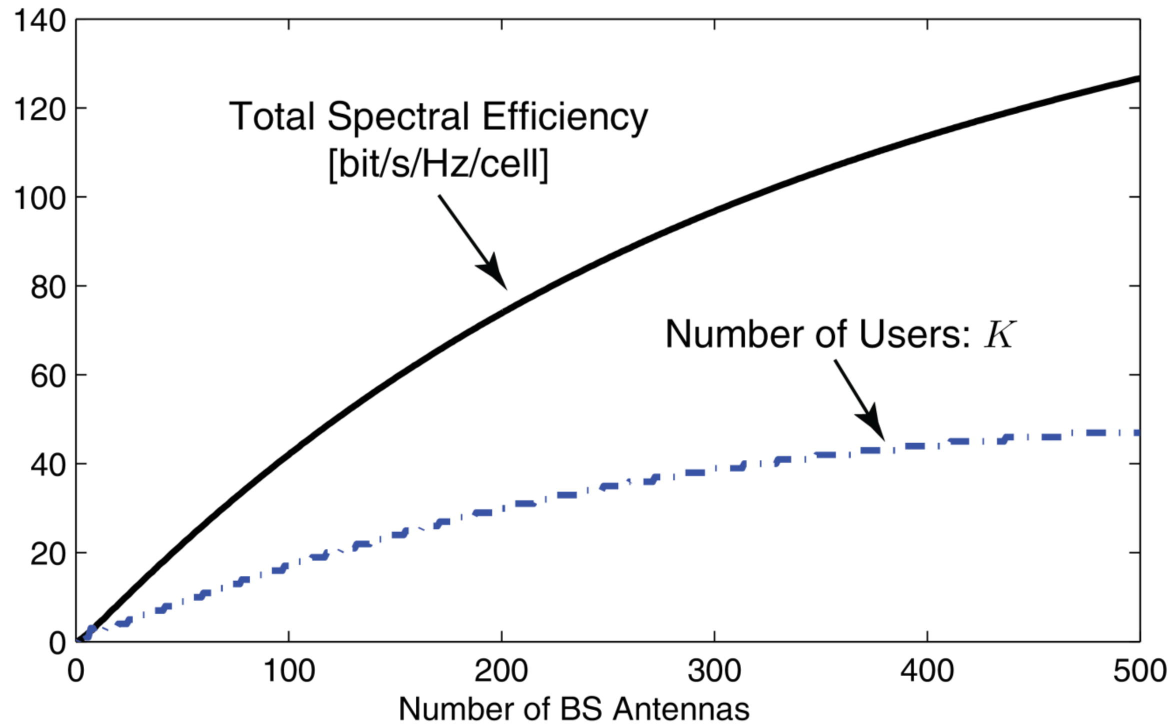
$$R_{jk} = \underbrace{\left(1 - \frac{\beta K}{S}\right)}_{\text{Pilot overhead}} \log_2 \left(1 + \frac{\overbrace{M \rho^u_{jk} \, g^j_{jk} \, \gamma^j_{jk}}^{\text{Desired signal}}}{\underbrace{\sum_{l,m} \rho^u_{lm} \, g^j_{lm}}_{\text{Conventional interference}} + \underbrace{M \sum_{l \in \mathcal{P}_j \setminus \{j\}} \rho^u_{lk} g^j_{lk} \, \gamma^j_{lk}}_{\text{Pilot-contaminated interference}} + 1} \right)$$

Array gain

Similar expressions for downlink and with ZF processing

How Much can Spectral Efficiency be Improved?





Uplink Simulation

LTE-like system parameters

Coherence block: $S = 500$

SNR 5 dB, i.i.d. Rayleigh

ZF and $\beta = 3$

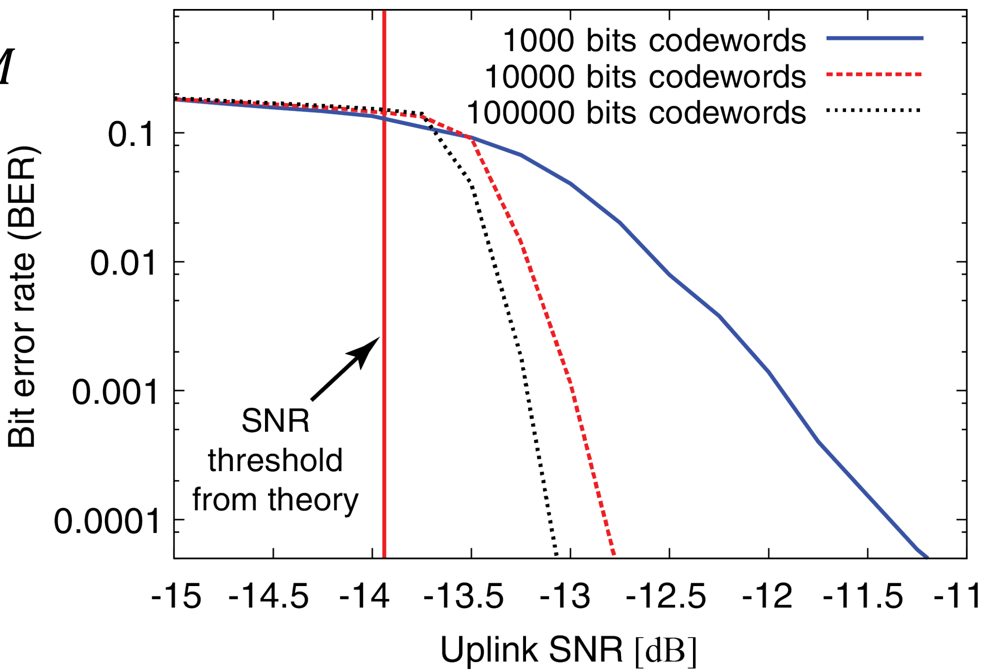
Observations

- *Baseline: 2.25 bit/s/Hz/cell (IMT-Advanced)*
- *Massive MIMO,* $M = 100$*: x20 gain* $(M/K \approx 6)$
- *Massive MIMO,* $M = 400$*: x50 gain* $(M/K \approx 9)$
- *Per scheduled user:* ≈ 2.5 *bit/s/Hz*

4 MYTHS AND MISCONCEPTIONS

Massive MIMO Relies on Asymptotic Results

- No! Accurate theory for any M
  - **Example**:
    $M = 100, K = 30$
    QPSK with ½-LDPC code
    Total: 30 bit/s/Hz
  - **Reality**:
    Rate expressions reached
    at modest codeword lengths



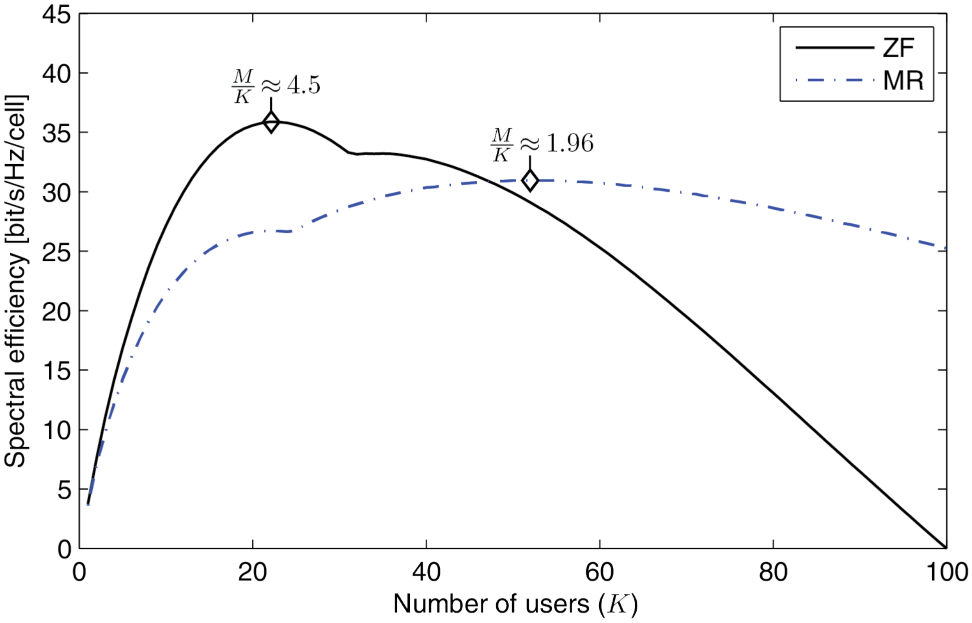
Signal Processing Complexity is Overwhelming

- **Reality:** It is higher, but manageable:
  - Most processing can be parallelized per antenna
    (e.g., FFTs, channel estimation, precoding/detection of data)
  - Not parallelized: Computing ZF/MMSE matrix inversion, but:
    - Not the main complexity (happens only $1/S$ of the time)
    - Good inversion approximations (diagonal-dominant matrices)

We Need Orders of Magnitude More Antennas than Users

- No! It all depends on the goal
  - **Example**:
    $M = 100,\ S = 200$
    SNR = -5 dB from own BS
  - **Reality**:
    Any $M/K > 2$ is desirable

    Choose between:
    Many or few users →
    Low or high rates/user



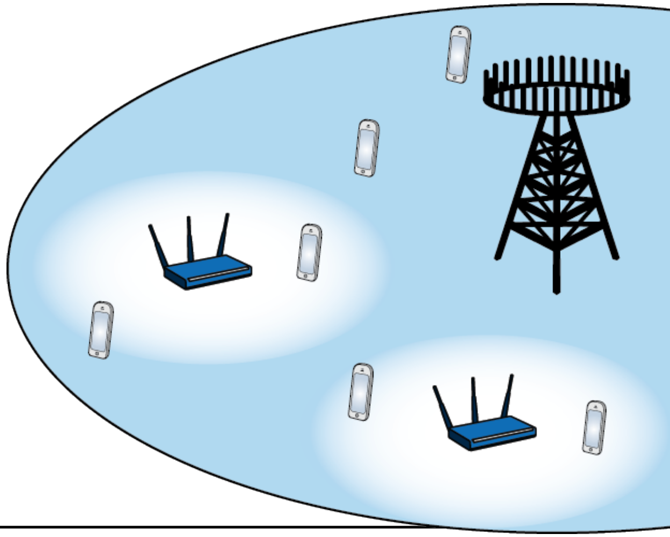
Resource Allocation is Impossible with so Many Antennas

- **Classically:** Allocate time/frequency blocks based on current fading
- **Reality:** Not needed in Massive MIMO!
  - Small-scale fading does not impact channel quality
  - Everyone can get the whole bandwidth whenever needed!
  - Remaining issue: Power control, but can be handled as a convex problem

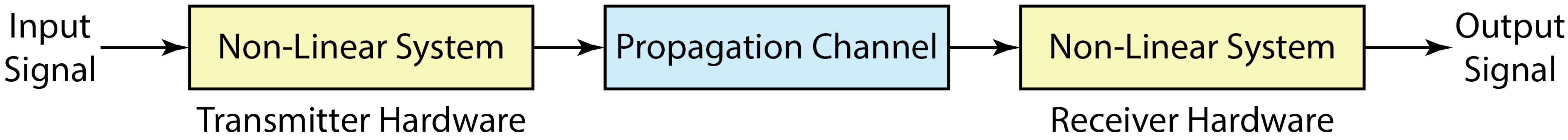
RESEARCH TRENDS AND OPEN QUESTIONS

Research Trends and Open Questions

- Coexistence with Other Technologies
  - How to balance traffic load over BS types?
  - Can it tolerate underlaying (e.g., D2D)?
  - Is TDD time synchronization manageable?
  - Are mmWave frequencies suitable?



- Implementation-Based Hardware and Algorithmic Design
  - Transceiver hardware is imperfect and affects the system:



  - Fact: Massive MIMO tolerates larger hardware impairments
  - How to utilize this to tailor the transceiver hardware?
  - Data shuffling is a bottleneck: Develop easily implementable algorithms

Can Massive MIMO work in FDD mode?

- Frequency-division duplex (FDD) is used in many systems
  - Different uplink/downlink frequencies → Two-way pilots & feedback needed

TDD versus FDD

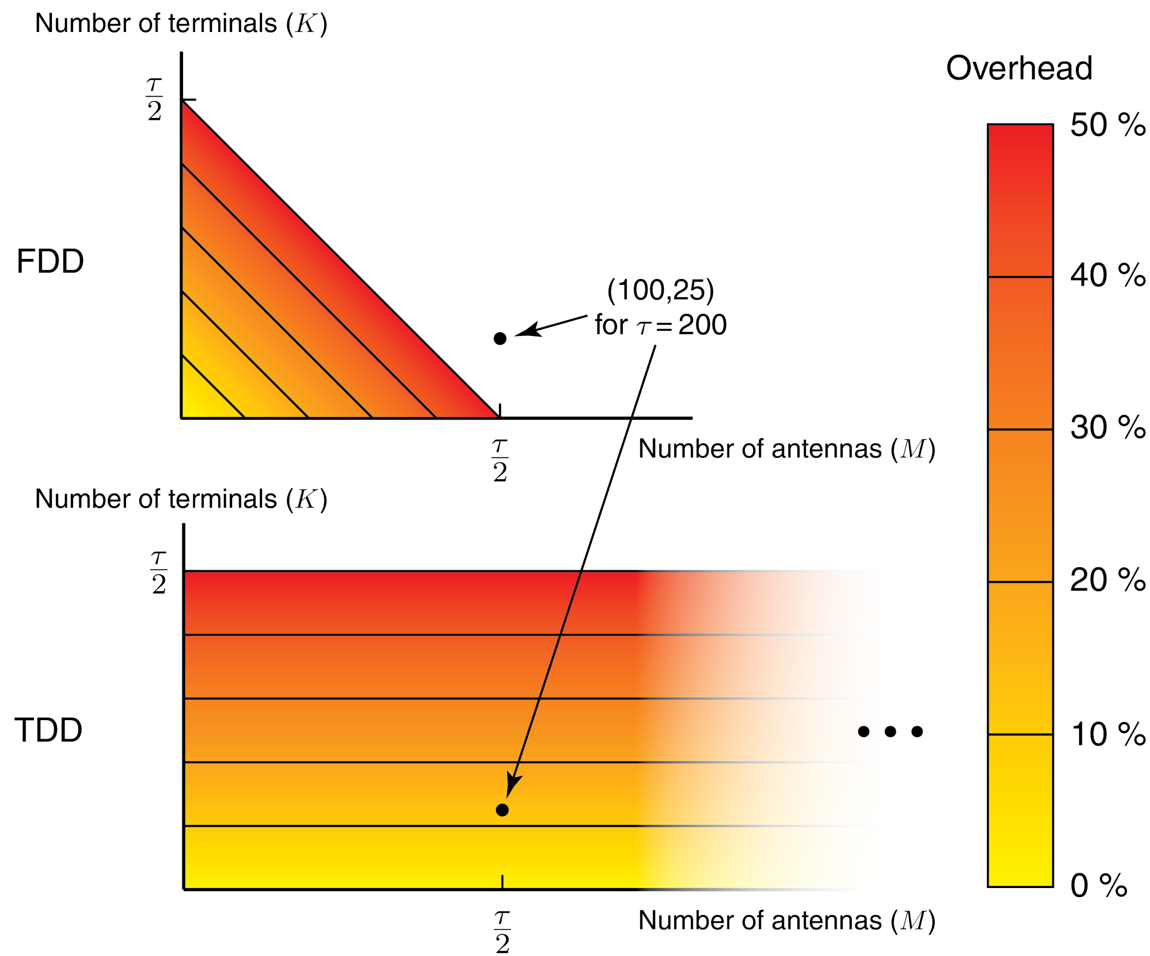
Channel coherence limits both antennas and users in FDD, but only users in TDD

FDD possible with low mobility

Open question

Can we reduce estimation and feedback load?

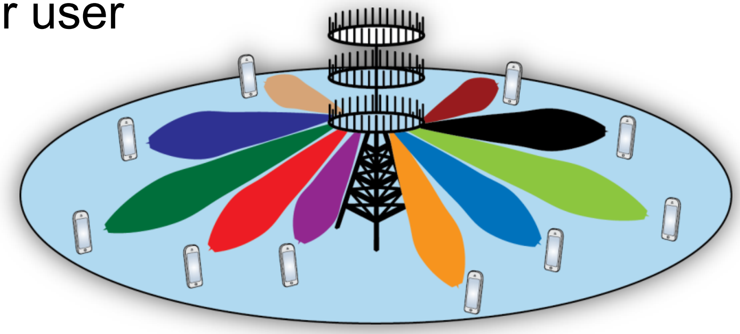
Yes, under channel sparsity: Exists in line-of-sight, but probably not in general!



SUMMARY

Summary

- Massive MIMO: The way to increase spectral efficiency in 5G networks
  - >20x gain over IMT-Advanced are foreseen
  - BSs with many small antennas and transceiver chains
  - Higher spectral efficiency per cell, not per user
  - Many potential deployment strategies



- Facts to Remember
  - Massive MIMO $\neq$ Massive size: TV sized panels at cellular frequencies
  - Favorable propagation in most propagation environments
  - Resource allocation and processing are simplified, not complicated
- Further Reading
  - Emil Björnson, Erik G. Larsson, Thomas L. Marzetta, "*Massive MIMO: 10 Myths and One Grand Question*," Submitted to IEEE Commun. Magazine.

QUESTIONS?

Dr. Emil Björnson

Slides and papers available online:
http://www.commsys.isy.liu.se/en/staff/emibj29

Would you like to learn more?

Come to Stockholm on June 29! I give a tutorial on this topic at IEEE SPAWC 2015!