IEEE International Symposium on Personal, Indoor and Mobile Radio Communications
13–16 September 2021 // Virtual Conference // By 6G Flagship

User-Centric Cell-Free Massive MIMO:

From Foundations to Scalable Implementation

Emil Björnson[*,***], Luca Sanguinetti[**], Özlem Tuğfe Demir[***]

[*]Department of Electrical Engineering (ISY), Linköping University, Sweden

[**]Dipartimento di Ingegneria dell’Informazione, University of Pisa, Italy

[***]Department of Computer Science, KTH Royal Institute of Technology, Sweden

IEEE International Symposium on Personal, Indoor and Mobile Radio Communications

13–16 September 2021 // Virtual Conference // By 6G Flagship

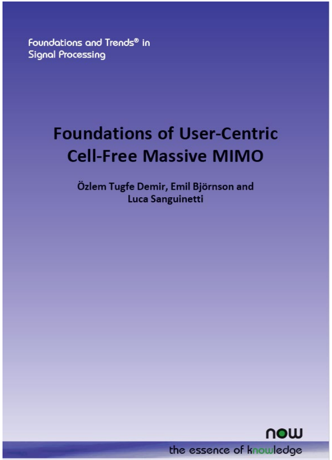
User-Centric Cell-Free Massive MIMO:

From Foundations to Scalable Implementation

Emil Björnson,
Visiting Professor,
KTH Royal Institute of Technology

Luca Sanguinetti,
Associate Professor,
University of Pisa

Özlem Tuğfe Demir,
Postdoc,
KTH Royal Institute of Technology



https://www.nowpublishers.com/article/Details/SIG-109

https://github.com/emilbjornson/cell-free-book

Outline

- Cellular Communications and Massive MIMO
- Background and Motivation of Cell-free Massive MIMO
- Foundation and System Model
- Uplink and Downlink System Operation
- Spatial Resource Allocation
- Takeaways and Open Problems

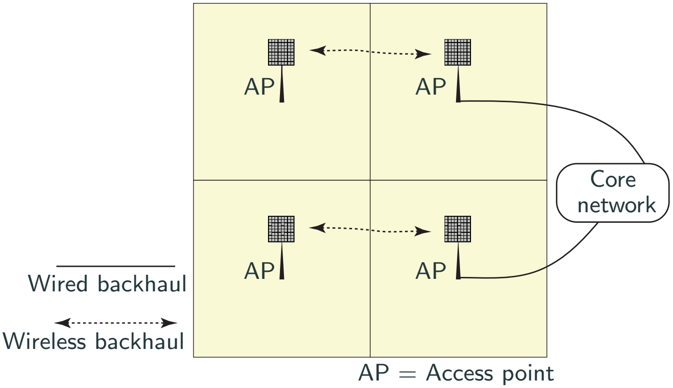
Cellular Communications and Massive MIMO

The Cellular Network Paradigm

Designed for mobile telephony

Bullington, K. (1953). "Frequency economy in mobile radio bands". The Bell System Technical Journal.

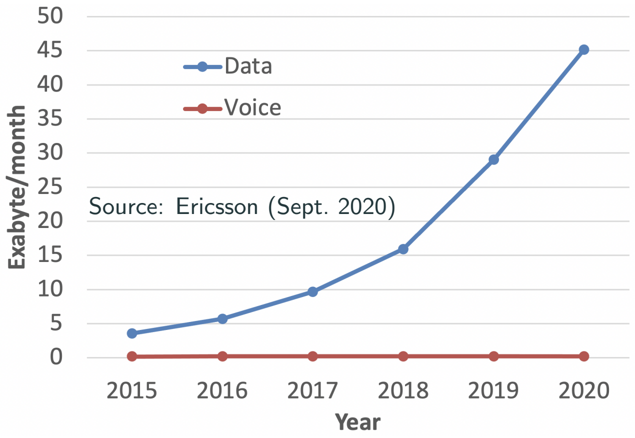
Schulte, H. J. and W. A. Cornell (1960). "Multi-area mobile telephone system". IEEE Trans. Veh. Technol.



Benefits

- Reuse of spectrum in space: Densify as usage increases
- Control interference by fractional spectrum reuse, reduce power

Current Trend: Exponential Traffic Growth

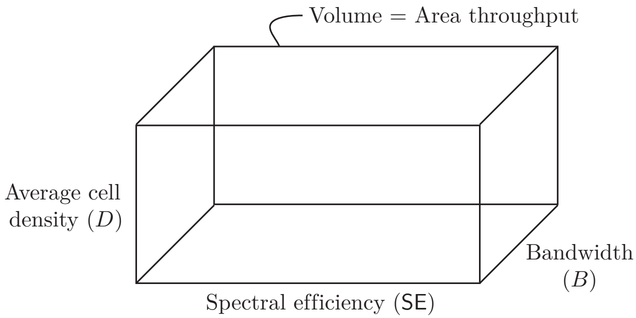


What grows?

- Voice: 1% per year
- Data: 60% per year

Service quality: Voice services require 10-100 kbit/s, not more
Data services better if data rate is higher, require 100 Mbit/s or more

Area Throughput Matters for Data Transmission



Spectral efficiency breakdown

$$\mathsf{SE} = \sum_{\text{stream } i} \mathsf{SE}_i$$

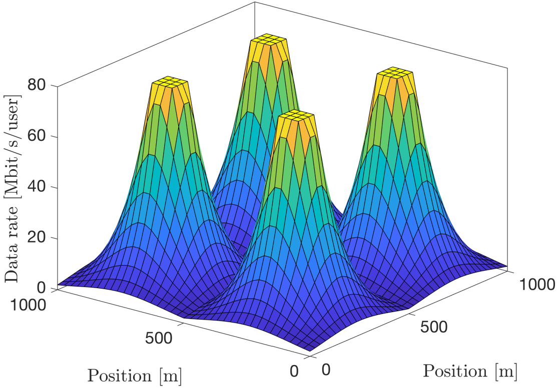
$$\mathsf{SE}_i = \begin{cases} 2 & \text{QPSK} \\ 4 & \text{16-QAM} \\ 6 & \text{64-QAM} \end{cases}$$

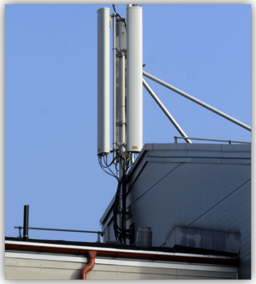
Definition: The *area throughput* of a cellular network is measured in bit/s/km^2:

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

where B = bandwidth, D = average cell density, SE = per-cell *spectral efficiency (SE)*

Sparse Deployment of Access Points



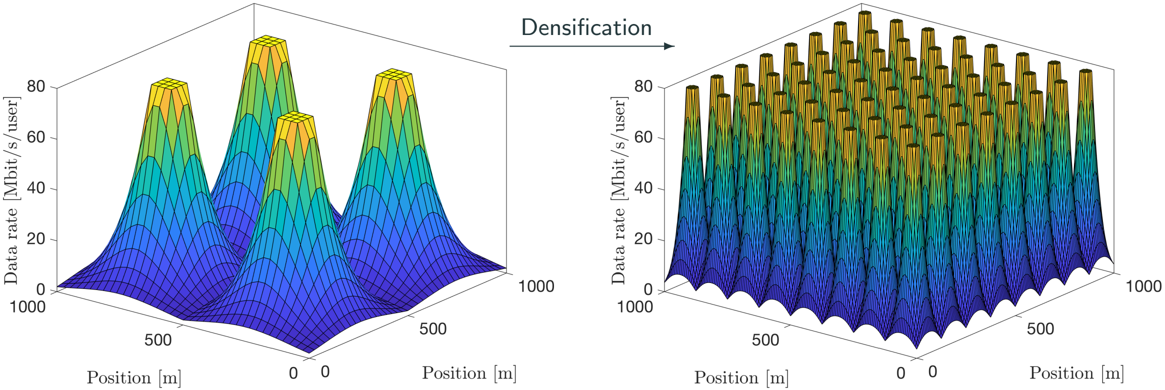


Large pathloss variations

- Most *active* users are the edge
- Regular service interruption at cell edges
- High power, visible installation

Suitable for coverage
Not for uniform data service

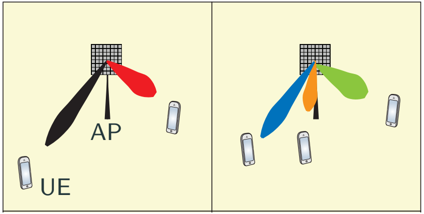
The End of Cell Densification Gains



Interference-limited regime

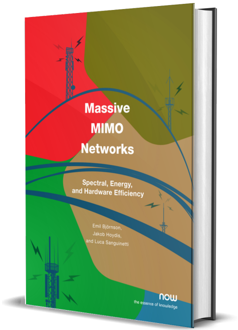
- Small APs cannot reject interference
- Most users get mediocre performance

Massive MIMO (multiple-input multiple-output) to Deal With Interference

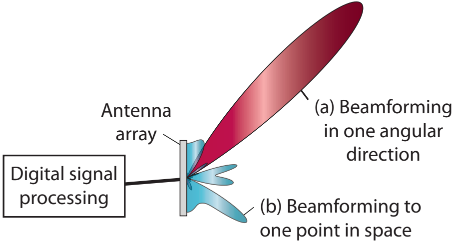


L APs, each with N antennas

- Many user equipments (UEs)
- Each UE served by one AP
- No information exchange
- Beamforming gains
- Multiplexing gains



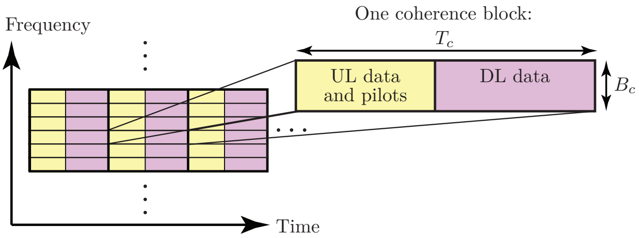
Free PDF:
`massivemimobook.com`



Massive MIMO in 5G

- ≥ 64 antennas per AP
- Up to 8 users per cell
- Mid and high bands

Massive MIMO TDD Operation in a Nutshell



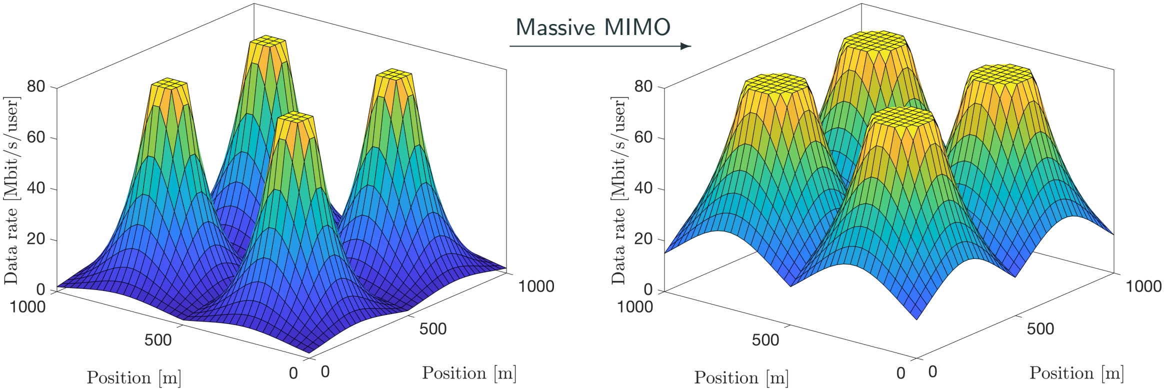
Channel change in time/frequency

$\approx$ Constant for T_c s
$\approx$ Constant for B_c Hz
Coherence block: $T_c B_c$ samples

Time-division duplex (TDD) operation of a coherence block

1. Uplink pilots: Each UE transmit a “pilot” to enable channel estimation
2. Uplink data: Each UE transmits data, AP decodes based on channel estimates
3. Downlink data: AP beamforms data to UEs based on channel estimates

Expected Gains from "Cellular" Massive MIMO



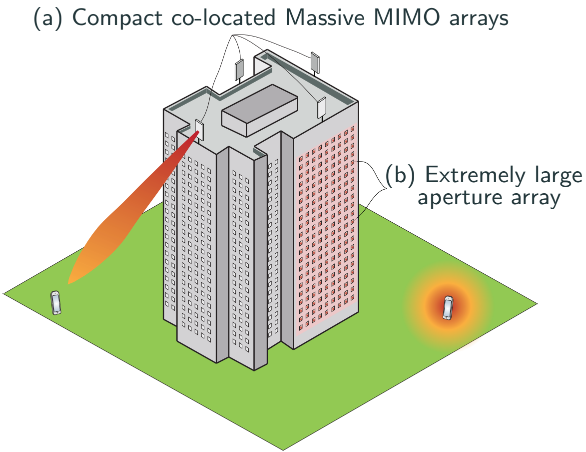
Beamforming and multiplexing

- Higher SNRs for every UE
- Multiple parallel data streams

Problems that remain

- Large SNR variations
- Interference due to "small" arrays

Evolution of Cellular Massive MIMO Beyond 5G



Physically large arrays

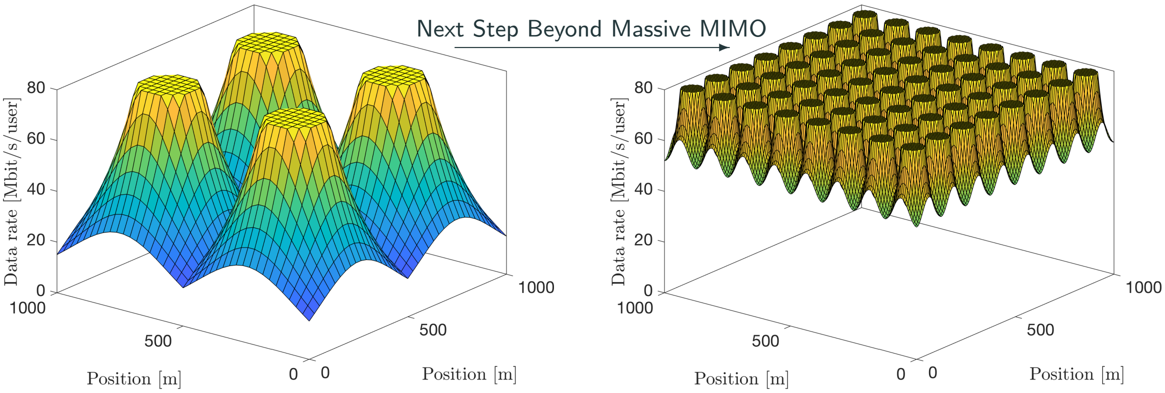
- More aggressive multiplexing
- Handle interference better
- Near-field propagation

SNR variations remain!

- Sensitive to blocking
- Cell-edge issues

[BSW+19] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, T. L. Marzetta, "Massive MIMO is a Reality—What is Next? Five Promising Research Directions for Antenna Arrays," DSP 2019.

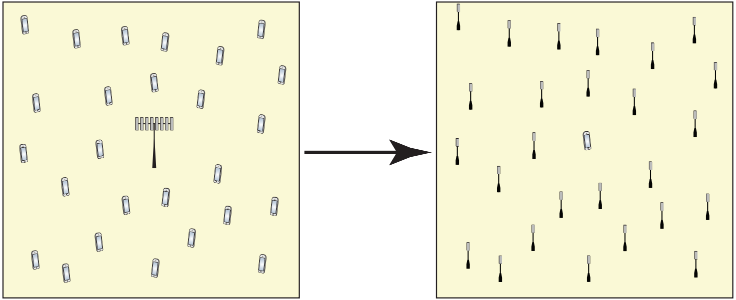
Ambition: Deliver Almost Uniformly Good Performance Everywhere



How can we achieve this?
Address both SNR variations and interference!

Background and Motivation of Cell-free Massive MIMO

Paradigm Shift: A User-centric Cell-free Networks

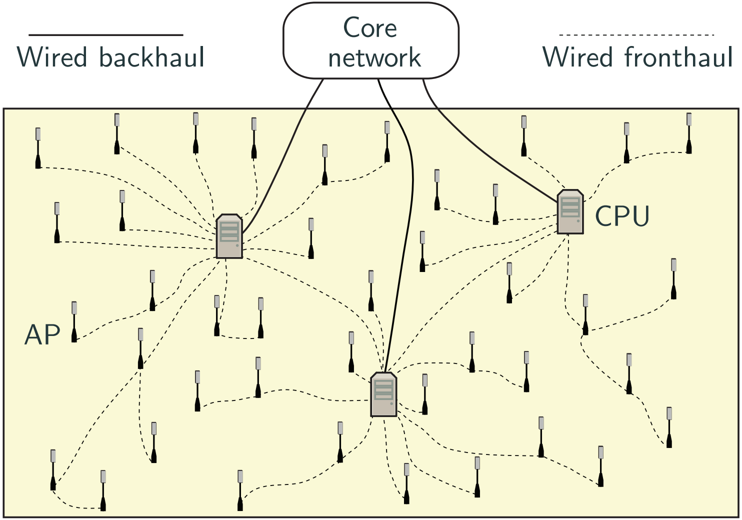


AP surrounded by users | UE surrounded by antennas

Benefits:

- No cells – no inter-cell interference
- Short distance to *some* of the antennas
- Massive macro-diversity against shadowing
- Each user served by *all* antennas that reach it

Implementation Using Central Processing Units

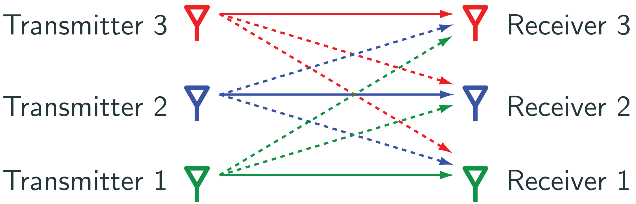


Three-level architecture

- Access points (APs)
- Central processing units (CPUs), edge-clouds
- Core network

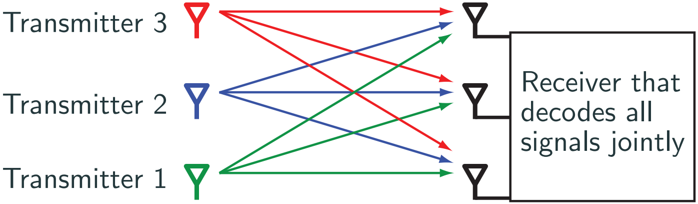
Essentially: A cell-free physical layer built on a cloud-RAN (C-RAN) architecture

Philosophy of Uplink Interference Rejection



Interference channel

- 1 observation
- 1 desired signal
- 2 interfering signals
- Too few observations



Multiaccess channel

- 3 observations
- 3 desired signals
- Can reject interference

[W94] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," IEEE Transactions on Information Theory, 1994.

Brief Historical Background

1. Interference avoidance and decoding, 1950–2005
    - Frequency reuse patterns, sectorization, etc.
    - Successive interference cancelation (e.g., NOMA)
2. Spatial interference rejection: 2000–2010
    - Turn interference channels into multiaccess/broadcast channels
    - Concept that underpins Cell-free Massive MIMO
    - Many names: Network MIMO, Group cells, multicell cooperative network, ...
3. Coordinated multipoint: 2005–2015
    - An attempt to implement these concepts in 4G
    - Add-on to cellular networks, not particularly successful
4. Cell-free Massive MIMO networks: 2015–
    - Restart of the topic in [NAY+15], using methodology from Cellular Massive MIMO
    - Focused on scalability, many more APs than UEs, distributed architecture...

[NAY+15] H. Q. Ngo and A. E. Ashikhmin and H. Yang and E. G. Larsson and T. L. Marzetta, "Cell-free Massive MIMO: Uniformly great service for everyone," IEEE SPAWC 2015.

One Possible Definition

Definition (Cell-free Massive MIMO)

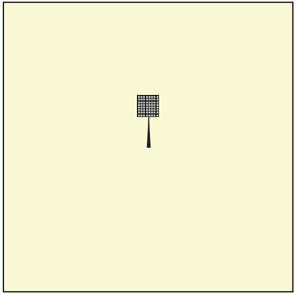
A network with L APs jointly serving all the UEs, where

- each AP has N antennas (N is fairly small);
- there are $K \ll L$ UEs;
- the APs are interconnected via fronthaul to one or multiple CPUs.

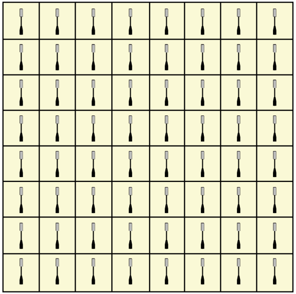
A few ways to think of it:

$$\text{Cell-free Massive MIMO} = \begin{cases} \text{Ultra-dense network} + \text{Network MIMO} \\ \text{Network MIMO} + \text{Massive MIMO operation} \\ \text{Network MIMO} + \text{C-RAN} + \text{Distributed antennas} \end{cases}$$

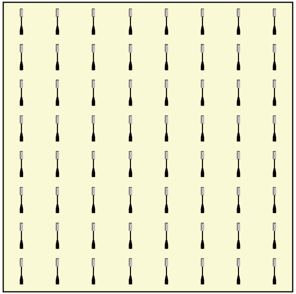
Basic Comparison Between Cellular and Cell-free Networks



Cellular Massive MIMO:
One 64-antenna AP



Small cells:
64 single-antenna APs

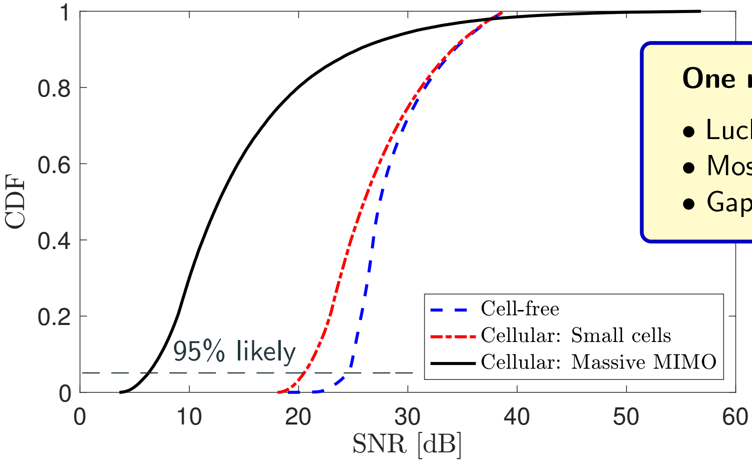


Cell-free:
64 single-antenna APs

Uplink simulation, area: $400 \times 400\,\text{m}$

Difference: SNR distribution and interference capabilities

Cell-free Benefit 1: Higher SNR With Smaller Variations

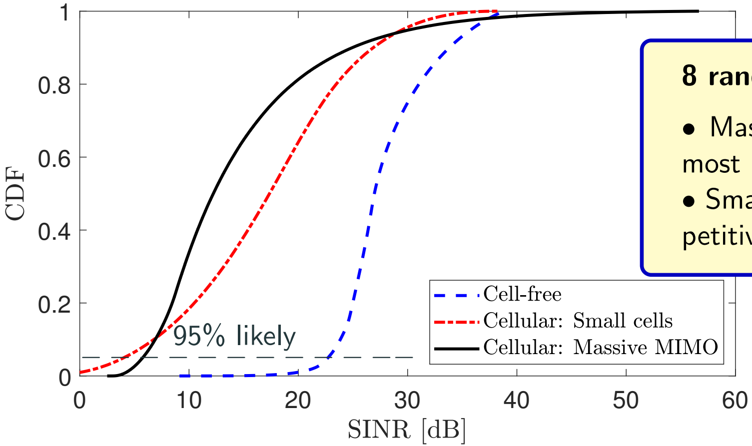


One randomly located user

- Lucky UEs prefer Massive MIMO
- Most UEs prefer Cell-free
- Gap to small cells is small

Transmit power: $10\,\text{dBm}$; Noise power: $-96\,\text{dBm}$
Channel gain: $\beta(d)\,[\text{dB}] = -30.5 - 36.7\log_{10}(d/1\,\text{m})$

Cell-free Benefit 2: Better Ability to Manage Interference

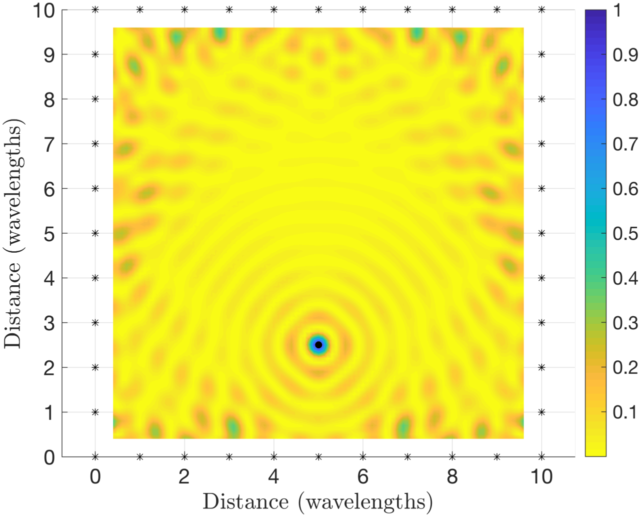


8 randomly located users

- Massive MIMO and cell-free almost unaffected by interference
- Small-cell system no longer competitive

Transmit power: $10\,\text{dBm}$; Noise power: $-96\,\text{dBm}$
Channel gain: $\beta(d)\,[\text{dB}] = -30.5 - 36.7\log_{10}(d/1\,\text{m})$

Cell-free Benefit 3: Precise Signal Focusing



Beamforming, compact array

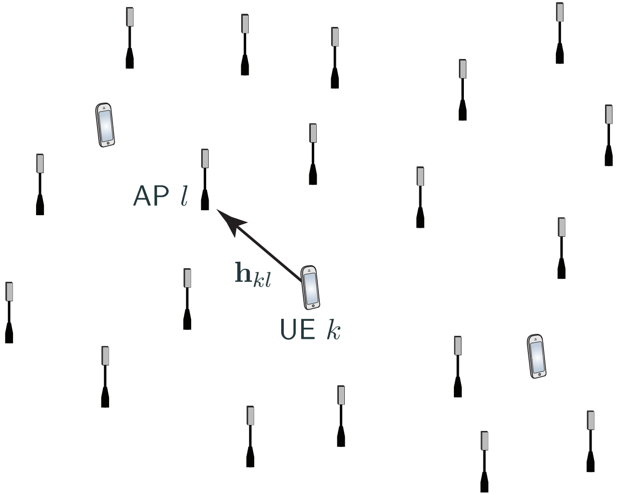
- Strong signal in angular interval
- Limited resolution

Beamforming, distributed array

- Focus signal on a point
- Size proportional to wavelength

Foundation and System Model

Channel Model and Notation



Channel notation:

$$\mathbf{h}_{\text{UE's number, AP's number}} \in \mathbb{C}^N$$

Correlated Rayleigh fading:

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl})$$

- Spatial correlation matrix and pathloss: $\mathbf{R}_{kl} \in \mathbb{C}^{N \times N}$
- One realization per coherence block

Uplink System Model and Receive Combining

Received complex baseband signal at AP l:

$$\mathbf{y}_l^{\text{ul}} = \sum_{i=1}^{K} \mathbf{h}_{il} s_i + \mathbf{n}_l \in \mathbb{C}^N$$

Assumption: Entire network is synchronized in time (e.g., OFDM with extended CP)

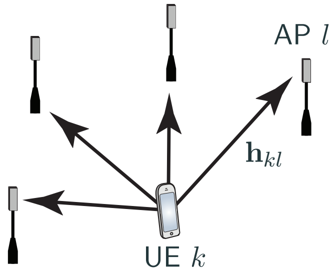
- $s_i \sim \mathcal{N}_{\mathbb{C}}(0, p_i)$ is data signal from UE i with power p_i
- $\mathbf{n}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\text{UL}}^2 \mathbf{I}_N)$ is independent receiver noise

Estimate of s_k at AP l:

$$\widehat{s}_{kl} = \mathbf{v}_{kl}^{\text{H}} \mathbf{y}_l^{\text{ul}}$$

- $\mathbf{v}_{kl} \in \mathbb{C}^N$ is receive combining vector
- Should be selected based on channel knowledge

Principle of Uplink Channel Estimation



UE l sends a predefined pilot signal $\phi \in \mathbb{C}$

- Received signal at AP l

$$\mathbf{y}_l^{\text{pilot}} = \mathbf{h}_{kl}\phi + \mathbf{n}_l$$

- Simple channel estimate:

$$\hat{\mathbf{h}}_{kl} = \frac{\phi^\star}{|\phi|^2}\mathbf{y}_l^{\text{pilot}} = \mathbf{h}_{kl} + \frac{\phi^\star}{|\phi|^2}\mathbf{n}_l$$

Any number of APs can estimate the channel from a single pilot transmission!

Cannot afford one pilot per UE: Share τ_p pilots

Downlink System Model and Transmit Precoding

Received complex baseband signal at UE k:

$$y_k^{\mathrm{dl}} = \sum_{l=1}^{L} \mathbf{h}_{kl}^{\mathrm{H}} \mathbf{x}_l + n_k$$

- $\mathbf{x}_l \in \mathbb{C}^N$ is signal transmitted from AP l
- $n_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{DL}}^2)$ is independent receiver noise

Data transmission from AP l:

$$\mathbf{x}_l = \sum_{i=1}^{K} \mathbf{w}_{il} \varsigma_i$$

- $\varsigma_i \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ is data signal to UE i
- $\mathbf{w}_{il} \in \mathbb{C}^N$ is precoding vector from AP l to UE i

What We Have Described Thus Far is Not Scalable

Each AP serves all K UEs

- Must estimate K channels
- Apply compute K combining/precoding vectors
- Must send/receive data on fronthaul for K UEs

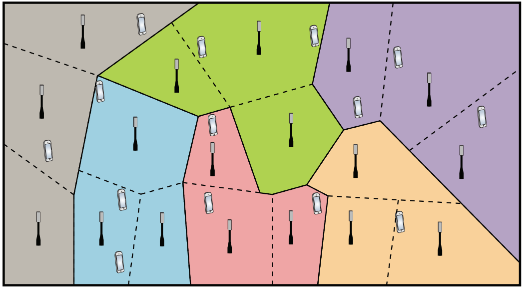
We cannot let
$K \to \infty$
Not scalable!

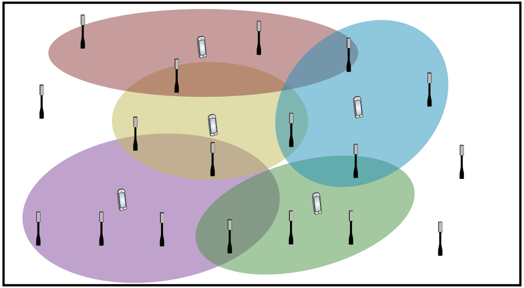
Assumption:
Each AP has a local processor – We should limit complexity per AP

Definition (Scalable operation)

A cell-free network is *scalable* if the computational complexity and fronthaul signaling per AP remains finite as $K \to \infty$

Approaches to Scalability: Network-centric and User-centric Clustering

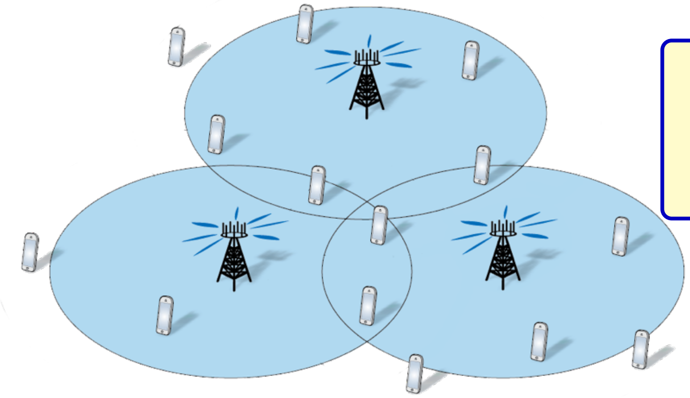




Network-centric clustering
Main approach in 4G
Not successful, not cell-free

User-centric clustering
Each UE served by all surrounding APs
Cell-free philosophy

Dynamic Cooperation Clusters (DCC)



A general framework enabling *"unified analysis of anything from interference channels to ideal network MIMO"*

- Originally proposed to enable scalable implementation of Network MIMO

[BJ13] E. Björnson, E. Jorswieck, "Optimal Resource Allocation in Coordinated Multi-Cell Systems," Foundations and Trends in Communications and Information Theory, 2013.

Solving the Scalability Issue

Definition (Scalable operation)

A cell-free network is *scalable* if the computational complexity and fronthaul signaling per AP remains finite as $K \to \infty$

Recall:

- There are τ_p orthogonal pilots per coherence block
- Large network: $K \gg \tau_p$

Solution: Each AP

- Serves at most τ_p users (one per pilot)
- Computes local channel estimates for only those τ_p users
- Computes τ_p combining/precoding vectors only based on the τ_p estimates

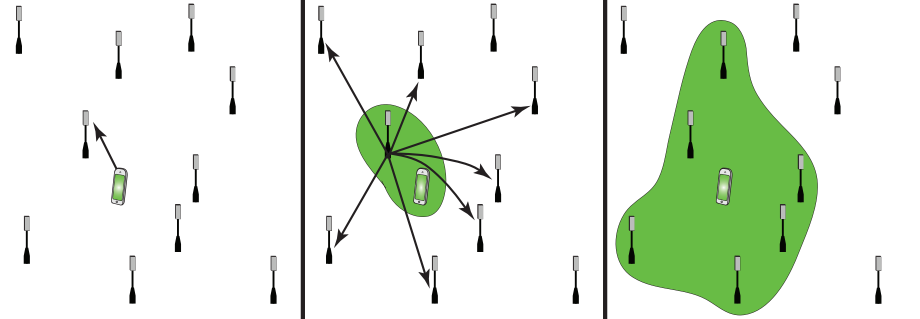
Scalable User-Centric Cooperation Clusters

One access-based algorithm:

Step 1:
UE appoints Master AP

Step 2:
Pilot assignment
invitation to other APs

Step 3:
Formation of UE cluster



[ES20] E. Björnson, L. Sanguinetti, “Scalable Cell-Free Massive MIMO Systems,” IEEE Trans. Communications, 2020.

Include DCC in System Models

UE k in served by APs with indices in $\mathcal{M}_k \subset \{1, \ldots, L\}$.

Uplink operation: AP l estimates s_k as

$$\widehat{s}_{kl} = \begin{cases} \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{y}_l^{\mathrm{ul}} & l \in \mathcal{M}_k \\ 0 & l \notin \mathcal{M}_k \end{cases} = \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{y}_l^{\mathrm{ul}}, \qquad \mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & l \in \mathcal{M}_k \\ \mathbf{0} & l \notin \mathcal{M}_k \end{cases}$$

Downlink operation: AP l transmits

$$\mathbf{x}_l = \sum_{i=1}^{K} \mathbf{D}_{il} \mathbf{w}_{il} \varsigma_i, \qquad \underbrace{\mathbf{D}_{il} \mathbf{w}_{il}}_{\text{Effective precoder}} = \begin{cases} \mathbf{w}_{il} & l \in \mathcal{M}_i \\ \mathbf{0} & l \notin \mathcal{M}_i \end{cases}$$

Channel Modeling

Channel Modeling

Deterministic and stochastic channel models can be used in wireless.

A classical stochastic model for NLoS communications is *uncorrelated Rayleigh fading*:

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \beta_{kl}\mathbf{I}_N)$$

- The complex Gaussian distribution accounts for the random small-scale fading;
- The variance β_{kl} describes the large-scale fading.

Inadequate in practice since *spatial correlation* is originated:

1. Some spatial directions are more likely than other directions;
2. Array geometry makes it more suited to transmit/receive signals in some directions than in other directions.

Correlated Rayleigh Fading

The two phenomena can be captured by the *correlated Rayleigh fading model*:

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{kl})$$

where $\mathbf{R}_{kl} \in \mathbb{C}^{N \times N}$ is the spatial correlation matrix between AP l and UE k, and

$$\beta_{kl} = \frac{1}{N} \mathrm{tr}\left(\mathbf{R}_{kl}\right).$$

Channel vectors of different APs are independently distributed (APs are separated by tens of wavelengths or more). The collective channel is thus:

$$\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_k)$$

where $\mathbf{R}_k = \mathrm{diag}(\mathbf{R}_{k1}, \ldots, \mathbf{R}_{kL}) \in \mathbb{C}^{M \times M}$ is block-diagonal.

Local Scattering Model (1/2)

The spatial correlation matrix depends on two main factors:

1. the array geometry;
2. the angular distribution of the multipath components.

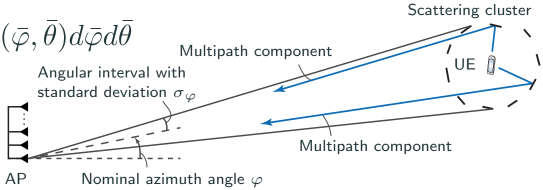
Cell-free networks envision small-sized APs

- ...common to utilize a ULA with N antennas equally spaced

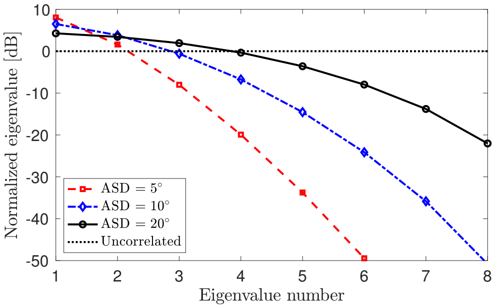
The (m,ℓ)th element of a generic spatial correlation matrix $\mathbf{R}$ can be computed as

$$[\mathbf{R}]_{m,\ell} = \beta \int\int e^{\mathrm{j}\pi(m-\ell)\sin(\bar{\varphi})\cos(\bar{\theta})} f(\bar{\varphi},\bar{\theta}) d\bar{\varphi} d\bar{\theta}$$

- valid for any PDF
- $f(\bar{\varphi},\bar{\theta}) = \dfrac{e^{-\frac{(\bar{\varphi}-\varphi)^2}{2\sigma_{\varphi}^2}} e^{-\frac{(\bar{\theta}-\theta)^2}{2\sigma_{\theta}^2}}}{2\pi\sigma_{\varphi}\sigma_{\theta}}$



Local Scattering Model (2/2)



- Spatial correlation is fundamental in Cellular Massive MIMO: large number of antennas.
- Since N is fairly small in Cell-free Massive MIMO, it cannot be utilized as efficiently.

- Eigenvalues are widely different when ASD is small;
- Eigenvalue spread decreases with an increasing ASD

- With $\sigma_\varphi = \sigma_\theta = 5^\circ$, the first four eigenvalues contribute to 99.99%;
- Correlation matrix is nearly rank-deficient;
- All channel realizations are linear combinations of the corresponding four eigenvectors.

Large-scale Fading Model

3GPP Urban Microcell model for the 2 GHz band

- APs are deployed in urban environments to serve a dense population of UEs.
- APs are deployed ten meters above the plane where the UEs are located.

The large scale fading coefficient (channel gain) in dB is:

$$\beta_{kl}\,[\mathrm{dB}] = -30.5 - 36.7\log_{10}\left(\frac{d_{kl}}{1\,\mathrm{m}}\right) + \underbrace{F_{kl}}_{\text{Shadow fading}}$$

where d_{kl} is the 3D distance between AP l and UE k, and $F_{kl} \sim \mathcal{N}(0, 4^2)$ with

$$\mathbb{E}\{F_{kl}F_{ij}\} = \begin{cases} 4^2 2^{-\delta_{ki}/9\,\mathrm{m}} & l = j \\ 0 & l \neq j \end{cases}$$

where δ_{ki} is the distance between UE k and UE i.

- Correlated shadowing if UEs closely located;
- Independent shadowing from different APs.

Channel Estimation

Preliminaries

To perform coherent processing (between AP antennas and across cooperating APs):

- AP l needs estimates of the channel vector $\mathbf{h}_{kl}$ from UE k if $l \in \mathcal{M}_k$;
- Block fading model: $\{\mathbf{h}_{kl}\}$ needs to be estimated once per each coherence block.

According to the TDD protocol:

- τ_p samples are reserved for uplink pilot signaling in each coherence block;
- each UE can transmit a pilot sequence that spans these τ_p samples;
- each AP can use the received signals to estimate the channel.

Ideally, orthogonal pilots should be used for all UEs (no interference)

- Only a set of at most τ_p mutually orthogonal sequences can be found!
- Since $\tau_p \leq \tau_c$, we cannot assign mutually orthogonal pilots to $K \gg \tau_c$ UEs.

Uplink Pilot Transmission

- Orthogonal signals: $\phi_1, \ldots, \phi_{\tau_p} \in \mathbb{C}^{\tau_p}$ with $\|\phi_t\|^2 = \tau_p$
- Pilot assigned $t_k \in \{1, \ldots, \tau_p\}$ to UE k
- UEs sharing pilot with UE k: $\mathcal{P}_k \subset \{1, \ldots, K\}$

Received uplink signal when UE k transmits (remove interference from UEs using orthogonal pilots):

$$\begin{aligned}
\mathbf{y}_{t_k l}^{\text{pilot}} &= \sum_{i=1}^{K} \frac{\sqrt{\eta_i}}{\sqrt{\tau_p}} \mathbf{h}_{il} \boldsymbol{\phi}_{t_i}^{\mathrm{T}} \boldsymbol{\phi}_{t_k}^{*} + \frac{1}{\sqrt{\tau_p}} \mathbf{N}_l \boldsymbol{\phi}_{t_k}^{*} \\
&= \underbrace{\sqrt{\eta_k \tau_p} \mathbf{h}_{kl}}_{\text{Desired part}} + \underbrace{\sum_{i \in \mathcal{P}_k \setminus \{k\}} \sqrt{\eta_i \tau_p} \mathbf{h}_{il}}_{\text{Interference}} + \underbrace{\mathbf{n}_{t_k l}}_{\text{Noise}}
\end{aligned}$$

MMSE Channel Estimation

Corollary

The MMSE estimate of $\mathbf{h}_{kl}$ based on $\mathbf{y}_{t_k l}^{\text{pilot}}$ is

$$\widehat{\mathbf{h}}_{kl} = \sqrt{\eta_k \tau_p} \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{y}_{t_k l}^{\text{pilot}}$$

where

$$\boldsymbol{\Psi}_{t_k l} = \mathbb{E}\left\{ \mathbf{y}_{t_k l}^{\text{pilot}} (\mathbf{y}_{t_k l}^{\text{pilot}})^{\mathrm{H}} \right\} = \underbrace{\sum_{i \in \mathcal{P}_k} \eta_i \tau_p \mathbf{R}_{il}}_{\text{Pilot-sharing UEs}} + \sigma_{\text{UL}}^2 \mathbf{I}_N .$$

The estimate $\widehat{\mathbf{h}}_{kl}$ and estimation error $\widetilde{\mathbf{h}}_{kl} = \mathbf{h}_{kl} - \widehat{\mathbf{h}}_{kl}$ are independent:

$$\widehat{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}, \eta_k \tau_p \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl} \right) \quad \widetilde{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{C}_{kl})$$

with error correlation matrix $\mathbf{C}_{kl} = \mathbb{E}\{\widetilde{\mathbf{h}}_{kl} \widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} = \mathbf{R}_{kl} - \eta_k \tau_p \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl}$

Alternative Channel Estimation Schemes

MMSE is optimal when having full statistical knowledge

- Note that the computational complexity is proportional to N^2.

Alternative channel estimators (LS, EW-MMSE) provide larger MSEs

- Useful to limit computational complexity or deal with incomplete statistics.
- Of interest in Cellular Massive MIMO systems, where N is large.

Not an issue in Cell-free Massive MIMO where N is small!

Pilot Contamination

The "pilot interference" from pilot-sharing UEs is known as *pilot contamination*

- it does not only reduce the estimation quality but...
- also makes the channel estimates of pilot-sharing UEs correlated

$$\widehat{\mathbf{h}}_{il} = \sqrt{\frac{\eta_i}{\eta_k}} \mathbf{R}_{il} (\mathbf{R}_{kl})^{-1} \widehat{\mathbf{h}}_{kl}, \quad i \in \mathcal{P}_k$$

although $\{\mathbf{h}_{kl}, \mathbf{h}_{il}\}$ are statistically independent (i.e., $\mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{il}^{\mathrm{H}}\} = \mathbf{0}_N$, for $i \neq k$).

With no spatial correlation (i.e., $\mathbf{R}_{kl} = \beta_{kl}\mathbf{I}_N$ and $\mathbf{R}_{il} = \beta_{il}\mathbf{I}_N$), channels will be exactly the same except for a scaling factor, i.e., fully correlated.

As in Cellular Massive MIMO, the induced correlation makes it harder to mitigate interference between UEs that use the same pilot in UL and DL.

MMSE Estimation of the Collective Channel

Full knowledge of $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_k)$ is needed for coherent processing at the CPU

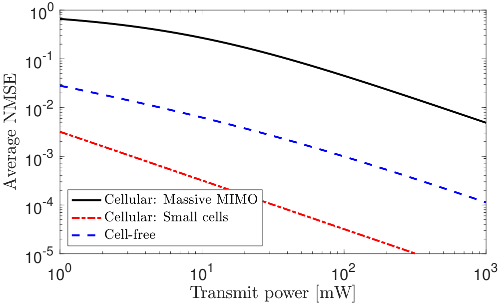
- ...but only the APs in $\mathcal{M}_k \subset \{1, \dots, L\}$ compute channel estimates and/or send pilot signals to CPU.

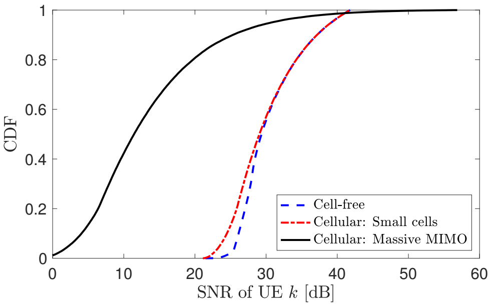
Only the following partial channel estimate is known in the cell-free network:

$$\mathbf{D}_k \widehat{\mathbf{h}}_k \triangleq \begin{bmatrix} \mathbf{D}_{k1} \widehat{\mathbf{h}}_{k1} \\ \vdots \\ \mathbf{D}_{kL} \widehat{\mathbf{h}}_{kL} \end{bmatrix} \sim \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}, \eta_k \tau_p \mathbf{D}_k \mathbf{R}_k \boldsymbol{\Psi}_{t_k}^{-1} \mathbf{R}_k \mathbf{D}_k \right) \tag{1}$$

where $\mathbf{D}_k = \mathrm{diag}(\mathbf{D}_{k1}, \dots, \mathbf{D}_{kL})$ with $\mathbf{D}_{kl} = \mathbf{I}_N$ for APs that serve UE k and zero otherwise.

Impact of the Cell-Free Architecture ($L = 64$ single-antenna APs, $K = 1$)

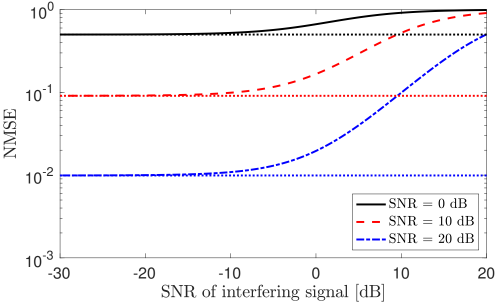


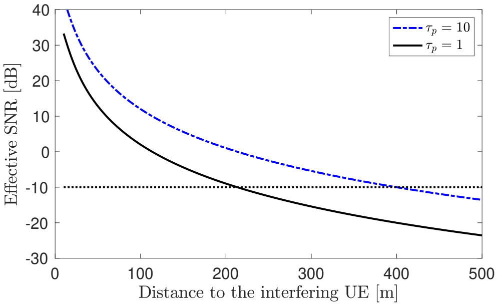


- Have many more APs is better;
- Small-cells achieve better estimation accuracy: AP with the lowest NMSE is always chosen.

- With MR combining, higher SE is however achieved with cell-free due to coherent processing;
- Larger benefit with multiple UEs.

Pilot Contamination Impact ($L = 64$ single-antenna APs, $K = 2$ pilot-sharing UEs)





- Negligible when interference is 10 dB weaker than noise.
- Sensitive to UEs that are close to serving APs;
- UEs that an AP serves should use different pilots.

- Larger if close to the AP, but above -10 dB for hundred meters.
- Increasing pilot length helps, but less pilot reuse.
- Pilots must be assigned properly.

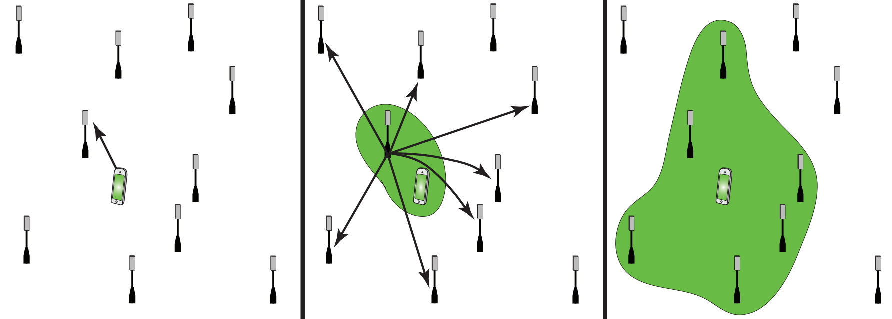
Scalable User-Centric Cooperation Clusters

One access-based algorithm:

Step 1:
UE appoints Master AP

Step 2:
Pilot assignment
invitation to other APs

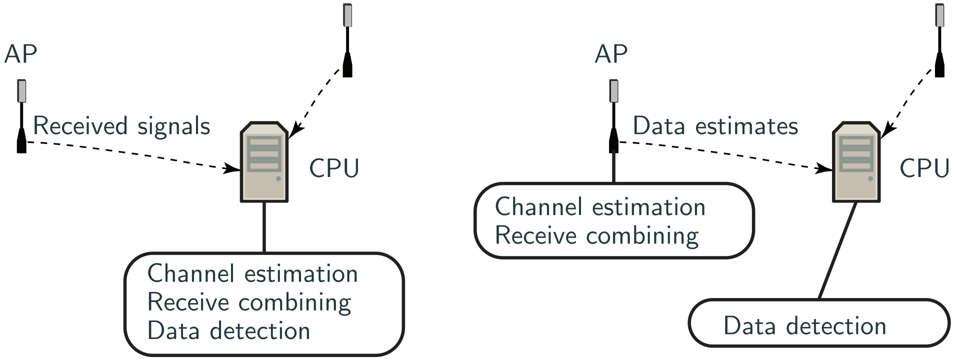
Step 3:
Formation of UE cluster



[BS20] E. Björnson, L. Sanguinetti, “Scalable Cell-Free Massive MIMO Systems,” IEEE Trans. Communications, 2020.

Uplink Operation

Centralized vs. Decentralized



(a) Centralized operation

(b) Distributed operation

Figure 1: Uplink signal processing tasks can be divided between APs and CPU. In the centralized operation, the channel estimation, receive combining, and data detection are done at the CPU. In the distributed operation, everything, except the data detection, is done at APs.

Centralized Uplink Operation

Most advanced uplink implementation of Cell-free Massive MIMO is

- a fully centralized operation;
- APs only act as relays that forward their signals to the CPU for processing.

- Each AP l sends τ_p received pilot signals $\{\mathbf{y}_{tl}^{\text{pilot}} : t = 1, \ldots, \tau_p\}$ to CPU;
- Each AP l sends received uplink data signal $\mathbf{y}_l^{\text{ul}}$ to CPU;
- CPU uses all the signals to perform channel estimation and data detection.

Highest level of AP cooperation.

Uplink System Model with Centralized Operation

The estimate $\widehat{s}_k$ of the signal s_k transmitted by UE k is

$$\widehat{s}_k = \sum_{l=1}^{L} \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{y}_l^{\mathrm{ul}} = \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{y}^{\mathrm{ul}}$$

where $\mathbf{v}_k = [\mathbf{v}_{k1}^{\mathrm{T}}, \ldots, \mathbf{v}_{kL}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{LN}$ is the centralized combining vector and

$$\mathbf{y}^{\mathrm{ul}} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{ul}} \\ \vdots \\ \mathbf{y}_L^{\mathrm{ul}} \end{bmatrix} = \sum_{i=1}^{K} \mathbf{h}_i \sqrt{p_i} s_i + \mathbf{n} \tag{2}$$

with $\mathbf{n} = [\mathbf{n}_1^{\mathrm{T}}, \ldots, \mathbf{n}_L^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{LN}$ being the collective noise vector.

Since $\mathbf{D}_{kl} = \mathbf{0}_N$ implies $\mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} = \mathbf{0}_N$, only AP signals with $\mathbf{D}_{kl} \neq \mathbf{0}_N$ are used.

Uplink System Model with Centralized Operation

If one treats the CPU as a receiver equipped with LN antennas:

> Mathematically equivalent to the signal model of an uplink single-cell Massive MIMO system with correlated fading.

However, there are some key differences :

- Multiple UEs served by CPU use the same pilot (normally avoided in single-cell systems);
- Antennas are distributed at different geographical locations. Hence, $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_k)$ where

$$\mathbf{R}_k = \mathrm{diag}(\mathbf{R}_{k1}, \ldots, \mathbf{R}_{kL}) \in \mathbb{C}^{LN \times LN}$$

  has a block-diagonal structure (normally not the case in single-cell systems).

Spectral Efficiency with Centralized Operation

From Massive MIMO literature...

Theorem

An achievable SE of UE k in the centralized operation is

$$\mathsf{SE}_k^{(\mathrm{ul,c})} = \frac{\tau_u}{\tau_c}\mathbb{E}\left\{\log_2\left(1 + \mathsf{SINR}_k^{(\mathrm{ul,c})}\right)\right\} \quad \mathrm{bit/s/Hz} \tag{3}$$

where the expectation is with respect to the channel estimates and

$$\mathsf{SINR}_k^{(\mathrm{ul,c})} = \frac{p_k \left|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k\right|^2}{\sum\limits_{\substack{i=1 \\ i\neq k}}^{K} p_i \left|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i\right|^2 + \mathbf{v}_k^{\mathrm{H}}\mathbf{Z}_k\mathbf{v}_k + \sigma_{\mathrm{UL}}^2\|\mathbf{D}_k\mathbf{v}_k\|^2} \tag{4}$$

with $\mathbf{Z}_k = \sum_{i=1}^{K} p_i\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k$.

Optimal Receive Combining

Corollary

The instantaneous SINR for UE k is maximized by the MMSE combining vector

$$\mathbf{v}_k^{\mathrm{MMSE}} = p_k \left(\sum_{i=1}^{K} p_i \mathbf{D}_k (\widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{C}_i) \mathbf{D}_k + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k$$

which leads to the maximum value

$$\mathsf{SINR}_k^{(\mathrm{ul,c})} = p_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \left(\sum_{i=1, i \neq k}^{K} p_i \mathbf{D}_k \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} \mathbf{D}_k + \mathbf{Z}_k + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k .$$

Not scalable as K grows!

Scalable Combining Schemes: MR combining

Let's develop centralized receive combining schemes that are scalable

- the computational complexity per UE is independent of K.

The simplest solution is MR combining:

$$\mathbf{v}_k^{\mathrm{MR}} = \mathbf{D}_k \widehat{\mathbf{h}}_k$$

- Key benefit: combining follows directly from the channel estimates;
- No additional computations are needed;
- Performance can be poor with interference;
- High degree of favorable propagation cannot be guaranteed between all UEs

Scalable Combining Schemes: P-MMSE combining

- Interference mainly generated by UEs located in the neighborhood of UE k.
- Consider only UEs that are served by partially the same APs as UE k:

$$\mathcal{S}_k = \{i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN}\}.$$

By utilizing $\mathcal{S}_k$, an alternative *partial MMSE (P-MMSE) combining* scheme is:

$$\mathbf{v}_k^{\mathrm{P-MMSE}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} \mathbf{D}_k + \mathbf{Z}_{\mathcal{S}_k} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k$$

with $\mathbf{Z}_{\mathcal{S}_k} = \sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k$.

Not optimal...but scalable approximation of optimal MMSE combining.

Fronthaul Signal Load with Centralized Operation

In each coherence block, AP l needs to send

- $\tau_p N$ complex scalars representing the pilot signals $\{\mathbf{y}_{tl}^{\text{pilot}} : t = 1, \ldots, \tau_p\}$;
- $\tau_u N$ complex scalars representing the received data signal $\mathbf{y}_l^{\text{ul}}$.

Does not grow with K...fronthaul signaling is scalable.

Knowledge of $\{\mathbf{R}_{kl} : \forall k, l\}$ is required for MMSE processing:

- It can be acquired directly at the CPU by using pilot signaling methods [SBH20];
- The situation will be different in the distributed implementation.

[SBH20] Sanguinetti, Björnson, and Hoydis. "Towards Massive MIMO 2.0". TCOM 2020.

Distributed Uplink Operation

Instead of sending pilot and received data signals to CPU, each AP l can:

- Compute *local* channel estimates $\{\widehat{\mathbf{h}}_{il} : i \in \mathcal{D}_l\}$;
- For each UE k, compute its *local* estimate of s_k given by $\widehat{s}_{kl} = \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{y}_l^{\mathrm{ul}}$.

Local data estimates are *linearly* combined at the CPU for final detection:

$$\widehat{s}_k = \sum_{l=1}^{L} a_{kl}^{\star} \widehat{s}_{kl} = \sum_{l=1}^{L} a_{kl}^{\star} \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{y}_l^{\mathrm{ul}}.$$

To limit fronthaul signaling, APs send only $\widehat{s}_{kl}$ (not channel estimates)

Weights $\{a_{kl}\}$ must be deterministic functions of channel statistics, designed on the basis of SNR, interference situation and local receive combining.

Spectral Efficiency with Distributed Uplink Operation (1/2)

Rewrite the global estimate of s_k as

$$\widehat{s}_k = \left(\sum_{l=1}^{L} a_{kl}^{\star} \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{h}_{kl} \right) \sqrt{p_k} s_k + \sum_{i=1, i \neq k}^{K} \left(\sum_{l=1}^{L} a_{kl}^{\star} \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{h}_{il} \right) \sqrt{p_i} s_i + n_k'$$

$$= \mathbf{a}_k^{\mathrm{H}} \mathbf{g}_{kk} \sqrt{p_k} s_k + \sum_{i=1, i \neq k}^{K} \mathbf{a}_k^{\mathrm{H}} \mathbf{g}_{ki} \sqrt{p_i} s_i + n_k'$$

where $\mathbf{a}_k = [a_{k1} \ \ldots \ a_{kL}]^{\mathrm{T}} \in \mathbb{C}^L$ is the LSFD weight vector of UE k and

$$\mathbf{g}_{ki} = \left[\mathbf{v}_{k1}^{\mathrm{H}} \mathbf{D}_{k1} \mathbf{h}_{i1}, \ldots, \mathbf{v}_{kL}^{\mathrm{H}} \mathbf{D}_{kL} \mathbf{h}_{iL} \right]^{\mathrm{T}}$$

collects the receive-combined channels between UE i and all APs that serve UE k.

Notice that $\mathbf{g}_{ki}$ changes in every coherence block, while $\mathbf{a}_k$ is deterministic.

Spectral Efficiency with Distributed Uplink Operation (2/2)

By using similar tools from Cellular Massive MIMO...

Theorem

An achievable SE of UE k in the distributed operation is

$$\mathsf{SE}_k^{(\mathrm{ul,d})} = \frac{\tau_u}{\tau_c} \log_2 \left(1 + \mathsf{SINR}_k^{(\mathrm{ul,d})} \right) \quad \mathrm{bit/s/Hz}$$

with the effective SINR given by

$$\mathsf{SINR}_k^{(\mathrm{ul,d})} = \frac{p_k \left| \mathbf{a}_k^{\mathrm{H}} \mathbb{E}\{\mathbf{g}_{kk}\} \right|^2}{\mathbf{a}_k^{\mathrm{H}} \left(\sum\limits_{i=1}^{K} p_i \mathbb{E}\{\mathbf{g}_{ki} \mathbf{g}_{ki}^{\mathrm{H}}\} - p_k \mathbb{E}\{\mathbf{g}_{kk}\} \mathbb{E}\{\mathbf{g}_{kk}^{\mathrm{H}}\} + \mathbf{F}_k \right) \mathbf{a}_k}$$

where $\mathbf{F}_k = \sigma_{\mathrm{UL}}^2 \mathrm{diag} \left(\mathbb{E}\left\{ \|\mathbf{D}_{k1}\mathbf{v}_{k1}\|^2 \right\}, \ldots, \mathbb{E}\left\{ \|\mathbf{D}_{kL}\mathbf{v}_{kL}\|^2 \right\} \right) \in \mathbb{R}^{L \times L}$.

Optimal Large Scale Fading Decoding (1/2)

Corollary

The effective SINR for UE k is maximized by

$$\mathbf{a}_k^{\text{opt}} = p_k \left(\sum_{i=1}^{K} p_i \mathbb{E}\left\{ \mathbf{g}_{ki} \mathbf{g}_{ki}^{\text{H}} \right\} + \mathbf{F}_k + \widetilde{\mathbf{D}}_k \right)^{-1} \mathbb{E}\left\{ \mathbf{g}_{kk} \right\}$$

where $\widetilde{\mathbf{D}}_k \in \mathbb{R}^{L \times L}$ is diagonal with $[\widetilde{\mathbf{D}}_k]_{l,l}$ being one if $l \notin \mathcal{M}_k$ and zero otherwise.

This leads to the maximum value

$$p_k \mathbb{E}\left\{ \mathbf{g}_{kk}^{\text{H}} \right\} \left(\sum_{i=1}^{K} p_i \mathbb{E}\left\{ \mathbf{g}_{ki} \mathbf{g}_{ki}^{\text{H}} \right\} - p_k \mathbb{E}\left\{ \mathbf{g}_{kk} \right\} \mathbb{E}\left\{ \mathbf{g}_{kk}^{\text{H}} \right\} + \mathbf{F}_k + \widetilde{\mathbf{D}}_k \right)^{-1} \mathbb{E}\left\{ \mathbf{g}_{kk} \right\}.$$

Optimal Large Scale Fading Decoding (2/2)

The optimal $\mathbf{a}_k^{\text{opt}}$ has a structure resembling MMSE combining:

- Computed to maximize a generalized Rayleigh quotient

Its evaluation requires knowledge of:

- the L-dimensional *statistical* vector $\mathbb{E}\{\mathbf{g}_{kk}\}$;
- the L diagonal elements of the real-valued *statistical* matrix $\mathbf{F}_k$;
- the $L \times L$ Hermitian complex matrix $\mathbb{E}\{\mathbf{g}_{ki}\mathbf{g}_{ki}^{\text{H}}\}$, for $i = 1, \dots, K$.

APs need to send statistical parameters to the CPU through the fronthaul links.

Number of parameters grows with K, thus making the optimal LSFD unscalable!

Scalable Local Combiners

The simplest solution is MR combining $\mathbf{v}_{kl}^{\mathrm{MR}} = \mathbf{D}_{kl}\widehat{\mathbf{h}}_{kl}$

- Key benefit: combining follows directly from the channel estimates;
- Another benefit: expectations can be computed analytically.

Inspired by the P-MMSE combining, an alternative is local P-MMSE (LP-MMSE)

$$\mathbf{v}_{kl}^{\mathrm{LP-MMSE}} = p_k \left(\sum_{i \in \mathcal{D}_l} p_i \left(\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{C}_{il} \right) + \sigma_{\mathrm{UL}}^2 \mathbf{I}_N \right)^{-1} \mathbf{D}_{kl}\widehat{\mathbf{h}}_{kl}.$$

- Complexity independent of K: scalable solution;
- Lower complexity than P-MMSE: inverse of an $N \times N$.

Scalable Large Scale Fading Decoding (1/2)

The optimal LSFD vector is not scalable

- All UEs in the network affect the interference levels at all APs;
- A scalable way needs to limit how many interfering UEs are considered.

The simplest solution is to give all APs equal importance by setting $\mathbf{a}_k = [1 \; \ldots \; 1]^{\mathrm{T}}$:

$$\widehat{s}_k = \sum\nolimits_{l \in \mathcal{M}_k} \widehat{s}_{kl}.$$

- Equal importance is given to all APs, even if some APs might be subject to high interference or low SNR.
- It can even happen that an AP reduces rather than increases the SE due to the suboptimal fusing.

Scalable Large Scale Fading Decoding (2/2)

The CPU should use some side-information of how accurate local data estimates are.

- This is what the optimal LSFD vector does.

Similar to P-MMSE, take only the UEs that are served by partially the same APs:

- If the AP selection is done properly, these UEs cause majority of interference.

Approximate the optimal LSFD vector as

$$\mathbf{a}_k^{\text{n-opt}} = p_k \Big(\sum_{i \in \mathcal{S}_k} p_i \mathbb{E}\left\{ \mathbf{g}_{ki} \mathbf{g}_{ki}^{\mathrm{H}} \right\} + \mathbf{F}_k + \widetilde{\mathbf{D}}_k \Big)^{-1} \mathbb{E}\left\{ \mathbf{g}_{kk} \right\}$$

which we call the *nearly optimal (n-opt) LSFD vector*.

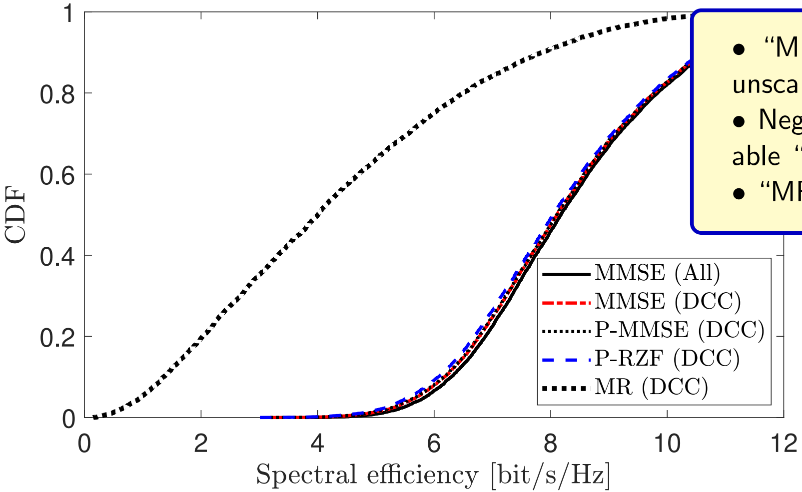
Number of statistical parameters independent of K. Fronthaul signaling scalable!

Running Example

Parameter	Value
Network area	$1\,\text{km} \times 1\,\text{km}$
Network layout	Random deployment
Number of APs	400 or 100
Number of antennas per AP	1 or 4
Number of total antennas	$M = LN = 400$
Bandwidth	$B = 20\,\text{MHz}$
Receiver noise power	$\sigma^2_{\mathrm{UL}} = -94\,\text{dBm}$
Maximum uplink transmit power	$100\,\text{mW}$
Maximum downlink transmit power	$200\,\text{mW}$
Samples per coherence block	$\tau_c = 200$
Channel gain at 1 km	$\Upsilon = -140.6\,\text{dB}$
Pathloss exponent	$\alpha = 3.67$
Height difference between AP and UE	$10\,\text{m}$
Standard deviation of shadow fading	$\sigma_{\mathrm{sf}} = 4$

Table 1: Key parameters of running example.

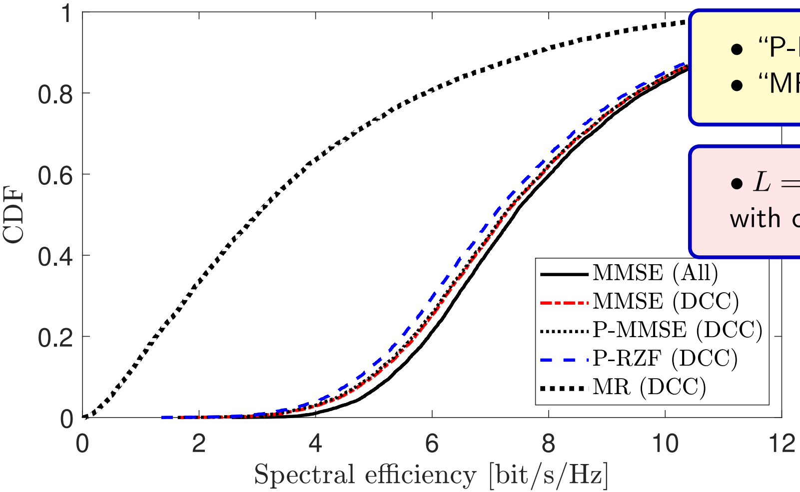
Evaluation with Centralized Operation (1/2)



- "MMSE (All)" is the best, but unscalable;
- Negligible gap in using the scalable "P-MMSE";
- "MR (DCC)" is not competitive.

APs $L = 400$, antennas per AP $N = 1$

Evaluation with Centralized Operation (2/2)

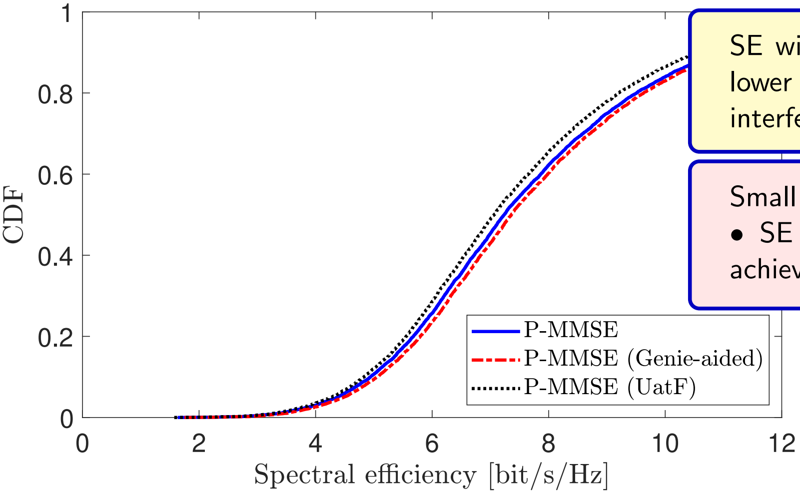


- "P-MMSE (DCC)" is still good;
- "MR (DCC)" not competitive.

- $L = 400, N = 1$ is more desirable with centralized operation.

APs $L = 100$, antennas per AP $N = 4$.

Tightness of the provided SE



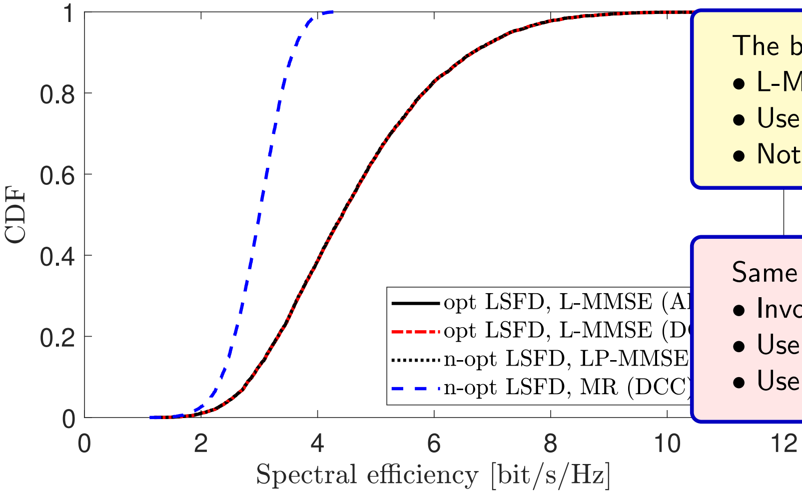
SE with centralized operation is a lower bound of the capacity where interference is treated as noise.

Small gap with genie-aided.

- SE representative of practically achievable performance.

APs $L = 100$, antennas per AP $N = 4$

Evaluation with Decentralized Operation (1/2)



The best solution is:

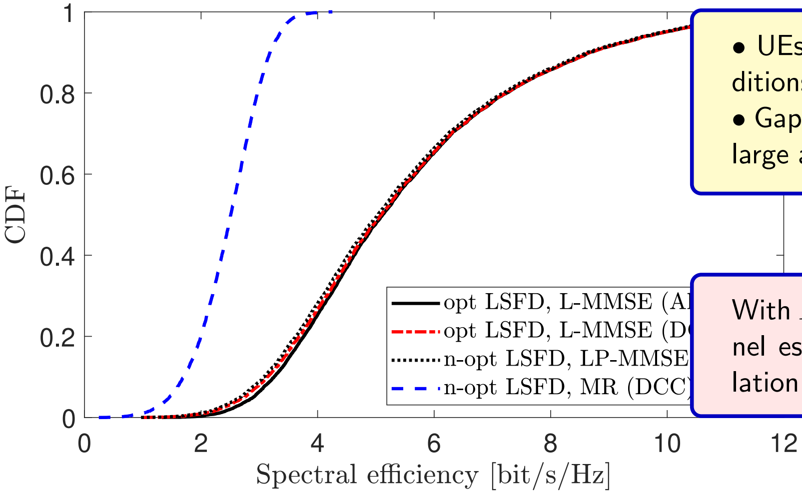
- L-MMSE combining at each AP;
- Use "opt-LSFD" at the CPU;
- Not scalable.

Same SE with three simplifications:

- Involve only a subset of APs;
- Use "n-opt-LSFD";
- Use LP-MMSE (not L-MMSE).

APs $L = 400$, antennas per AP $N = 1$

Evaluation with Decentralized Operation (2/2)

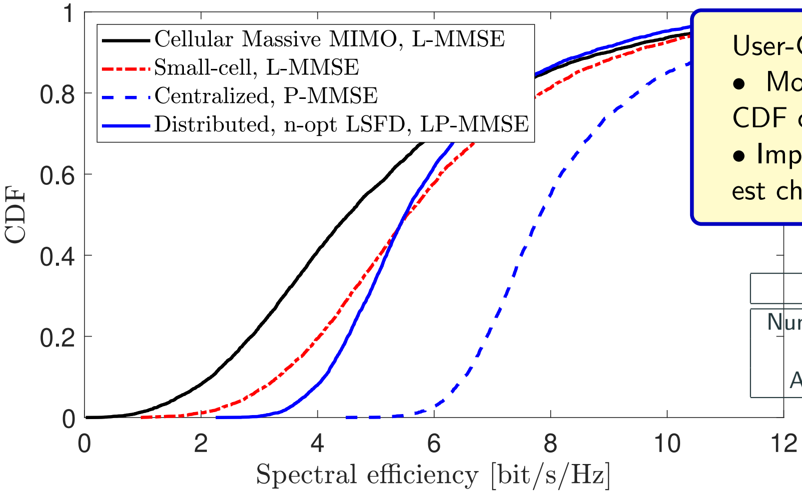


- UEs with the worst channel conditions experience lower SE;
- Gap between two scenarios not as large as with centralized operation.

With $L = 100, N = 4$, better channel estimates due to spatial correlation.

APs $L = 100$, antennas per AP $N = 4$

Comparison between Cell-free and Cellular Networks



User-Centric CF Massive MIMO

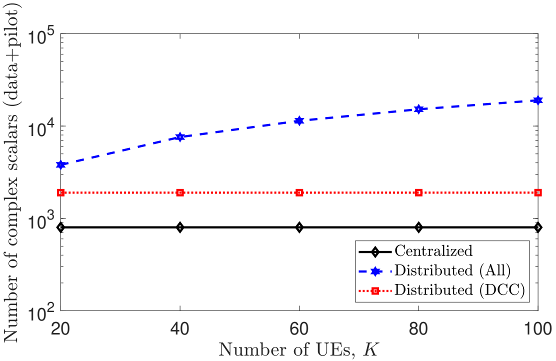
- More uniform service (steeper CDF curves);
- Improved SE for UEs with weakest channels.

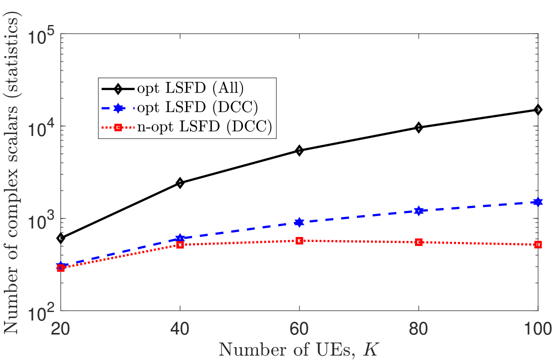
Parameter	Value
Number of small cells	100
Small cell area	$100\,\mathrm{m} \times 100\,\mathrm{m}$
Antennas per AP	4

APs $L = 100$, antennas per AP $N = 4$

Scalability of Fronthaul and Computations

Fronthaul Signaling

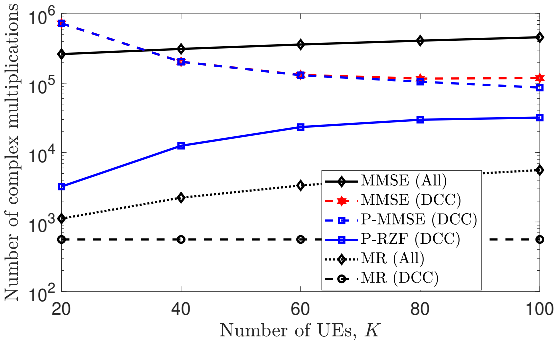


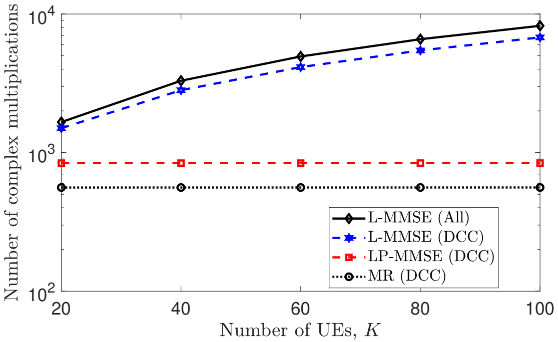


- With centralized operation, scalability is not an issue (it is constant!).
- With distributed operation, it grows with K only if all APs serve all UEs.

- No statistical parameters with centralized operation;
- Using optimal LSFD and serving all UEs using all APs is NOT scalable;
- n-opt-LSFD using DCC is scalable!.

Computational Complexity





- Much higher complexity when all APs serve all UEs;
- Due to shared matrix inversion, the complexity of MMSE-type combiners decreases with K.

- Does not grow with K for the scalable combining schemes;
- For the other schemes, it grows with K since they are unscalable.

Downlink Operation

Downlink Operation

The received downlink signal at UE k is

$$y_k^{\mathrm{dl}} = \sum_{l=1}^{L} \mathbf{h}_{kl}^{\mathrm{H}} \left(\sum_{i=1}^{K} \mathbf{D}_{il} \mathbf{w}_{il} \varsigma_i \right) + n_k$$

$$= \sum_{i=1}^{K} \begin{bmatrix} \mathbf{h}_{k1} \\ \vdots \\ \mathbf{h}_{kL} \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} \mathbf{D}_{i1} \mathbf{w}_{i1} \\ \vdots \\ \mathbf{D}_{iL} \mathbf{w}_{iL} \end{bmatrix} \varsigma_i + n_k$$

$$= \sum_{i=1}^{K} \mathbf{h}_k^{\mathrm{H}} \mathbf{D}_i \mathbf{w}_i \varsigma_i + n_k$$

- $\varsigma_i \in \mathbb{C}$ is data signal
- $\mathbb{E}\{|\varsigma_i|^2\} = 1$
- $n_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{DL}}^2)$

$$\underbrace{\mathbf{D}_{il} \mathbf{w}_{il}}_{\text{Effective precoder}} = \begin{cases} \mathbf{w}_{il} & l \in \mathcal{M}_i \\ \mathbf{0} & l \notin \mathcal{M}_i \end{cases}$$

- The same system model for both centralized and distributed operation
- Selection of the collective precoding vector $\mathbf{D}_i \mathbf{w}_i$ for UE i differs

Centralized Downlink Operation

- Design of $\mathbf{D}_i \mathbf{w}_i$ at the CPU like in an LN-antenna single-cell massive MIMO **but**
- Considering the DCC, block correlated Rayleigh fading, and per-AP power constraints

$$\mathbb{E}\left\{\|\mathbf{x}_l\|^2\right\} \leq \rho_{\max} \quad l = 1, \ldots, L$$

where $\mathbf{x}_l = \sum_{i=1}^{K} \mathbf{D}_{il} \mathbf{w}_{il} \varsigma_i$ is the transmitted signal from AP l.

- An achievable SE of UE k is

$$\mathsf{SE}_k^{(\mathrm{dl,c})} = \frac{\tau_d}{\tau_c} \log_2\left(1 + \mathsf{SINR}_k^{(\mathrm{dl,c})}\right) \quad \mathrm{bit/s/Hz}$$

where $\mathsf{SINR}_k^{(\mathrm{dl,c})}$ is the effective SINR

$$\mathsf{SINR}_k^{(\mathrm{dl,c})} = \frac{\left|\mathbb{E}\left\{\mathbf{h}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{w}_k\right\}\right|^2}{\sum_{i=1}^{K} \mathbb{E}\left\{\left|\mathbf{h}_k^{\mathrm{H}} \mathbf{D}_i \mathbf{w}_i\right|^2\right\} - \left|\mathbb{E}\left\{\mathbf{h}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{w}_k\right\}\right|^2 + \sigma_{\mathrm{DL}}^2}$$

Duality between Uplink and Downlink

- The SE of UE k is affected by the transmit precoding vector $\mathbf{D}_i \mathbf{w}_i \ \forall i$.
- More complicated precoding selection than in the uplink receive combining

Uplink-Downlink Duality: Let $\mathbf{D}_i \mathbf{v}_i$ and p_i denote the combining vector and transmit power for UE i, respectively, used in the uplink. If the precoding vectors are selected as

$$\mathbf{w}_i = \sqrt{\rho_i} \frac{\mathbf{v}_i}{\sqrt{\mathbb{E}\left\{\|\mathbf{D}_i \mathbf{v}_i\|^2\right\}}}$$

then there exists a downlink power allocation policy $\{\rho_i : i = 1, \ldots, K\}$ with $\sum_{i=1}^{K} \rho_i / \sigma_{\mathrm{DL}}^2 = \sum_{i=1}^{K} p_i / \sigma_{\mathrm{UL}}^2$ for which the same effective SINRs are attained.

Distributed Downlink Operation

- An achievable SE of UE k is

$$\mathsf{SE}_k^{(\mathrm{dl,d})} = \frac{\tau_d}{\tau_c} \log_2\left(1 + \mathsf{SINR}_k^{(\mathrm{dl,d})}\right) \quad \text{bit/s/Hz}$$

with effective SINR given by

$$\mathsf{SINR}_k^{(\mathrm{dl,d})} = \frac{\left|\sum\limits_{l=1}^{L} \mathbb{E}\left\{\mathbf{h}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{w}_{kl}\right\}\right|^2}{\sum\limits_{i=1}^{K} \mathbb{E}\left\{\left|\sum\limits_{l=1}^{L} \mathbf{h}_{kl}^{\mathrm{H}} \mathbf{D}_{il} \mathbf{w}_{il}\right|^2\right\} - \left|\sum\limits_{l=1}^{L} \mathbb{E}\left\{\mathbf{h}_{kl}^{\mathrm{H}} \mathbf{D}_{kl} \mathbf{w}_{kl}\right\}\right|^2 + \sigma_{\mathrm{DL}}^2}$$

- The same SE as in the centralized downlink
- Each AP l designs the local transmit precoders $\mathbf{D}_{il}\mathbf{w}_{il}$ using local channel estimates
- Use duality for local transmit precoder design

Fronthaul Signaling Load in Downlink

Centralized Operation

- The CPU sends $\mathbf{x}_l \in \mathbb{C}^N$ to AP l, for $l = 1, \ldots, L$.
- Number of complex scalars to be shared over the fronthaul is

Scheme	Each coherence block
Any precoding	$\tau_d N L$

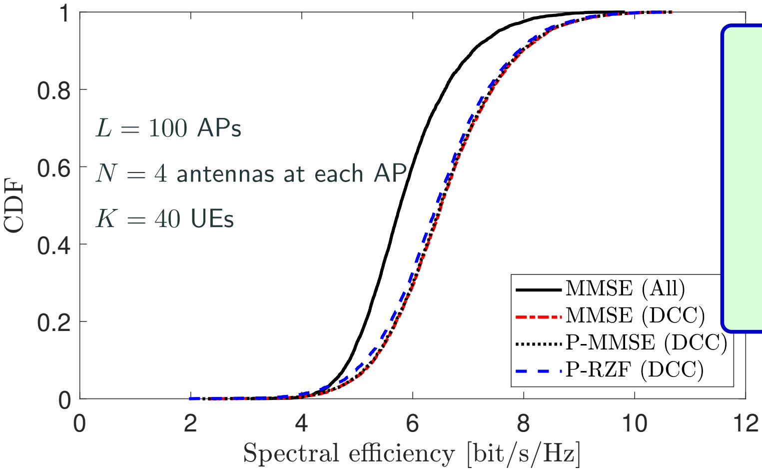
- The APs send the received pilot signals, so the CPU can compute the long-term statistical matrices.

Distributed Operation

- The CPU sends the data signals $\{\varsigma_k : k \in \mathcal{D}_l\}$ to each AP l, for $l = 1, \ldots, L$.
- Number of complex scalars to be shared over the fronthaul is

Scheme	Each coherence block
Any precoding	$\tau_d \sum_{l=1}^{L} \lvert\mathcal{D}_l\rvert$

SE Performance: Downlink Centralized Operation

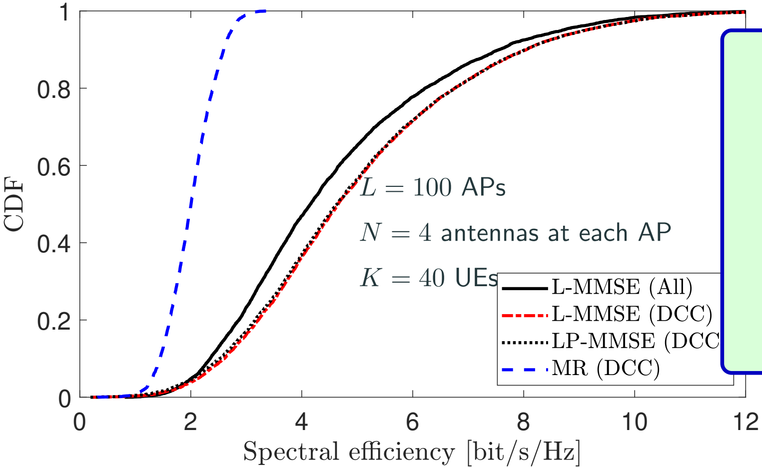


Recall the centralized transmit precoding vectors:

$$\mathbf{w}_k = \sqrt{\rho_k} \frac{\bar{\mathbf{w}}_k}{\sqrt{\mathbb{E}\left\{\|\bar{\mathbf{w}}_k\|^2\right\}}}$$

A new fractional-type scalable power allocation

SE Performance: Downlink Distributed Operation



Recall the local transmit precoding vectors:

$$\mathbf{w}_{kl} = \sqrt{\rho_{kl}} \frac{\bar{\mathbf{w}}_{kl}}{\sqrt{\mathbb{E}\left\{ \|\bar{\mathbf{w}}_{kl}\|^2 \right\}}}$$

Fractional power allocation [IFL+19]

[IFL+19] G. Interdonato, P. Frenger and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," IEEE ICC 2019.

Spatial Resource Allocation

Transmit Power Optimization

- Uplink, UE transmit powers: p_k, for $k = 1, \ldots, K$
- Centralized downlink, transmit powers ρ_k, for $k = 1, \ldots, K$
- Distributed downlink, AP transmit powers ρ_{kl}, for $k = 1, \ldots, K$ and $l = 1, \ldots, L$

1) Heuristically selected

2) Optimized: Maximize a network-wide utility function, e.g.,

- max-min fairness
- sum SE

- The same power coefficients for a given long-term channel statistics
- No need to change power control on a coherence block basis
- Dependent on the channel statistics based on choice of combining/precoding

Uplink Power Optimization

Power control: controlling how much each UE cuts down on its power from the maximum $p_{\max}$:

$$0 \leq p_k \leq p_{\max}$$

- $\mathbf{p} = [p_1 \ \ldots \ p_K]^{\mathrm{T}}$
- The effective SINR for UE k has the generic form (in both centralized and distributed operation)

$$\mathrm{SINR}_k(\mathbf{p}) = \frac{b_k p_k}{\mathbf{c}_k^{\mathrm{T}} \mathbf{p} + \sigma_k^2}$$

- $b_k \geq 0$: the average channel gain of the desired signal
- $\mathbf{c}_k = [c_{k1} \ \ldots \ c_{kK}]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^K$: the average channel gains for the interfering signals
- $\sigma_k^2 \geq 0$: the effective noise variance

Uplink Max-Min Fair SE Maximization

Maximize the minimum SE under UE transmit power constraints:

$$\begin{aligned} \underset{\mathbf{p} \geq \mathbf{0}_K}{\text{maximize}} \quad & \min_{k \in \{1,\ldots,K\}} \frac{b_k p_k}{\mathbf{c}_k^{\mathrm{T}} \mathbf{p} + \sigma_k^2} \\ \text{subject to} \quad & p_k \leq p_{\max}, \quad k = 1, \ldots, K \end{aligned}$$

- Solved optimally by a bisection search: sequence of linear feasibility programs
- All UEs will have equal SE at the optimum and at least one UE will use the maximum power $p_{\max}$

Fixed-point algorithm that converges to the optimal solution

- $p_k \leftarrow \frac{p_k}{\mathrm{SINR}_k(\mathbf{p})}, \ k = 1, \ldots, K$
- $\mathbf{p} \leftarrow \frac{p_{\max}}{\max\limits_{k \in \{1,\ldots,K\}} p_k} \mathbf{p}$

It converges quickly, but the complexity grows with K, which makes it not scalable!

Uplink Sum SE Maximization

Maximize the sum SE under UE transmit power constraints:

$$\begin{aligned} \underset{\mathbf{p} \geq \mathbf{0}_K}{\text{maximize}} \quad & \sum_{k=1}^{K} \log_2 \left(1 + \frac{b_k p_k}{\mathbf{c}_k^{\mathrm{T}} \mathbf{p} + \sigma_k^2} \right) \\ \text{subject to} \quad & p_k \leq p_{\max}, \quad k = 1, \ldots, K \end{aligned}$$

- A local optimum by a block coordinate descent algorithm based on weighted MMSE (WMMSE) reformulation [SRL+11].

Block coordinate descent algorithm based on WMMSE reformulation

- $u_k \leftarrow \frac{\sqrt{b_k p_k}}{b_k p_k + \mathbf{c}_k^{\mathrm{T}} \mathbf{p} + \sigma_k^2}$ • $d_k \leftarrow \frac{1}{u_k^2 \left(b_k p_k + \mathbf{c}_k^{\mathrm{T}} \mathbf{p} + \sigma_k^2 \right) - 2 u_k \sqrt{b_k p_k} + 1}$
- $p_k \leftarrow \min \left(p_{\max}, b_k d_k^2 u_k^2 \Big/ \left(d_k u_k^2 b_k + \sum_{i=1}^{K} d_i u_i^2 c_{ik} \right)^2 \right)$

Closed-form updates but not scalable!

[SRL+11] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," IEEE TSP, 2011.

Fractional Uplink Power Control

Fractional power control: classical heuristic in the uplink of multi-user systems

- Balance between SE of the cell-center and the cell-edge UEs
- Different situation in cell-free networks; each UE has multiple serving APs

- A fractional power control for cell-free networks [NL+19]:

$$p_k = p_{\max} \frac{\left(\sum\limits_{l \in \mathcal{M}_k} \beta_{kl} \right)^{\upsilon}}{\max\limits_{i \in \{1,\ldots,K\}} \left(\sum\limits_{l \in \mathcal{M}_i} \beta_{il} \right)^{\upsilon}}$$

- $\upsilon = 0 \rightarrow p_k = p_{\max}$
- $\upsilon = -1 \rightarrow$ fully compensating for the variations in the total channel gain
- $\upsilon \in [-1, 0]$ in [NL19]
- $\upsilon > 0 \rightarrow$ more power used by the UEs with good channel conditions

[NL19] R. Nikbakht and A. Lozano, "Uplink fractional power control for cell-free wireless networks,", ICC 2019.

Scalable Uplink Power Control

For obtaining p_k, the maximum of K terms is computed

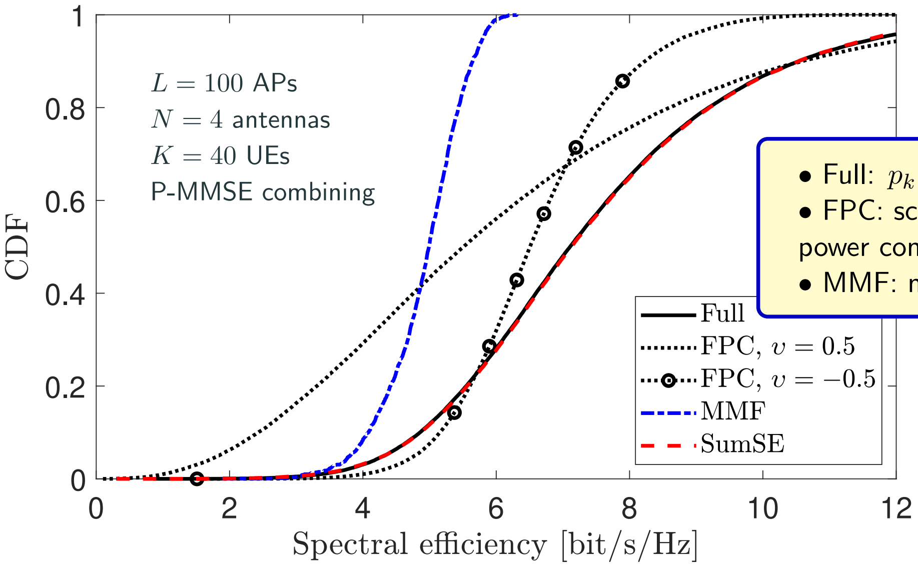
- The original fractional power control is not scalable

- Modify it such that the number of the terms in the denominator does not grow with K:

$$p_k = p_{\max} \frac{\left(\sum_{l \in \mathcal{M}_k} \beta_{kl} \right)^{\upsilon}}{\max_{i \in \mathcal{S}_k} \left(\sum_{l \in \mathcal{M}_i} \beta_{il} \right)^{\upsilon}}$$

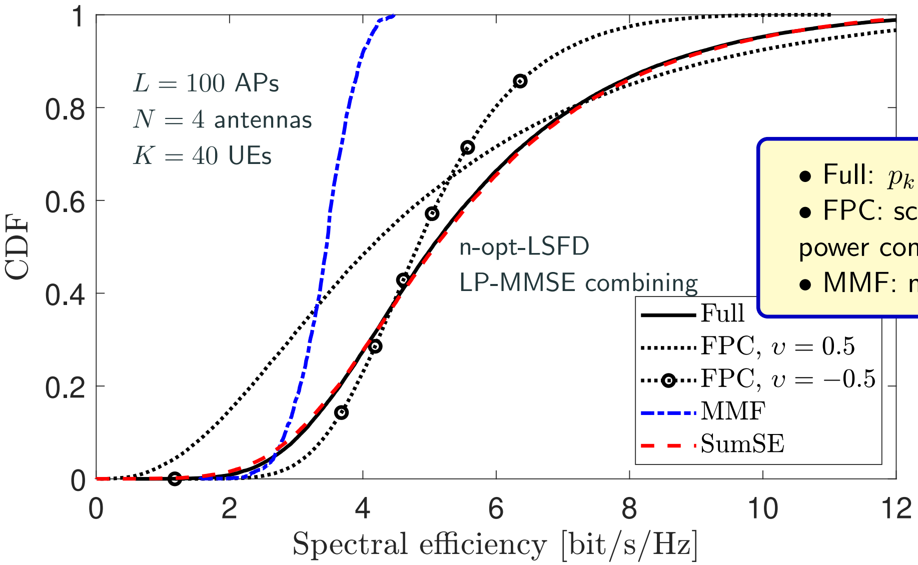
- Recall: $\mathcal{S}_k = \{ i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN} \}$, the UEs that are served by partially the same APs as UE k

Comparison: Uplink Centralized Operation



- Full: $p_k = p_{\max}$
- FPC: scalable fractional power control
- MMF: max-min fairness

Comparison: Uplink Distributed Operation



- Full: $p_k = p_{\max}$
- FPC: scalable fractional power control
- MMF: max-min fairness

Downlink Power Optimization

Power allocation: Dividing the total power $\rho_{\max}$ at each AP between the UEs

- The same per-AP power constraints in centralized and distributed operations
- Handled separately due to
  - Centralized: Coupled precoding vectors between the APs
  - Distributed: No coupling of precoding

AP l transmits the signal $\mathbf{x}_l = \sum_{i=1}^{K} \mathbf{D}_{il}\mathbf{w}_{il}\varsigma_i$.

Centralized operation:

$$\mathbf{w}_k = \sqrt{\rho_k}\frac{\bar{\mathbf{w}}_k}{\sqrt{\mathbb{E}\left\{\left\|\bar{\mathbf{w}}_k\right\|^2\right\}}}$$

- K power coefficients

Distributed operation:

$$\mathbf{w}_{kl} = \sqrt{\rho_{kl}}\frac{\bar{\mathbf{w}}_{kl}}{\sqrt{\mathbb{E}\left\{\left\|\bar{\mathbf{w}}_{kl}\right\|^2\right\}}}$$

- $\sum_{k=1}^{K} |\mathcal{M}_k|$ power coefficients

Centralized Downlink Power Optimization

- $\boldsymbol{\rho} = [\rho_1 \ \ \ldots \ \ \rho_K]^{\mathrm{T}}$
- The effective SINR for UE k in the generic form is

$$\mathsf{SINR}_k^{(\mathrm{dl,c})}(\boldsymbol{\rho}) = \frac{\tilde{b}_k \rho_k}{\tilde{\mathbf{c}}_k^{\mathrm{T}} \boldsymbol{\rho} + \sigma_{\mathrm{DL}}^2}$$

- $\tilde{b}_k \geq 0$: the average channel gain of the desired signal
- $\tilde{\mathbf{c}}_k = [\tilde{c}_{k1} \ \ \ldots \ \ \tilde{c}_{kK}]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^K$: the average channel gains for the interfering signals
- $\sigma_{\mathrm{DL}}^2 \geq 0$: the noise variance

$$\mathbf{w}_k = \sqrt{\rho_k} \frac{\bar{\mathbf{w}}_k}{\sqrt{\mathbb{E}\left\{\|\bar{\mathbf{w}}_k\|^2\right\}}} = \sqrt{\rho_k} \left[\ \bar{\mathbf{w}}_{k1}' \ \ \ldots \ \ \bar{\mathbf{w}}_{kL}' \ \right]^{\mathrm{T}}$$

Per-AP power constraints: $\sum_{k \in \mathcal{D}_l} \rho_k \mathbb{E}\left\{\|\bar{\mathbf{w}}_{kl}'\|^2\right\} \leq \rho_{\max}, \quad l = 1, \ldots, L$

Centralized Downlink Max-Min Fair SE Maximization

- The same SINR structure with K power coefficients as in uplink
- Almost the same optimization algorithms
- Difference: one power constraint per AP

The max-min fairness optimization problem is

$$\begin{aligned}
\underset{\boldsymbol{\rho} \geq \mathbf{0}_K}{\text{maximize}} \quad & \min_{k \in \{1,\ldots,K\}} \quad \frac{\tilde{b}_k \rho_k}{\tilde{\mathbf{c}}_k^{\mathrm{T}} \boldsymbol{\rho} + \sigma_{\mathrm{DL}}^2} \\
\text{subject to} \quad & \sum_{k \in \mathcal{D}_l} \rho_k \mathbb{E}\left\{ \left\| \bar{\mathbf{w}}_{kl}' \right\|^2 \right\} \leq \rho_{\max}, \quad l = 1, \ldots, L
\end{aligned}$$

Fixed-point algorithm that converges to the optimal solution

- $\rho_k \leftarrow \frac{\rho_k}{\mathrm{SINR}_k(\boldsymbol{\rho})}, \; k = 1, \ldots, K$
- $\boldsymbol{\rho} \leftarrow \frac{\rho_{\max}}{\max\limits_{l \in \{1,\ldots,L\}} \sum\limits_{k \in \mathcal{D}_l} \rho_k \mathbb{E}\{\|\bar{\mathbf{w}}_{kl}'\|^2\}} \boldsymbol{\rho}$

It converges quickly, but the complexity grows with K, which makes it not scalable!

Centralized Downlink Sum SE Maximization

- The sum SE maximization problem is

$$
\begin{aligned}
\underset{\boldsymbol{\rho} \geq \mathbf{0}_K}{\text{maximize}} \quad & \sum_{k=1}^{K} \log_2 \left(1 + \frac{\tilde{b}_k \rho_k}{\tilde{\mathbf{c}}_k^{\mathrm{T}} \boldsymbol{\rho} + \sigma_{\mathrm{DL}}^2} \right) \\
\text{subject to} \quad & \sum_{k \in \mathcal{D}_l} \rho_k \mathbb{E}\left\{ \left\| \bar{\mathbf{w}}'_{kl} \right\|^2 \right\} \leq \rho_{\max}, \quad l = 1, \ldots, L.
\end{aligned}
$$

- A local optimum by the block coordinate descent algorithm based on WMMSE reformulation [SRL+11]

- No closed-form updates unlike the uplink counterpart
- A convex quadratically-constrained quadratic programming problem at each iteration
- Not scalable!

[SRL+11] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," IEEE TSP, 2011.

Scalable Centralized Downlink Power Allocation

- The transmit powers of different APs are coupled through the centralized precoding vector $\frac{\bar{\mathbf{w}}_k}{\sqrt{\mathbb{E}\left\{\|\bar{\mathbf{w}}_k\|^2\right\}}}$
- Direction of the precoding vector matters to cancel the interference!
- If one AP increases its power allocation to the UE, then all the other serving APs need to do the same

Network-wide equal power allocation:

$$\rho_k = \frac{\rho_{\max}}{\tau_p}$$

- Each AP serves at most τ_p UEs using the joint pilot assignment and DCC algorithm described earlier $\rightarrow$ Each per-AP constraint is satisfied
- Computation does not grow with K $\rightarrow$ Scalable!

Another Scalable Centralized Downlink Power Allocation

Inspired by the fractional power control

$$\rho_k \propto \left(\sum_{l \in \mathcal{M}_k} \beta_{kl} \right)^{\upsilon}$$

- $\upsilon = 0 \rightarrow$ Network-wide equal power allocation
- $\upsilon > 0 \rightarrow$ More power to the UEs with higher total channel gain
- $\upsilon < 0 \rightarrow$ More power to the UEs with less total channel gain

- Difficulty: ρ_k affects the transmit power of all the APs with indices $l \in \mathcal{M}_k$ as

$$\sum_{k \in \mathcal{D}_l} \rho_k \mathbb{E}\left\{ \left\| \bar{\mathbf{w}}'_{kl} \right\|^2 \right\} \leq \rho_{\max}$$

- AP l consumes at most the power $\rho_k \omega_k$ for each UE $k \in \mathcal{D}_l$ where

$$\omega_k = \max_{\ell \in \mathcal{M}_k} \mathbb{E}\left\{ \left\| \bar{\mathbf{w}}'_{k\ell} \right\|^2 \right\}$$

Another Scalable Centralized Downlink Power Allocation

An additional tuning:

$$\rho_k \propto \frac{\left(\sum_{l \in \mathcal{M}_k} \beta_{kl}\right)^{\upsilon}}{\omega_k^{\kappa}}$$

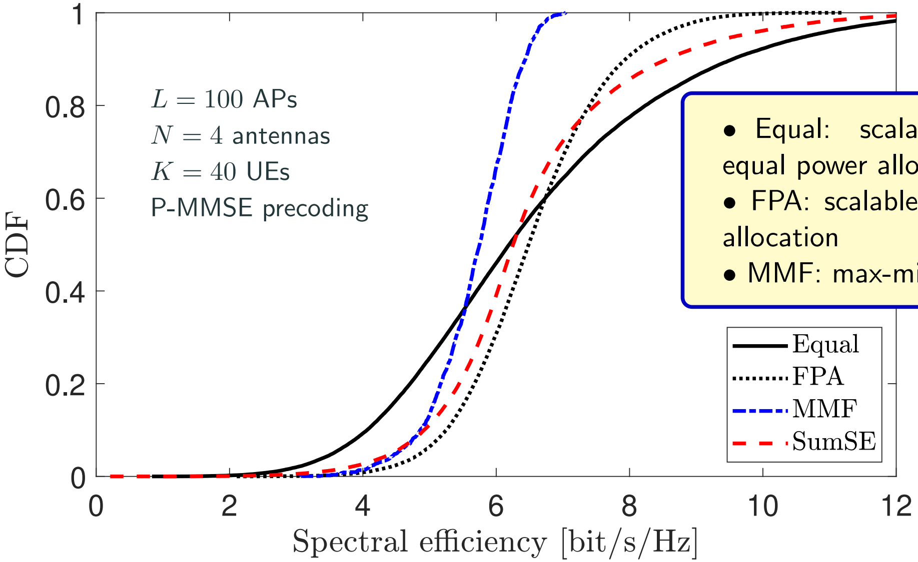
- $0 \leq \kappa \leq 1$ reshapes the ratio of power
- Motivation: When ω_k is too high, i.e., very close to 1, then most of the APs that serve UE k will cut down their power by supposing they use much more power than expected.

To satisfy the power limit $\rho_{\max}$ of each AP, we normalize the power coefficients as

$$\rho_k = \rho_{\max} \frac{\left(\sum_{l \in \mathcal{M}_k} \beta_{kl}\right)^{\upsilon} \omega_k^{-\kappa}}{\max_{\ell \in \mathcal{M}_k} \sum_{i \in \mathcal{D}_\ell} \left(\sum_{l \in \mathcal{M}_i} \beta_{il}\right)^{\upsilon} \omega_i^{1-\kappa}}$$

- Computational complexity does not grow with K
- Scalable!

Comparison: Downlink Centralized Operation



- Equal: scalable network-wide equal power allocation
- FPA: scalable fractional power allocation
- MMF: max-min fairness

Distributed Downlink Power Allocation

- $\{\rho_{kl} : l \in \mathcal{M}_k, k = 1, \dots, K\}$
- Introduce $\tilde{\rho}_{kl} = \sqrt{\rho_{kl}} \geq 0$ for tractable formulations
- The effective SINR for UE k in the generic form is

$$\mathsf{SINR}_k^{(\mathrm{dl},\mathsf{d})}\left(\{\tilde{\boldsymbol{\rho}}_i\}\right) = \frac{\left|\tilde{\mathbf{b}}_k^{\mathrm{T}} \tilde{\boldsymbol{\rho}}_k\right|^2}{\sum_{i=1}^{K} \tilde{\boldsymbol{\rho}}_i^{\mathrm{T}} \tilde{\mathbf{C}}_{ki} \tilde{\boldsymbol{\rho}}_i - \left|\tilde{\mathbf{b}}_k^{\mathrm{T}} \tilde{\boldsymbol{\rho}}_k\right|^2 + \sigma_{\mathrm{DL}}^2}$$

- $\tilde{\boldsymbol{\rho}}_k = [\tilde{\rho}_{k1} \ \ \dots \ \ \tilde{\rho}_{kL}]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^{L}$
- $\tilde{\mathbf{b}}_k \in \mathbb{R}_{\geq 0}^{L}$
- $\tilde{\mathbf{C}}_{ki} \in \mathbb{C}^{L \times L}$: positive semidefinite matrices

$$\mathbf{w}_{kl} = \sqrt{\rho_{kl}} \frac{\bar{\mathbf{w}}_{kl}}{\sqrt{\mathbb{E}\left\{\|\bar{\mathbf{w}}_{kl}\|^2\right\}}} = \tilde{\rho}_{kl} \frac{\bar{\mathbf{w}}_{kl}}{\sqrt{\mathbb{E}\left\{\|\bar{\mathbf{w}}_{kl}\|^2\right\}}}$$

Per-AP power constraints: $\sum_{k \in \mathcal{D}_l} \tilde{\rho}_{kl}^2 \leq \rho_{\max}, \quad l = 1, \dots, L$

Distributed Downlink Max-Min Fair SE Maximization

Previous fixed-point algorithm cannot be used!

The max-min fairness optimization problem in the epigraph form is

$$\begin{aligned} \underset{\tilde{\boldsymbol{\rho}}_k \geq \mathbf{0}_L, \forall k, t \geq 0}{\text{maximize}} \quad & t \\ \text{subject to} \quad & \frac{\left| \tilde{\mathbf{b}}_k^{\mathrm{T}} \tilde{\boldsymbol{\rho}}_k \right|^2}{\sum_{i=1}^{K} \tilde{\boldsymbol{\rho}}_i^{\mathrm{T}} \tilde{\mathbf{C}}_{ki} \tilde{\boldsymbol{\rho}}_i - \left| \tilde{\mathbf{b}}_k^{\mathrm{T}} \tilde{\boldsymbol{\rho}}_k \right|^2 + \sigma_{\mathrm{DL}}^2} \geq t, \quad k = 1, \ldots, K \\ & \sum_{k \in \mathcal{D}_l} \tilde{\rho}_{kl}^2 \leq \rho_{\max}, \quad l = 1, \ldots, L \end{aligned}$$

Obviously not scalable!

- A quasi-convex problem $\rightarrow$ Bisection search over t finds the optimal solution by solving a sequence of second-order cone programming (SOCP) problems

[NAY+17] H. Q. Ngo and A. E. Ashikhmin and H. Yang and E. G. Larsson and T. L. Marzetta, "Cell-free massive MIMO versus small cells," IEEE Transactions on Wireless Communications, 2017.

Distributed Downlink Sum SE Maximization

- The sum SE maximization problem is

$$\begin{aligned} \underset{\tilde{\boldsymbol{\rho}}_k \geq \mathbf{0}_L, \forall k}{\text{maximize}} \quad & \sum_{k=1}^{K} \log_2 \left(1 + \frac{\left| \tilde{\mathbf{b}}_k^{\mathrm{T}} \tilde{\boldsymbol{\rho}}_k \right|^2}{\sum_{i=1}^{K} \tilde{\boldsymbol{\rho}}_i^{\mathrm{T}} \tilde{\mathbf{C}}_{ki} \tilde{\boldsymbol{\rho}}_i - \left| \tilde{\mathbf{b}}_k^{\mathrm{T}} \tilde{\boldsymbol{\rho}}_k \right|^2 + \sigma_{\mathrm{DL}}^2} \right) \\ \text{subject to} \quad & \sum_{k \in \mathcal{D}_l} \tilde{\rho}_{kl}^2 \leq \rho_{\max}, \quad l = 1, \ldots, L \end{aligned}$$

- A local optimum by the block coordinate descent algorithm based on WMMSE reformulation [SRL+11]

- A convex quadratically-constrained quadratic programming problem at each iteration
- Not scalable!

[SRL+11] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," IEEE TSP, 2011.

Scalable Distributed Downlink Power Allocation

- Extending the power allocation scheme in [IFL+19] for arbitrary N and DCC,

$$\rho_{kl} = \begin{cases} \rho_{\max} \frac{f(\mathcal{G}_{kl})}{\sum_{i \in \mathcal{D}_l} f(\mathcal{G}_{il})} & \text{if } k \in \mathcal{D}_l \\ 0 & \text{otherwise} \end{cases}$$

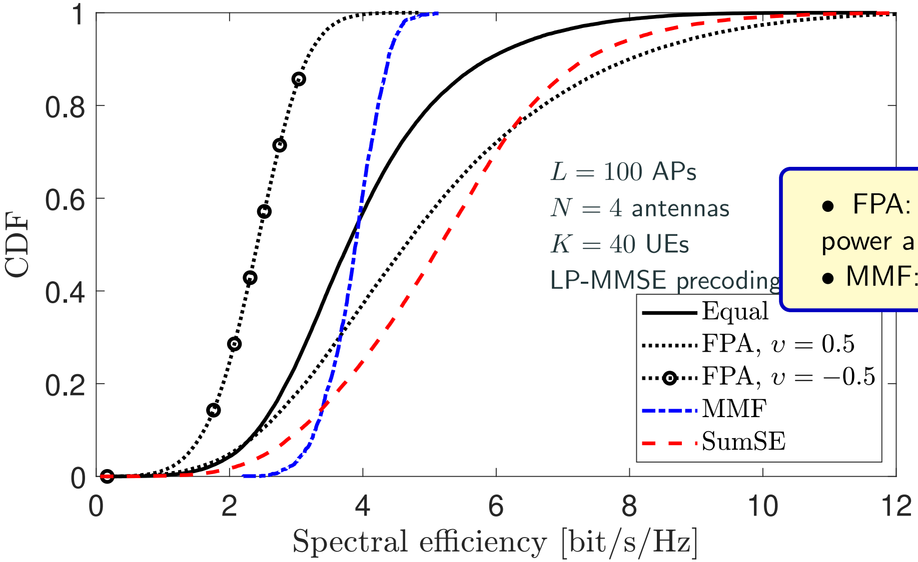
  where $f(\cdot)$ is a pre-determined function of the channel statistics $\mathcal{G}_{il} = \{\mathbf{R}_{il}, \mathbf{C}_{il}\}$
- The function $f(\mathcal{G}_{il})$ determines the relative importance of transmitting to UE i
- Denominator is for the normalization for per-AP power constraints

- $f(\mathcal{G}_{il}) = 1 \rightarrow$ Per-AP equal power allocation with $\rho_{kl} = \frac{\rho_{\max}}{|\mathcal{D}_l|}$
- $f(\mathcal{G}_{il}) = (\beta_{il})^{\upsilon} \rightarrow \rho_{kl} = \begin{cases} \rho_{\max} \frac{(\beta_{kl})^{\upsilon}}{\sum_{i \in \mathcal{D}_l} (\beta_{il})^{\upsilon}} & \text{if } k \in \mathcal{D}_l \\ 0 & \text{otherwise} \end{cases}$

$\upsilon \in [0, 1]$ is preferred [IFL19]

[IFL19] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," IEEE ICC 2019.

Comparison: Downlink Distributed Operation



- FPA: scalable fractional power allocation
- MMF: max-min fairness

Takeaways and Open Problems

Takeaways

1. Cell-free mMIMO attempts to provide uniformly good service everywhere
   - Massive macro-diversity from distributed antennas
   - Beyond 5G or 6G technology?

2. MMSE processing is a key prerequisite for Cell-free mMIMO
   - to outperform Cellular mMIMO and small cells (roughly double SE per UE)

3. Centralized implementation is theoretically preferable:
   - It can simultaneously increase the SE and reduce the fronthaul signaling
   - C-RAN methods can be utilized for implementation

4. Distributed implementation is practically preferable:
   - Must utilize serial fronthaul to make sense – Radio stripes concept
   - Processing complexity is reduced and shared among APs.

Future Development

The physical layer is fairly well understood
...but the topic is still in its infancy

Open problems:

- Initial access and synchronization
- Hardware impairments due to low-cost components
- Quantization of fronthaul signaling
- Fronthaul deployment
- Lack of channel models
- Prototype development
- ...

Acknowledgments

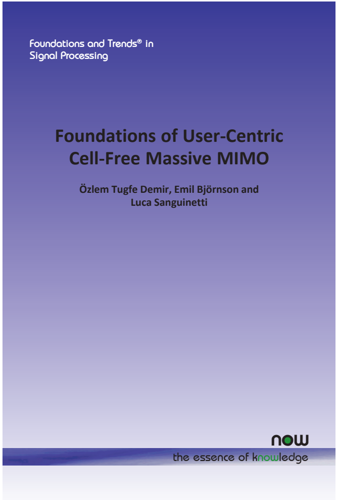
Thanks to our colleagues and collaborators:

- Zheng Chen, Sucharita Chakraborty, Giovanni Interdonato, Erik G. Larsson, Özgecan Özdogan (Linköping University)
- Pål Frenger, Martin Hessler (Ericsson Research)
- Hien Quoc Ngo (Queen's University Belfast)
- Jiayi Zhang (Beijing Jiaotong University)
- Cicek Cavdar (KTH Royal Institute of Technology)

Thanks to our funding agencies:

- Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation
- Swedish Foundation for Strategic Research (SSF)

Questions?



https://www.nowpublishers.com/article/Details/SIG-109

https://github.com/emilbjornson/cell-free-book