Machine Type Communications (MTC) in 5G: Challenges, Algorithmic Design and Performance Analysis

Mohamad Assaad¹
¹TCL Chair on 5G, CentraleSupélec, Gif-sur-Yvette, France.

September 13, 2018
5G Chair Holder at CentraleSupelec since 2017
Overview of 5G Massive MIMO Spatial division multiplexing based user scheduling Coherence time based user grouping in massive MIMO Multiple Access Techniques Random Access for multi-service scenarios Transmission and Control of devices in dense networks Conclusion

TCL Corporation

HANDSET

TABLET

PANEL

TV

AIR CONDITIONING

WASHING MACHINE

REFRIGERATOR

⑦ Global

⑦ Global

⑥ Global

③ Global

⑤ China

⑥ China

⑦ China
TCL Corporation

16.1 billion $

More than 11,000 patents (#3 in China)

Presence in 160 countries
Outline

Overview of 5G

Massive MIMO

Spatial division multiplexing based user scheduling

Coherence time based user grouping in massive MIMO

Multiple Access Techniques

Random Access for multiservice scenarios

Transmission and Control of devices in dense networks

Conclusion
Section 1

Overview of 5G
Mobile communications have been fundamental in producing our contemporary information societies.

A new generation of mobile networks is periodically required to provide a considerable leap in performance.

Wireless communications are at crossroads:
- Proliferation of smart devices.
- Ever growing demand for high data capacity.
- Data hungry applications (VR, AR, ...).

5G is envisioned to provide a unifying and flexible connectivity fabric for a wide range of industries with heterogeneous service requirements.
5G: Enabler for Smart World

- Previous generations focused on increasing data rates and spectral efficiency
- 5G is also focusing on IoT requirements: connected cars, robot control, massive connection of sensors, etc.
- Powerful enabler for smart worlds:
  - Smart Home
  - Smart Healthcare
  - Smart Car
- Support Diverse Multimedia Services

**Figure:** 5G requirements: creating a smart world
Vertical industries business cases

- Media & Entertainment
  - Ultra high fidelity media
  - On-site live event experience
  - Cooperative media production
  - Collaborative gaming

- Factories of the future
  - Time-critical process optimization inside factory to support zero-defect manufacturing
  - Remote maintenance and control optimizing
  - Energy management such as monitoring energy consumption and conditions and resource management
  - Stock control
  - Factory automation such as production line control, security and safety, and video surveillance
  - Optimization of logistic flows, etc.

- eHealth
  - Robotics (remote surgery, etc)
  - Remote monitoring of health or wellness data (Blood pressure monitors, Glucose monitoring, Hearing aids, Electrocardiogram (ECG) monitors, Insulin pumps, etc.)
  - etc.
Vertical industries business cases

- **Automotive**
  - Automated driving
  - Road safety and traffic efficiency services
  - Information society on the road
  - Electronic Toll Collection Services
  - Taxi adverts
  - Vehicle Data Services
  - Vehicle Diagnostic & Maintenance Report
  - Remote Maintenance services
  - Traffic Accident Information collection

- **Public services**
  - Street Light Automation
  - Devices, Virtual devices and Things
  - Car/Bicycle Sharing Services
  - Smart parking
  - Information Delivery service in the devastated area
Vertical industries business cases

Figure: Connected Car View

Figure: Global Connected Car Shipment

2 BI Intelligence ScotiaBank
Vertical industries business cases

- Residential
  - Home Energy Management System
  - Plug-In Electrical Charging Vehicles and power feed in home scenario
  - Real-time Audio/Video Communication
  - Event Triggered Task Execution
  - Semantic Home Control

- Energy
  - Smart Grid (Grid access, backhaul, etc.)
  - Wide area Energy related measurement /control system for advanced transmission and distribution automation
  - Smart Meter Reading
  - Oil and Gas Pipeline Cellular/Satellite Gateway

- Agriculture (Smart Irrigation System, etc.)

- New type of devices: smart lenses, robots, Drones, bracelets, etc.
IoT Business Market

Figure: IoT connectivity revenue by industry sector and total connectivity ARPU, worldwide, 2011?2021

\(^3\) Analysis Mason, 2013
5G Use Cases 1

**Figure:** 5G use cases

---

4 NGMN 5G White Paper
5G Use Cases 2

![5G Use Cases Diagram](image)

**Figure:** 5G use cases

---

5NGMN 5G White Paper
5G Use Cases 3

Figure: 5G use cases

---

NGMN 5G White Paper
5G User Experience KPI’s

<table>
<thead>
<tr>
<th>Use case category</th>
<th>User Experienced Data Rate</th>
<th>E2E Latency</th>
<th>Mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadband access in dense areas</td>
<td>DL: 300 Mbps, UL: 50 Mbps</td>
<td>10 ms</td>
<td>On demand, 0-100 km/h</td>
</tr>
<tr>
<td>Indoor ultra-high broadband access</td>
<td>DL: 1 Gbps, UL: 500 Mbps</td>
<td>10 ms</td>
<td>Pedestrian</td>
</tr>
<tr>
<td>Broadband access in a crowd</td>
<td>DL: 25 Mbps, UL: 50 Mbps</td>
<td>10 ms</td>
<td>Pedestrian</td>
</tr>
<tr>
<td>50+ Mbps everywhere</td>
<td>DL: 50 Mbps, UL: 25 Mbps</td>
<td>10 ms</td>
<td>0-120 km/h</td>
</tr>
<tr>
<td>Ultra-low cost broadband access for low ARPU areas</td>
<td>DL: 10 Mbps, UL: 10 Mbps</td>
<td>50 ms</td>
<td>on demand: 0-50 km/h</td>
</tr>
<tr>
<td>Mobile broadband in vehicles (cars, trains)</td>
<td>DL: 50 Mbps, UL: 25 Mbps</td>
<td>10 ms</td>
<td>On demand, up to 500 km/h</td>
</tr>
<tr>
<td>Airplanes connectivity</td>
<td>DL: 15 Mbps per user, UL: 7.5 Mbps per user</td>
<td>10 ms</td>
<td>Up to 1000 km/h</td>
</tr>
<tr>
<td>Massive low-cost/long-range/low-power MTC</td>
<td>Low (typically 1-100 kbps)</td>
<td>Seconds to hours</td>
<td>on demand: 0-500 km/h</td>
</tr>
</tbody>
</table>

**Figure:** 5G User Experience KPI's

---

### 5G User Experience KPI’s

<table>
<thead>
<tr>
<th>Broadband MTC</th>
<th>Ultra-low latency</th>
<th>Resilience and traffic surge</th>
<th>Ultra-high reliability &amp; Ultra-low latency</th>
<th>Ultra-high availability &amp; reliability</th>
<th>Broadcast like services</th>
</tr>
</thead>
</table>
| See the requirements for the Broadband access in dense areas and 50+Mbps everywhere categories | DL: 50 Mbps  
UL: 25 Mbps | DL: 0.1-1 Mbps  
UL: 0.1-1 Mbps | DL: From 50 kbps to 10 Mbps;  
UL: From a few bps to 10 Mbps | DL: 10 Mbps  
UL: 10 Mbps | DL: Up to 200 Mbps  
UL: Modest (e.g. 500 kbps) |
| | <1 ms  
Pedestrian | Regular communication: not critical  
0-120 km/h | 1 ms  
on demand: 0-500 km/h | 10 ms  
on demand, 0-500 km/h | <100 ms  
on demand: 0-500 km/h |

**Figure:** 5G User Experience KPI’s

---

8 NGMN 5G White Paper
### System Performance KPI's

<table>
<thead>
<tr>
<th>Use case category</th>
<th>Connection Density</th>
<th>Traffic Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadband access in dense areas</td>
<td>200-2500 /km²</td>
<td>DL: 750 Gbps /km²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UL: 125 Gbps /km²</td>
</tr>
<tr>
<td>Indoor ultra-high broadband access</td>
<td>75,000 / km²</td>
<td>DL: 15 Tbps/km²</td>
</tr>
<tr>
<td>(75/1000 m² office)</td>
<td></td>
<td>(15 Gbps / 1000 m²)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UL: 2 Tbps/km²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2 Gbps / 1000 m²)</td>
</tr>
<tr>
<td>Broadband access in a crowd</td>
<td>150,000 / km²</td>
<td>DL: 3.75 Tbps/km²</td>
</tr>
<tr>
<td>(30,000 / stadium)</td>
<td></td>
<td>(DL: 0.75 Tbps/stadium)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UL: 7.5 Tbps/km²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.5 Tbps / stadium)</td>
</tr>
<tr>
<td>60+ Mbps everywhere</td>
<td>400 / km² in suburban</td>
<td>DL: 20 Gbps / km²</td>
</tr>
<tr>
<td></td>
<td>100 / km² in rural</td>
<td>UL: 10 Gbps / km²</td>
</tr>
<tr>
<td>Ultra-low cost broadband access for low</td>
<td>16 / km²</td>
<td>DL: 5 Gbps / km²</td>
</tr>
<tr>
<td>ARPU areas</td>
<td></td>
<td>UL: 2.5 Gbps / km²</td>
</tr>
<tr>
<td>Mobile broadband in vehicles (cars, trains)</td>
<td>2000 / km²</td>
<td>DL: 100 Gbps / km²</td>
</tr>
<tr>
<td>(500 active users per train x 4 trains,</td>
<td>(25 Gbps per train, 50 Mbps per car)</td>
<td></td>
</tr>
<tr>
<td>or 1 active user per car x 2000 cars)</td>
<td>UL: 50 Gbps / km²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.5 Gbps per train, 25 Mbps per car)</td>
<td></td>
</tr>
<tr>
<td>Airplanes connectivity</td>
<td>80 per plane</td>
<td>DL: 1.2 Gbps / plane</td>
</tr>
<tr>
<td></td>
<td>60 airplanes per 18,000 km²</td>
<td>UL: 600 Mbps / plane</td>
</tr>
<tr>
<td>Massive low-cost/long-range/low-power</td>
<td>Up to 200,000 / km²</td>
<td>Non critical</td>
</tr>
<tr>
<td>MTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broadband MTC</td>
<td>See the requirements for the Broadband access in dense areas and 50+Mbps everywhere categories</td>
<td></td>
</tr>
<tr>
<td>Ultra-low latency</td>
<td>Not critical</td>
<td>Potentially high</td>
</tr>
<tr>
<td>Resilience and traffic surge</td>
<td>10,000 / km²</td>
<td>Potentially high</td>
</tr>
</tbody>
</table>

**Figure:** System Performance KPI's

---

9 NGMN 5G White Paper
# 5G Network Capabilities

<table>
<thead>
<tr>
<th>Attribute</th>
<th>3GPP Release-12 capability</th>
<th>Improvement needed to meet NGMN requirements</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data rate (per user)</td>
<td>Up to 100 Mb/s on average Peaks of 600 Mb/s (Cat 11/12)</td>
<td>&gt; 10X expected on average and peak rates</td>
<td>Technology should allow operators to optimize topology to achieve 1 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; 100X expected on cell edge</td>
<td>end-to-end</td>
</tr>
<tr>
<td>End-to-end latency</td>
<td>10 ms for two-way RAN (prescheduled) Typically, up to 50 ms end-to-end if other factors are considered (e.g., transmission, CN, internet, proxy servers)</td>
<td>&gt; 10X (smaller)</td>
<td></td>
</tr>
<tr>
<td>Mobility</td>
<td>Functional up to 350 km/h (for certain bands up to 500 km/h) No support for civil aviation</td>
<td>&gt; 1.5X</td>
<td>Functional in 5G means sustained service quality for the considered use case. 5G in addition should support civil aviation use case.</td>
</tr>
<tr>
<td>Spectral efficiency</td>
<td>DL: 0.074 – 6.1 b/s/Hz UL: 0.07 – 4.3 b/s/Hz depending on cell edge or average, deployment scenario, and FDD or TDD</td>
<td>Pushing the envelope for substantial increase</td>
<td>Requirements should be specified by NGMN operators jointly with the industry in due course.</td>
</tr>
<tr>
<td>Connection density</td>
<td>Typically ~2,000 active users/km²</td>
<td>&gt; 100X</td>
<td></td>
</tr>
</tbody>
</table>

**Figure:** 5G Network Capabilities \(^\text{10}\)

\(^\text{10}\) NGMN 5G White Paper
5G mobile communication systems: requirements and promises (1)

- **Human Type Communications**
  - User Bit rate: 1Gbps
  - 100x Capacity increase
  - High area throughput

- **Machine Type Communications**
  - Massive density, Small data size (10 - 300 bytes) Smart city (Metering, etc)
  - Critical machine type communications, URLLC: very low latency (1ms) industry automation, self driving car, etc

- Ubiquitous wireless connectivity (many papers and talks)
- Complexity of the network architecture
- Several Metrics should be improved in 5G
  - High user data rates
  - High area throughput
  - Great scalability in number of connected devices
  - High reliability and lower latency
  - Better coverage with more fairness
  - Improvement of energy efficiency

- Conflicting Metrics!
- Target use cases: Enhanced Mobile Broadband (eMBB), massive Machine Type Communications (mMTC) and Ultra-Reliable and Low Latency Communications (URLCC)
5G mobile communication systems: requirements and promises (2)

5G envisioned improvements:

- 10 folds increase in throughput.
- 10 folds decrease in latency.
- 10 folds connection density.
- 100 folds traffic capacity.
- 100 folds network efficiency.

Figure: 5G requirements: A need for a scalable and adaptive system [1]
Challenges

- **MTC**
  - Low throughput smart machines
  - M2M requirements differ from mobile broadband: Make technology scalable to a wide range of use cases
  - Years of battery life: Battery-powered and deployed in field for years
  - Low cost/complexity of the devices
  - Signaling overhead
  - Coverage: Could be deployed in challenging environment

- **URLCC**
  - Challenging E2E latency: 1 to 10 ms
  - High reliability: > 99%

- **eMBB**
  - Higher bit rate
  - Higher capacity
  - Enhanced Quality of Experience (QoE)
  - etc.
5G: A concentration of new paradigms and innovative technologies

- Meeting the aforementioned requirements necessitates
  - drastic changes in the network paradigm.
  - disruptive innovations

- 5G networks can call upon a wide range of new technologies [2]:
  - Millimeter-Wave Mobile Communications.
  - Massive MIMO Communications.
  - Non-Orthogonal Multiple Access (NOMA).
  - Full-Duplex Wireless Communications.
  - Carrier aggregation and Multicarrier Modulations.
  - Larger spectrum.
  - Sidelink communication (coverage, energy, traffic offload, etc).
  - New waveform and heterogeneous OFDM numerology.

- 5G will change how networking is performed
  - Cloud radio access network (CRAN).
  - Energy Harvesting.
  - V2X.
  - Self Organizing Network (SON).
  - Network Slicing
  - Fog Computing.
  - User-Centric Wireless Network (proactive caching, etc).
  - New Access techniques (collision issues, extension of battery life, massive density of devices, etc).
3GPP standardization timeline

Figure: 3GPP standardization timeline [1]; SA: Stand Alone, NR: New Radio
Section 2

Massive MIMO
Massive MIMO: The power of numbers

Massive MIMO

- Meeting the spectral efficiency (SE) requirement (a ten-fold improvement in SE [2]) and can result in power consumption gain [25].
- Enabling the spatial multiplexing of a considerable number of mobile devices.
- Arrays of some hundred cheap antenna elements at the BS.

Massive MIMO $\mapsto$ Interesting gains that come with their toll of challenges:

- Spectral efficiency gain
- Energy efficiency gain (power scaling of cellular user $1/\sqrt{M}$)
- Simple processing
- User scheduling
- CSI acquisition
Massive MIMO: The power of numbers

Massive MIMO

- Meeting the spectral efficiency (SE) requirement (a ten-fold improvement in SE [2]) and can result in power consumption gain [25].
- Enabling the spatial multiplexing of a considerable number of mobile devices.
- Arrays of some hundred cheap antenna elements at the BS.

Massive MIMO $\rightarrow$ Interesting gains that come with their toll of challenges:

- Spectral efficiency gain
- Energy efficiency gain (power scaling of cellular user $1/\sqrt{M}$)
- Simple processing
- User scheduling
- CSI acquisition
Massive MIMO: The power of numbers

Massive MIMO

- Meeting the spectral efficiency (SE) requirement (a ten-fold improvement in SE [2]) and can result in power consumption gain [25].
- Enabling the spatial multiplexing of a considerable number of mobile devices.
- Arrays of some hundred cheap antenna elements at the BS.

Massive MIMO → Interesting gains that come with their toll of challenges:

- Spectral efficiency gain
- Energy efficiency gain (power scaling of cellular user $1/\sqrt{M}$)
- Simple processing
- User scheduling
- CSI acquisition
massive MIMO Model

- **Time-Division Duplex (TDD) Protocol**
  - Uplink and downlink separated in time
  - Uplink fraction $\zeta^{(ul)}$ and downlink fraction $\zeta^{(dl)}$

- **Coherence Block**
  - $B$ Hz bandwidth = $B$ “channel uses” per second (symbol time $1/B$)
  - Channel stays fixed for $U$ channel uses (symbols) = Coherence block
  - Determines how often we send pilot signals to estimate channels

Assumption: Perfect channel estimation (relaxed later)
massive MIMO Model

- Flat-Fading Channels
  - Channel between BS and User $k$: $h_k \in \mathbb{C}^M$
  - Rayleigh fading: $h_k \sim CN(0, \lambda_k I)$
  - Channel variances $\lambda_k$: Random variables, pdf $f_\lambda(x)$

- Uplink Transmission
  - User $k$ transmits signal $s_k$ with power $E(|s_k|^2) = p_k^{(ul)}$ [Joule/channel use]
  - Received signal at BS:
    \[
    y = h_k s_k + \sum_{i \neq k} h_i s_i + n
    \]
    Signals from other users (interference)
    Noise $\sim CN(0, \sigma^2 I)$
  - Recover $s_k$ by receive beamforming $g_k$ as $g_k^H y$:
    \[
    SINR_k^{(ul)} = \frac{E(|s_k|^2 |g_k^H h_k|^2)}{\sum_{i \neq k} E(|s_i|^2 |g_k^H h_i|^2) + E(|g_k^H n|^2)} = \frac{p_k^{(ul)} |g_k^H h_k|^2}{\sum_{i \neq k} p_i^{(ul)} |g_k^H h_i|^2 + \sigma^2 |g_k|^2}
    \]
Motivation of massive MIMO

Consider a $N \times K$ MIMO MAC:

$$ y = \sum_{k=1}^{K} h_k x_k + n $$

where $h_k, n$ are i.i.d. with zero mean and unit variance.

By the strong law of large numbers:

$$ \frac{1}{N} h_m^H y \xrightarrow{a.s.} x_m $$

With an unlimited number of antennas,

- uncorrelated interference and noise vanish,
- the matched filter is optimal,
- the transmit power can be made arbitrarily small.

massive MIMO Model

Overview of 5G Massive MIMO Spatial division multiplexing based user scheduling Coherence time based user grouping in massive MIMO Multiple Access

About some fundamental assumptions

- The receiver has perfect channel state information (CSI). What happens if the channel must be estimated?

- The number of interferers $K$ is small compared to $N$. What does small mean?

- The channel provides infinite diversity, i.e., each antenna gives an independent look on the transmitted signal. What if the degrees of freedom are limited?

- The received energy grows without bounds as $N \to \infty$. Clearly wrong, but might hold up to very large antenna arrays if the aperture scales with $N$. 
A main problem in TDD massive MIMO: Pilot contamination

- The aggressive spatial multiplexing gain of mMIMO, which in turns relies on the knowledge of the (CSI)s at the Base Station (BS).
- CSI acquisition: feedback cost in FDD and Time resources in TDD
  - In TDD, the number of pilots scales with the number of users.
- The problem of pilot contamination:
  - Limited coherence slot \implies Uplink reference signal reuse \implies Pilot contamination \implies inaccurate CSI & Reduced Beamforming precision.
A main problem in TDD massive MIMO: Pilot contamination

- The aggressive spatial multiplexing gain of mMIMO, which in turns relies on the knowledge of the (CSI)s at the Base Station (BS).
- CSI acquisition: feedback cost in FDD and Time resources in TDD
  - In TDD, the number of pilots scales with the number of users.
- The problem of pilot contamination:
  - Limited coherence slot $\Rightarrow$ Uplink reference signal reuse $\Rightarrow$ Pilot contamination $\Rightarrow$ inaccurate CSI & Reduced Beamforming precision.
A main problem in TDD massive MIMO: Pilot contamination

- The aggressive spatial multiplexing gain of mMIMO, which in turns relies on the knowledge of the (CSI)s at the Base Station (BS).
- CSI acquisition: feedback cost in FDD and Time resources in TDD
  - In TDD, the number of pilots scales with the number of users.
- The problem of pilot contamination:
  - Limited coherence slot ⇒ Uplink reference signal reuse ⇒ Pilot contamination ⇒ inaccurate CSI & Reduced Beamforming precision.
On channel estimation and pilot contamination

1. The receiver estimates the channels based on pilot sequences.
2. The number of orthogonal sequences is limited by the coherence time.
3. Thus, the pilot sequences must be reused.

Assume that transmitter $m$ and $j$ use the same pilot sequence:

$$\hat{h}_{m} = h_{m} + \underbrace{h_{j}}_{\text{pilot contamination}} + \underbrace{n_{m}}_{\text{estimation noise}}$$

Thus,

$$\frac{1}{N} \hat{h}_{m} \hat{y} \xrightarrow{\scriptscriptstyle N \to \infty, K = \text{const.}} x_{m} + x_{j}$$

With an unlimited number of antennas,

- uncorrelated interference, noise and estimation errors vanish,
- the matched filter is optimal,
- the transmit power can be made arbitrarily small ($\sim 1/\sqrt{N}$ [Ngo'11]),
- but the performance is limited by pilot contamination.

massive MIMO Model

System model and channel estimation

Uplink: $L$ BSs with $N$ antennas, $K$ UTs per cell. Received signal at BS $j$:

$$y_j = \sqrt{\rho} \sum_{l=1}^{L} H_{j_l} x_l + n_j$$

The columns of $H_{j_l}$ ($N \times K$) are modeled as

$$h_{j_lk} = R_{j_lk}^{\frac{1}{2}} w_{j_lk}, \quad w_{j_lk} \sim \mathcal{CN}(0, I_N)$$

Channel estimation:

$$y_{jk}^T = h_{jk} + \sum_{l \neq j} h_{jk} + \frac{1}{\sqrt{\rho}} n_{jk}$$

MMSE estimate:

$$h_{jk} = \hat{h}_{jk} + \tilde{h}_{jk}$$

$$\hat{h}_{jk} \sim \mathcal{CN}(0, \Phi_{jk}), \quad \tilde{h}_{jk} \sim \mathcal{CN}(0, \Phi_{jk} - \Phi_{jk})$$

$$\Phi_{jk} = R_{jlk} Q_{jk} R_{jlk}, \quad Q_{jk} = \left(\frac{1}{\rho} I_N + \sum_l R_{jlk}\right)^{-1}$$
massive MIMO Model

A simple multi-cell scenario

- intercell interference factor $\alpha \in [0, 1]$
- transmit power per UT: $p$
- $H_{ji} = [h_{j1}, \ldots, h_{jK}] = \sqrt{N/P}AW_{ji}$
- $A \in \mathbb{C}^{N \times P}$ composed of $P \leq N$ columns of a unitary matrix
- $W_{ij} \in \mathbb{C}^{P \times K}$ have i.i.d. elements with zero mean and unit variance

Assumptions:
- $P$ channel degrees of freedom, i.e., rank ($H_{ji}$) = $\min(P, K)$ [Ngo’11]
- energy scales linearly with $N$, i.e., $E[tr(H_{ji}H_{ji}^H)] = KN$
- only pilot contamination, i.e., no estimation noise:
  $$\hat{h}_{ij} = h_{ij} + \sqrt{\alpha} \sum_{i \neq j} h_{ijk}$$
massive MIMO Model

Asymptotic performance of the matched filter

Assume that $N$, $K$ and $P$ grow infinitely large at the same speed:

$$\text{SINR}^{\text{MF}} \approx \frac{\bar{L}}{\rho N} + \frac{K \bar{I}^2}{P} + \frac{\alpha(\bar{I} - 1)}{\rho N}$$

where $\bar{I} = 1 + \alpha(L - 1)$.

Observations:

- The effective SNR $\rho N$ increases linearly with $N$.
- The multiuser interference depends on $P/K$ and not on $N$.
- Ultimate performance limit:

$$\text{SINR}^{\text{MF}} \xrightarrow{\text{as}, N,P \to \infty, K=\text{const.}} \text{SINR}^{\infty} = \frac{1}{\alpha(\bar{I} - 1)}$$

Asymptotic performance of the MMSE detector

Assume that $N$, $K$ and $P$ grow infinitely large at the same speed:

$$\text{SINR}^{\text{MMSE}} \approx \frac{1}{\frac{L}{\rho N} X + \frac{K}{P} \frac{L^2}{Y} + \frac{\alpha(L - 1)}{\text{noise}} + \frac{\text{multi-user interference}}{\text{pilot contamination}}}$$

where $L = 1 + \alpha(L - 1)$ and $X, Y$ are given in closed-form.

Observations:

- As for the MF, the performance depends only on $\rho N$ and $P/K$.
- The ultimate performance of MMSE and MF coincide:

$$\text{SINR}^{\text{MMSE}} \xrightarrow{\text{as}} \frac{1}{\text{SINR}^\infty} = \frac{1}{\alpha(L - 1)}$$

massive MIMO Model

![Numerical results](image)

**Figure:** Uplink performance results
massive MIMO Model

Figure: Downlink performance results
Section 3

Spatial division multiplexing based user scheduling
Pilot contamination reduction: User grouping

- **Spatial User grouping**
  - Spatial User grouping (covariance based methods): Joint Spatial and Division Multiplexing (JSDM)\(^{11}\). The users are separated based on the covariance of their channels.
  - revolves around the idea of splitting the precoder \(V\) into two stages: \(V = BP\), where \(B\) and \(P\) are referred to as the outer and inner precoders respectively.
  - The outer precoder can be designed to match the covariance space of scheduled groups in order to obtain the highest useful signal possible.
  - The inner precoder \(P\) depends on the instantaneous channel realizations and is intended to suppress intra-group interference (ZF, MMSE, etc.).
  - Extension to the case where the covariance matrix of the users' channels is not available\(^{12}\).

- **New user grouping framework**\(^{13}\)

---


\(^{13}\) S. Hajri and M. Assaad: A spatial basis coverage approach for uplink training and scheduling in Massive MIMO systems
Joint Spatial Division and Multiplexing: General Model

Downlink Multi-User MIMO System with $K$ Users and $N_t$ Antennas

$$y = H^H x + z$$  \hspace{1cm} (1)

- $x \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal vector
- $z \in \mathbb{C}^{K \times 1}$ is the additive noise
- $H \in \mathbb{C}^{N_t \times K}$ is the channel matrix

Assumptions
- Gaussian Additive Noise i.e $z \sim \mathcal{CN}(0, \sigma^2 I_K)$
- No LOS i.e Rayleigh Fading with $h_k \sim \mathcal{CN}(0, R_k)$
- Equal-Power-Allocation \textbf{EPA} to all streams
- Total Downlink Power Budget $P$
Joint Spatial And Division Multiplexing

The transmitted signal vector is actually a precoded version of a data vector $\mathbf{x} = \mathbf{Vd}$ where:

- $\mathbf{V} \in \mathbb{C}^{N_t \times S}$ is the precoder
- $\mathbf{d} \in \mathbb{C}^{S \times 1}$ is the data vector
- $S \leq \min\{N_t, K\}$ is the number of independent stream

JSDM Approach:

- Users partitioned into $G$ groups & $\mathbf{V} = \mathbf{BP}$
- $\mathbf{B} \in \mathbb{C}^{N_t \times b}$: depends on channel statistics
- $\mathbf{P} \in \mathbb{C}^{b \times S}$: depends on instantaneous effective channel $\tilde{\mathbf{H}} = \mathbf{B}^H \mathbf{H} \in \mathbb{C}^{b \times K}$
Figure: JSDM
Necessity Of Clustering

Figure: Per Group Processing
Inner Precoder

Assumptions

- Perfect effective CSI available
- Zero forcing (ZF) precoder

The inner precoder is therefore:

\[ P_g = \zeta_g \tilde{H}_g (\tilde{H}_g^H \tilde{H}_g)^{-1} \in \mathbb{C}^{b_g \times S_g} \]  \hspace{1cm} (2)

where \( \zeta_g \) being a normalization factor to ensure that the power budget constraint is satisfied:

\[ \zeta_g^2 = \frac{S_g}{\text{tr}(B_g \tilde{H}_g (\tilde{H}_g^H \tilde{H}_g)^{-2} \tilde{H}_g^H B_g^H)} \]  \hspace{1cm} (3)
Centroid Representation

The centroid that would be taken as a representative of each group’s equivalent covariance:

\[ R_g = \frac{1}{K_g} \sum_{k=1}^{K_g} R_{g_k} \]  

(4)

Let the EVD of the centroid be:

\[ R_g = U_g \Lambda_g U_g^H \]  

(5)

- \( \Lambda_k \) is an \( r_g \times r_g \) diagonal eigenvalues matrix with \( r_g \) being the rank
- \( U_k \in \mathbb{C}^{N_t \times r_g} \) being the set of eigenvectors corresponding to the nonzero eigenvalues.
Outer Precoder

Build the interference matrix:

$$\Xi_g = [U_1^*, \ldots, U_{g-1}^*, U_{g+1}^*, \ldots, U_G^*]$$ \hspace{1cm} (6)

with $U_{g'}^* \in \mathbb{C}^{N_t \times r_{g'}}$ where $r_{g'}^*$ is a design parameter.

Let $[E_g^{(1)}, E_g^{(0)}]$ denotes the set of left eigenvectors of $\Xi_g$ such as:

- $E_g^{(0)}$ of dimension $N_t \times (N_t - \sum_{g' \neq g} r_{g'}^*)$ form a unitary basis for $\text{Span}^\perp (U_{g'}^* : g' \neq g)$

- Project the channel matrix of group $g$ on this space:

$$\hat{R}_g = (E_g^{(0)})^H U_g \Lambda_g U_g^H E_g^{(0)} = G_g \Phi_g G_g^H$$

- Match the $b_g$ strongest eigenmodes of our projected channel $B_g = E_g^{(0)} G_g^{(1)}$
Consider a user terminal (UT) at an azimuth angle $\theta$ and distance $s$ is surrounded by a ring of scatterers of radius $r$ with an angular spread $\Delta$.

- Appropriate model when the base station is elevated and seldom obstructed.
Correlation Entries

By adopting the one-ring model we have:

- \( \Delta \approx \arctan\left( \frac{r}{s} \right) \)

The correlation entry is calculated \( 1 \leq m, p \leq N_t \) for each user using:

\[
[R]_{m,p} = \frac{1}{2\Delta} \int_{\theta-\Delta}^{\theta+\Delta} e^{j k^T(\alpha)(u_m-u_p)} d\alpha
\] (7)

- \( k(\alpha) = -\frac{2\pi}{\lambda} (cos(\alpha), sin(\alpha))^T \) is the planar wave vector
- \( \lambda \) is the wavelength
- \( u_m, u_p \in \mathbb{R}^2 \) are the position vectors of the BS antennas in the 2D-coordinate system.
- the channel of user \( k \)

\[
h_k = U_k \Lambda_k^{1/2} w_k \quad (8)
\]

\[
R_g = U_g \Lambda_g U_g^H \quad (9)
\]

- \( \Lambda_k \) is an \( r_g \times r_g \) diagonal eigenvalues matrix with \( r_g \) being the rank
- \( U_k \in \mathbb{C}^{N_t \times r_g} \) being the set of eigenvectors corresponding to the nonzero eigenvalues. \( U_g \) is approximated by the columns of an MxM unitary DFT matrix.
- Two groups with disjoint AoA support have disjoint supports of their correlation Fourier Transform.
Goals & Adopted approach

Goals

- Training overhead scales with the number of users.
- Increasing connection density ⇒ Reduced SE as precious time-frequency resources need to be allocated to CSI estimation
- ⇒ Enable accurate CSI estimation in TDD massive MIMO with reduced training overhead.

Adopted approach\textsuperscript{14}

- Remove the obligation of pilots orthogonality within the cells ⇒ reduced training overhead.
- Allow the reuse of the same pilot sequences by users with semi-orthogonal spatial information.
- Construct copilot groups in each cell: Users in each group have minimum spatial signature overlapping with maximum coverage of the signal spatial basis ⇒ The possibility to distinguish copilot signals based on spatial information.
- Cross-cell pilot allocation: Leverage the previously formed copilot groups in a simplified training sequence allocation procedure.

\textsuperscript{14}S. Hajri and M. Assaad: A spatial basis coverage approach for uplink training and scheduling in Massive MIMO systems
System Model And Preliminaries

- A multi-cell multi-user massive MIMO system \((N_c \text{ cell})\)
- A large \(M\)-element ULA at each BS serving \(K\) users in each cell.
- TDD mode.
- Channel estimates are obtained through uplink training using orthogonal pilot sequences.
- LS channel estimation.
- The channel vector between user \(i\) in cell \(b\) and the BS of the \(r^{th}\) cell, \(g_{ib}^{[r]}\) is given by

\[
g_{ib}^{[r]} = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} a(\theta_{ib}^{[r,p]}) \gamma_{ib}^{[r,p]},
\]

- \(\gamma_{ib}^{[r,p]}\) represents the complex gain of the \(p^{th}\) ray \(\mathcal{CN}(0, \mu_{ib}^{[r,p]}^2)\) (\(\mu_{ib}^{[r,p]} = \text{average attenuation of the channel}\)).
- \(a(\theta_{ib}^{[r,p]}) \in \mathbb{C}^{M \times 1}\) is the array manifold vector (\(\theta_{ib}^{[r,p]} = \text{the direction of arrival (DOA) of the } p^{th} \text{ ray for } g_{ib}^{[r]}\)):

\[
a(\theta_{ib}^{[r,p]}) = [1, e^{j2\pi d \sin(\theta_{ib}^{[r,p]}/\lambda)}, \ldots, e^{j2\pi d \sin(\theta_{ib}^{[r,p]}/\lambda)(M-1)}],
\]

- As in [35], the incident angles of each user, with mean DOA \(\theta_{ib}^{[r]}\), are restrained in a narrow angular range \([\theta_{ib}^{[r]} - \Delta \theta_{ib}^{[r]}, \theta_{ib}^{[r]} + \Delta \theta_{ib}^{[r]}]\).
- \(a(\theta_{ib}^{[r,p]}), p = 1, \ldots, P, \forall i, b, r\) are mutually correlated (the covariance matrix of each channel \(g_{ib}^{[r]}\) \((R_{ib}^{[r]}\) possesses a low-rank property).
Spatial division multiplexing based user scheduling

- Exploiting the channel low-rank property in massive MIMO transmission strategies proved to provide non-negligible gains in performance for both TDD and FDD systems [35, 32, 44].
- Low-rank CSI estimation methods rely on different knowledge levels of the channel statistics.
- Owing to the slow varying nature of spatial information, efficient estimation can be obtained
  - UL spatial training [35].
  - DL spatial training followed by feedback [34].
- In this work, we need to quantify the importance of each DoA for each channel.
- For each user $i, b$, we construct a DFT-based decoding matrix $F_{ib}^{[b]}$

\[
F_{ib}^{[b]} = \{ f_s \in F, \frac{\| g_{ib}^{[b]*} f_s \|_2^2}{\text{Tr}(R_{ib}^{[b]})} \geq \alpha \}, \quad (12)
\]

$0 < \alpha < 1$ is a design parameter.
- $F_{ib}^{[b]}$ is used as the bases in which the user’s signal is detected and will henceforth be referred to as spatial signature of user $ib$. 
An Alternative Approach to Spatial User Grouping: A Spatial Basis Coverage Problem

- Previously proposed grouping approaches get the work done and provide considerable performance increase for both FDD and TDD modes.
- Nevertheless, they overlook a potentially important criterion:
  - Classical clustering approaches concentrate on the mutual distance between channels subspaces with little regard to the final coverage of independent streams.
- For a predefined training overhead of $\tau$:
  - Construct $\tau$ copilot user groups per cell.
  - The users in each copilot group provide
    - Minimum spatial signature overlapping
    - Maximum coverage of the signal spatial basis (DFT vectors)
- Two user grouping methods
  - Power agnostic
  - Power aware

Figure: Example of spatial basis coverage for $M = 100$
Power Agnostic Spatial Basis Coverage (1)

Power Agnostic Spatial Basis Coverage:

- Simpler.
- Faster.
- Better approximation ratio.

\[
\max \sum_{k=1}^{\tau} \sum_{b=1}^{C} \sum_{s=1}^{M} y_{s,b}^{[k]} \\
\text{subject to } \sum_{i} x_{\{i,b\}}^{[k]} \leq U_{b}^{[k]} \quad \forall k = 1...\tau, \quad \forall b = 1,...,N_{c}
\]

\[
\sum_{i,f_{s} \in F_{ib}} x_{\{i,b\}}^{[k]} \geq y_{s,b}^{[k]} \quad \forall k = 1...\tau, \quad \forall b = 1,...,N_{c},
\]

Note that \(U_{b}^{[k]}\) is a design parameter that defines the reuse factor of a given pilot sequence in each cell. Depending on the considered setting, \(U_{b}^{[k]}\) can be the same or differs from one cell to the other.

**Lemma**

The spatial basis coverage based copilot UE selection problem is NP-hard.
The proposed algorithm

- Two nested greedy phases:
  - In the upper phase, the algorithm produces $\tau$ maximum coverages of the DFT matrix vectors ($F$), in each cell.
  - Users are added to $C_k^{[b]}$ successively while selecting, at each iteration, the user with the spatial signature that cover a maximum of the uncovered DFT columns.

Theorem

The algorithm provides an $(1 - (\frac{\tau-1}{\tau})^\tau)(1 - \frac{1}{e})$-approximation of the optimal solution.
Power Aware Spatial Basis Coverage (1)

Power Aware Spatial Basis Coverage:
- More complex.
- Better in reducing interference.

\[
\max_Y \sum_{k=1}^{\tau} \sum_{b=1}^{N_c} \sum_{i \in \Gamma(b)} \sum_{f_s \in F} \zeta_{ib} y_{\{i,b\}}^{[s,k]} \\
\text{subject to } \sum_{i \in \Gamma(b), f_s \in F_{ib}} y_{\{i,b\}}^{[s,k]} \leq 1 \ \forall k = 1 \ldots \tau, \ \forall b = 1 \ldots N_c \\
\sum_{i \in \Gamma(b), f_s \in F_{ib}} x_{\{i,b\}}^{[k]} \geq y_{\{i,b\}}^{[s,k]} \ \forall k = 1 \ldots \tau, \ \forall b = 1 \ldots N_c, \\
\sum_{i \in \Gamma(b)} x_{\{i,b\}}^{[k]} \leq U_b^{[k]} \ \forall k = 1 \ldots \tau, \ \forall b = 1 \ldots N_c,
\]

Lemma

The spatial basis coverage based copilot UE selection problem is NP-hard.
Power Aware Spatial Basis Coverage (2)

The proposed algorithm

- Solving a generalized maximum coverage problem for each copilot group \( C_k^b, k = 1, \ldots, \tau, b = 1, \ldots, N_c \), successively.

- Two nested greedy phases:
  - In the upper phase, the maximum coverages \( C_k^b, k = 1, \ldots, \tau, b = 1, \ldots, N_c \) are computed successively in a greedy manner.
  - In the lower phase, users having the spatial signature with the highest density are greedily added, as long as the pilot reuse constraint is not violated.

Theorem

The proposed algorithm provides an \( (1 - \left( \frac{\tau - 1}{\tau} \right)^\tau) \frac{3 - e^{-2}}{2} - e^{-2} \)-approximation of the optimal solution.
Cross Cell Pilot allocation: a graphical approach

- Address the problems of intra-cell and inter-cell interference simultaneously, proves to be quite complex ⇒ The two problems are addressed successively.
- A major advantage of such division is the reduction of complexity.
- We propose a scheme in which pilot allocation is done such that high interference links are suppressed when spatial signature based receivers are used.
- infer inter-cell copilot interference from the spatial signatures of interference links.
- Spatial information of the interference links is required.

\[ F_{ib}^r = \{ f_s \in F, \frac{\| g_{ib}^r f_s \|^2}{Tr(R_{ib}^r)} \geq \alpha \} \]  
(15)
The first step to manage inter-cell copilot interference is to construct an appropriate interference graph.

We construct an undirected, weighted interference graph \( G(\mathcal{C}, \mathcal{E}) \).

Each edge in \( e_{C_b^j, C_l^k} \in \mathcal{E} \) represents an interference link and is associated with a given weight \( w_{C_b^j, C_l^k} \).

The weight of each edge quantifies the level of interference between two copilot groups.

We call upon hierarchical clustering where measuring distances between groups is commonly encountered.

\[
w_{C_b^j, C_l^k} = \min_{y \in C_b^j, z \in C_l^k} \left\{ \frac{1}{2} \| F_{yb} F_{yb}^\dagger - F_{zl} F_{zl}^\dagger \|_F^2 + \frac{1}{2} \| F_{yb} F_{yb}^\dagger - F_{zl} F_{zl}^\dagger \|_F^2 \right\},
\]

(16)
Graphical Modeling and proposed solution

An illustration of $G$ is presented in figure 2 for the case of $N_c = 2$ and $\tau = 2$.

The pilot allocation problem is equivalent to a $\text{MAX}-\tau$-$\text{CUT}$ problem [39].

The pilot allocation problem is NP-hard.

The proposed algorithm provides a $(1 - \frac{1}{\tau})$-approximation of the optimal cross-cell pilot allocation.
Numerical Results

Simulations setting:

- \( N_c = 4 \) hexagonal cells.
- Each cell has a radius 0.5 Km
- \( M = 128 \) equally spaced isotropic antennas.
- \( K = 25 \) per cell
- Number of paths \( P = 100 \)
- Path-loss coefficient = 3.5
- For each user \( i, b \), the angles of its rays \( \theta_{ib}^{[r,p]} \), \( p = 1, \ldots, P \) are uniformly distributed in the interval \( [\theta_{ib}^{[r]} - \Delta \theta_{ib}^{[r]}, \theta_{ib}^{[r]} + \Delta \theta_{ib}^{[r]}] \) where \( \Delta \theta_{ib}^{[r]} = \Delta = 4^\circ \).
- The coherence interval is set to \( T_s = 128 \) samples
The presented performances are attained with UL training overheads of $\tau$ and $U_b^{[k]} \times \tau$, for the proposed algorithms and the conventional approach respectively.
Sum spectral efficiency

**Figure:** Comparison of the achievable average SE with $\tau = 10$ and $U_b^{[k]} = 2$, $\forall k, b$

For an SNR of 0 dB, the power aware and the power agnostic spatial basis coverage approaches achieve 124.75 bits/Hz/s and 93.095 bits/Hz/s respectively. This represent gains of 41.983bits/Hz/s and 10.328 bits/Hz/s, respectively, in comparison with a conventional TDD massive MIMO system.
Spectral efficiency CDFs comparison

Figure: Comparison of CDFs of achievable SE for different $\tau$ and $U_b^{[k]}$ values with $SNR = 10$ dB

For $\tau = 10$ and $U_b^{[k]} = 2 \forall k, b$, the power aware and the power agnostic spatial basis coverage approaches achieves 5% outage rate around 132 bit/s/Hz and 113 bit/s/Hz, respectively.

For $\tau = 5$ and $U_b^{[k]} = 4 \forall k, b$, 5% outage rate is attained around 186 bit/s/Hz and 126 bit/s/Hz.
Performance of cross-cell pilot allocation (max cut problem)

Figure: Comparison of CDFs of achievable SE for $\tau = 4$, $U_b^{[k]} = 5$ and with $SNR = 10$ dB

While the power aware spatial basis cover approach achieves 5% outage rate around 168 bit/s/Hz, the combination with the max-$\tau$-cut pilot assignment algorithm achieves 5% outage rate around 191 bit/s/Hz.
The goal in this work was to enable high connection density in TDD massive MIMO systems.

This requires, perforce, an adapted CSI acquisition scheme for high density scenarios.

Exploit spatial diversity in order to reduce the needed training overhead.

Addressing both intra and inter-cell copilot interference:
- Copilot user grouping based on spatial information.
- Cross-cell pilot allocation.

Efficient low complexity algorithms for uplink training optimization.

This scheme allows more reuse of pilots in the network which increases the density of users (MTC) served by the system.

The grouping is based on the second order statistics of the users' channel: what happens when if the users are mobile with different speed? what about the traffic patterns (all users in the same group may not be active at the same time)?
Section 4

Coherence time based user grouping in massive MIMO
Observations & Adopted approach

Observations & Intuition

▶ Current wireless systems assume the same time slot duration for all devices, regardless of the fact that users are subject to heterogeneous Doppler spreads.
▶ Uplink training is performed at each slot for all scheduled users.
▶ The wireless channels do not change at the same rate ⇔ Channel aging depends on the experienced Doppler spread.
▶ The network is going to spend precious resources on estimating somehow redundant information.
▶ We do not need to send uplink reference signal at each slot for all users as channels change at different rates \(^{15}\).

Adopted approach \(^{16}\)

▶ Exploit a degree of freedom that was previously neglected: \(\textit{CSI estimation periodicity}: \) Addressing the uplink training bottleneck through an adaptive TDD frame structure.
▶ Define the needed training resources dynamically depending on the \(\textit{Doppler spread}\) (equivalently, the channel \(\textit{coherence time}\)) while taking into consideration user mobility.
▶ This is in line with the emerging concept of Dynamic TDD.


\(^{16}\)S. Hajri, M. Larranaga, M. Assaad, ”Heterogeneous Doppler Spread-based CSI Estimation Planning for TDD Massive MIMO,” IEEE Transactions on Wireless Communications, 2018
Observations & Adopted approach

Observations & Intuition

▶ Current wireless systems assume the same time slot duration for all devices, regardless of the fact that users are subject to heterogeneous Doppler spreads.
▶ Uplink training is performed at each slot for all scheduled users.
▶ The wireless channels do not change at the same rate ⇔ Channel aging depends on the experienced Doppler spread.
▶ ⇒ The network is going to spend precious resources on estimating somehow redundant information.
▶ We do not need to send uplink reference signal at each slot for all users as channels change at different rates \(^{15}\).

Adopted approach \(^{16}\)

▶ Exploit a degree of freedom that was previously neglected:

*CSI estimation periodicity*: Addressing the uplink training bottleneck through an adaptive TDD frame structure.

▶ Define the needed training resources dynamically depending on the *Doppler spread* (equivalently, the channel *coherence time*) while taking into consideration user mobility.

▶ This is in line with the emerging concept of Dynamic TDD.


System Model And Preliminaries

- A multi-cell multi-user massive MIMO system (C cell)
- Each Base station has $M$ antennas and serves $K$ users.
- The network operates in time division duplexing mode.
- Channel estimates are obtained through uplink training using orthogonal pilot sequences.
- MMSE channel estimation.
- Capturing the impact of channel aging: a stationary ergodic Gauss-Markov block fading regular process.
- Linear receivers: MRC & ZF.
A new TDD protocol for adaptive uplink training (1)

- CSI estimation periodicity as a DoF ⇒ A need for a time varying TDD frame (Dynamic TDD).
- Important question: On which criterion should we define the dynamic TDD frame?
- Intuition:
  - Taking into account the impact of channel aging & asymptotic regime:
    \[
    \tilde{R}_u^\infty (d_g, g = 1, \ldots, N_g) \geq \sum_{l=1}^{C} \sum_{g=1}^{N_g} \left(1 - \frac{\tau}{T_s}\right) \log \left(1 + \frac{\beta[l]^2 \rho[l]^2 d_g}{\sum_{c\neq l} C_c \rho[c]^2 d_g \beta[c]^2} \right),
    \]
    (17)
  - The SINR decays slower as a function of CSI age if copilot users have similar channel autocorrelation coefficients ⇒ similar Doppler spreads & coherence times.

- Proposed solution:
  - Define a new TDD protocol that enables dynamic uplink training resource allocation.
  - Group users based on their channel coherence times and schedule them for training accordingly.
  - Allow the network to learn an optimized long term training policy taking into account large-scale statistics evolution.
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. The BSs estimate $\beta_{kc}^{[j]}$ and $\rho_{kc}^{[j]}$ for all $k = 1, \ldots, K$, and $c, j = 1, \ldots, C$. All coefficients are then fed back to a central processing unit (CPU).
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. Statistics estimation

2. CPU clusters users according to their autocorrelation coefficients using $K$-mean algorithm [?$]$. We choose the number of clusters $N_c$ according to

$$N_c = \left\lceil \frac{T_{\text{max}}}{T_s} \rightceil,$$

(18)

where $T_{\text{max}}$ represents the maximum coherence time of the users’ channels.
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. Statistics estimation
2. Coherence time-based user clustering
3. CPU allocates all users in the network to $N_g$ copilot user groups. Each group contains at maximum $C$ users from the same channel autocorrelation based cluster and from different cells.
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. Statistics estimation
2. Coherence time-based user clustering
3. Copilot groups formation
4. Uplink training ($\tau$ copilot groups at maximum)
5. Channel estimation + decoding
6. Precoding + Downlink transmission
7. Transmission and Control of devices in dense networks

4. At each coherence slot, the network schedules at maximum $\tau$ copilot user groups for uplink training synchronously.
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. **Statistics estimation**
2. **Coherence time-based user clustering**
3. **Copilot groups formation**
4. **Uplink training ($\tau$ copilot groups at maximum)**
5. All $N_g$ copilot groups transmit their uplink signal in a synchronous manner.
6. **Uplink data transmission**

Diagram:

- Hexagonal region representing the network area.
- Various points and lines indicating user distribution and signal transmission.
- Hexagonal clusters with different colors representing different groups or areas of operation.
- Arrows indicating the flow of the protocol steps.
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. Statistics estimation
2. Coherence time-based user clustering
3. Copilot groups formation
4. Uplink training (τ copilot groups at maximum)
5. Uplink data transmission
6. Channel estimation + decoding
7. Downlink transmission

6. The BSs process the received pilot signal and estimates the channels of the active users during uplink training using MMSE estimators. The BSs decode and precode the uplink and downlink data signals, respectively, using the last estimated version of CSI.
A new TDD protocol for adaptive uplink training (2)

The proposed TDD protocol consists of the following seven steps:

1. **Statistics estimation**
2. **Coherence time-based user clustering**
3. **Copilot groups formation**
4. **Uplink training ($\tau$ copilot groups at maximum)**
5. **Uplink data transmission**
6. **Channel estimation + decoding**
7. **Precoding + Downlink transmission**

7. All BSs synchronously transmit downlink data signals to the $N_g$ copilot groups.
Theorem

For $N_g$ active copilot groups during uplink transmission, $\tau$ of which are scheduled for uplink training and using a MRC receiver $u_{gl}(t)$ that is based on the latest available CSI estimates of each user $g, l$, the average achievable spectral efficiency in the uplink $\bar{R}_u^\text{MRC}$ is lower bounded by:

$$
\bar{R}_u^\text{MRC} \geq \sum_{l=1}^{C} \sum_{g=1}^{N_g} \left(1 - \frac{\tau}{T_s}\right) \log \left(1 + \frac{(M - 1)\beta_{gl}^{[l]^2} \rho_{gl}^{[l]^2dg}}{(M - 1) \times I_{gl}^p + I_{gl}^n}\right),
$$

(19)

where $d_g$, $g = 1 \ldots N_g$ represents the CSI delays of users in copilot groups $g, g = 1, \ldots, N_g$. $I_{gl}^p$ and $I_{gl}^n$ are given by:

$$
I_{gl}^p = \sum_{c \neq l}^{C} \rho_{gc}^{[l]2dg} \beta_{gc}^{[l]^2},
$$

(20)

$$
I_{gl}^n = \left(\sum_{c=1}^{C} \sum_{k \neq g}^{N_g} \beta_{kc}^{[l]} \right) + \sum_{c=1}^{C} \left(\beta_{gc}^{[l]} - \rho_{gc}^{[l]2dg} \frac{\beta_{gc}^{[l]^2}}{\frac{1}{P_p} + \sum_{b=1}^{C} \beta_{gb}^{[l]} + \frac{1}{P_u}}\right) \times \left(\frac{1}{P_p} + \sum_{b=1}^{C} \beta_{gb}^{[l]}\right).
$$
A new TDD protocol for adaptive uplink training: Performance (2)

Figure: Comparison of the CDFs of spectral efficiency

For 150 antennas, the gain in the 5%-outage rate attains 8 bit/s/Hz
A new TDD protocol for adaptive uplink training: Performance (3)

Figure: Comparison of Spectral efficiency for varying values of M with ZF and MRC receivers, respectively

For $M = 105$, the proposed training scheme achieves SE gains of 12.2 bit/s/Hz (ZF) and 3.6 bit/s/Hz (MRC).
A new TDD protocol for adaptive uplink training: When less is better?

Theorem

In the asymptotic regime (M grows large), with $\bar{\rho}_g^{[\text{min}]}$ and $\bar{\rho}_g^{[\text{max}]}$ denoting, respectively, the minimum and maximum autocorrelation coefficients in copilot group $g$, $g = 1, ..., N_G$, the proposed training framework enables to improve the SE of each user when (21) is satisfied

\[
\left( \frac{\bar{\rho}_g^{[\text{min}]}^2}{\bar{\rho}_g^{[\text{max}]}^2} \right)^{d_g} \geq \left( 1 + \frac{\text{SINR}_{g,l}^{[\infty]}}{\text{SINR}_{g,l}^{[\infty]}} \right) \frac{T_s - N_G}{T_s - \tau} - 1, \quad \text{with} \quad \text{SINR}_{g,l}^{[\infty]} = \frac{\beta_{g,l}^2}{\sum_{c \neq l} \beta_{g,c}^2} \tag{21}
\]

- SE is improved as long as the SINR degradation is compensated for by the spared resources from training.
- A high ratio between the minimum and maximum autocorrelation coefficients is required (tighter as the CSI time offset increases).
- Improvement in SE, even with random pilot allocation.
- One can do better if a coherence time adaptive scheduling for uplink training is implemented.
Learn the best training strategy: A two-time scale decision process (1)

- The network learns an optimized uplink training policy taking into account user mobility.
- User locations change slower than the actual wireless channel ⇒ A two-time scale decision process..

- Slow time scale:
  - Costly positioning (OTDOA [6], GPS [4]) \(\iff\) Need to limit the positioning signaling cost.
  - In every decision epoch, the users from \(U_{\text{max}}\) copilot groups can feedback their positions (with \(U_{\text{max}} < N_g\)). The positions of the rest will be inferred from previous estimations.

- Fast time scale:
  - Learning the best training strategy for large but finite time periods \(H\).
  - When a given copilot group is not scheduled for training, the last available CSI estimates is used.
The network learns an optimized uplink training policy taking into account user mobility.

User locations change slower than the actual wireless channel ⇒ A two-time scale decision process.

Slow time scale:
- Costly positioning (OTDOA [6], GPS [4]) ⇐ Need to limit the positioning signaling cost.
- In every decision epoch, the users from $U_{\text{max}}$ copilot groups can feedback their positions (with $U_{\text{max}} < N_g$). The positions of the rest will be inferred from previous estimations.

Fast time scale:
- Learning the best training strategy for large but finite time periods $H$.
- When a given copilot group is not scheduled for training, the last available CSI estimates is used.
Learn the best training strategy: A two-time scale decision process (1)

- The network learns an optimized uplink training policy taking into account user mobility.

- User locations change slower than the actual wireless channel ⇒ **A two-time scale decision process.**

  **Slow time scale:**
  - Costly positioning (OTDOA [6], GPS [4]) ⇐ Need to limit the positioning signaling cost.
  - In every decision epoch, the users from $U_{\text{max}}$ copilot groups can feedback their positions (with $U_{\text{max}} < N_g$). The positions of the rest will be inferred from previous estimations.

  **Fast time scale:**
  - Learning the best training strategy for large but finite time periods $H$.
  - When a given copilot group is not scheduled for training, the last available CSI estimates is used.
Learn the best training strategy: A two-time scale decision process (2)

In the Fast time scale

State space: $\vec{d} = (d_1, \ldots, d_{N_g}) \in X$

Action space: $\vec{a} = (a_1, \ldots, a_{N_g}) \in \mathcal{A} = \{0, 1\}^{N_g}$

In the Slow time scale

State space: $x(n) = (\vec{b}_1, \ldots, \vec{b}_{N_g}) \in X$

Action space: $\vec{u}(n) \in \mathcal{A} = \{0, 1\}^{N_g}$
Learn the best training strategy: A two-time scale decision process (2)

$x(0) = (\vec{b}_1(0), \ldots, \vec{b}_{N_g}(0))$

$x(1) = (\vec{b}_1(1), \ldots, \vec{b}_{N_g}(1))$

In the Slow time scale

State space: $x(n) = (\vec{b}_1, \ldots, \vec{b}_{N_g}) \in \mathcal{X}$

Action space: $\vec{u}(n) \in \mathcal{A} = \{0, 1\}^{N_g}$

In the Fast time scale

$t_0 \quad t_1 \quad t_{H-1} \quad t_H \quad t_{H+1} \quad t_{2H-1} \quad t_{2H}$
Learn the best training strategy: A two-time scale decision process (2)

\[ x(0) = (\vec{b}_1(0), \ldots, \vec{b}_{N_g}(0)) \quad x(1) = (\vec{b}_1(1), \ldots, \vec{b}_{N_g}(1)) \]

Slow time scale

\[ \vec{u}(0) \]

n=0 \quad n=1 \quad n=2

In the Fast time scale

State space: \( \vec{d} = (d_1, \ldots, d_{N_g}) \in X \)

Action space: \( \vec{a} = (a_1, \ldots, a_{N_g}) \in A = \{0, 1\}^{N_g} \)

Fast time scale

\[ \vec{d}(t_0) \quad \vec{d}(t_1) \]

\[ \vec{a}(t_0) \]

\[ t_0 \quad t_1 \quad t_{H-1} \quad t_H \quad t_{H+1} \quad t_{2H-1} \quad t_{2H} \]
Learn the best training strategy: A two-time scale decision process (2)

\[ x(0) = (\vec{b}_1(0), \ldots, \vec{b}_{N_g}(0)) \]
\[ x(1) = (\vec{b}_1(1), \ldots, \vec{b}_{N_g}(1)) \]

Slow time scale

Fast time scale

\[ \vec{a}(t_0) \quad \vec{a}(t_1) \quad \vec{a}(t_{H-1}) \]
\[ \vec{a}(t_H) \quad \vec{a}(t_{2H-1}) \]

\[ \vec{d}(t_0) \quad \vec{d}(t_1) \quad \vec{d}(t_H) \quad \vec{d}(t_{H-1}) \]
A two-time scale decision process: A two-time scale decision process

(3)

Problem formulation:

- Denote $\Phi^{up}$ the set of all possible stationary decision rules in the upper level, such that $\pi^{up} \in \Phi^{up}$, $\pi^{up} : X \times X \to A$. Consequently, the objective is to find $\pi^{up} \in \Phi^{up}$ and $\pi^{low} \in \Phi^{low}$ such that

$$\max_{\pi^{up} \in \Phi^{up}} \max_{\pi^{low} \in \Phi^{low}} \lim_{Z \to \infty} \frac{1}{Z} \sum_{n=0}^{Z-1} \mathbb{E} \left( R^{up}(\vec{d}(t_{nH}), \pi^{low}, x(n), \pi^{up}(\vec{d}(t_{nH}), x(n))) \right).$$

(22)

- The reward in the upper level:

$$R^{up}(\vec{d}, \vec{a}^{low}, x(n), \vec{u}(n)) = \sum_{t=t_{nH}}^{t_{(n+1)H-1}} R^{low}(\vec{d}(t), \phi^{low}_{t}(\vec{d}(t), x(n), \vec{u}(n)), x(n), \vec{u}(n)),$$

(23)

- The reward in the lower level:

$$R^{low}(\vec{d}(t), \vec{a}(t), x, \vec{u}) = \sum_{g=1}^{Ng} \sum_{l=1}^{C} \left( 1 - \frac{1}{T_{s}} \sum_{i=1}^{Ng} a_{i}(t) \right) \log \left( 1 + \text{SINR}_{gl}^{MRC}(\vec{d}(t), x, \vec{u}) \right),$$

(24)
A two-time scale decision process: POMDP (1)

The latter problem is a POMDP [5].

- The slow time scale sequential decision process is a POMDP with a reward that depends on the fast time scale decision process.
- Bellman’s optimality equations follows (for $0 < \alpha < 1$)

\[
V(\vec{d}, x) = \max_{\vec{u} \in \mathcal{A}} \left( \max_{\phi_{x,\vec{u}} \in \Phi_{low}} \left\{ R^{up}(\vec{d}, \phi_{x,\vec{u}}^\text{low}, x, \vec{u}) + \alpha \sum_{y \in \mathcal{X}} \mathbb{P}^{up}(y|x, \vec{u}) V(\vec{d}^{\phi_{x,\vec{u}}^\text{low}, y}) \right\} \right).
\]

(25)

Assume stationarity and define

\[
R^{\text{max}}(x, \vec{u}) = \max_{\phi_{x,\vec{u}}^\text{low} \in \Phi_{low}} \left\{ R^{up}(\phi_{x,\vec{u}}^\text{low}, x, \vec{u}) \right\},
\]

(26)

⇒ We obtain a standard one-time scale POMDP, and its optimality equation reduces to

\[
V(x) = \max_{\vec{u} \in \mathcal{A}} \left( R^{\text{max}}(x, \vec{u}) + \alpha \sum_{y \in \mathcal{X}} \mathbb{P}^{up}(y|x, \vec{u}) V(y) \right).
\]

(27)
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard
  - Belief-state monitoring is infeasible for large problems.

- Decompose problem
- Find two decision policies associated each with a time scale.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (2) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

- Decompose problem
- Find two decision policies associated each with a time scale.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

- Decompose problem
- Find two decision policies associated each with a time scale.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

- Decompose problem
- Find two decision policies associated each with a time scale.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

Solution:

- Decompose problem
  - Find two decision policies associated each with a time scale.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

Solution:

- Decompose problem
- Find two decision policies associated each with a time scale.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

Solution:

- Decompose problem
- Find two decision policies associated each with a time scale.
  - Fast time scale ↦ a finite horizon training policy.

- The combination of the latter will provide a solution to (22).
A two-time scale decision process: POMDP (2)

- One can solve using Bellman equations.
- Solving a two time scale POMDP (22) ⇒ Too complex to solve directly.
  - Complexity of exact algorithms grows exponentially with the number of state variables.
  - Finding the optimal policy is PSPACE-hard.
  - Belief-state monitoring is infeasible for large problems.

Solution:

- Decompose problem
- Find two decision policies associated each with a time scale.
  - Fast time scale \(\mapsto\) a finite horizon training policy.
  - Slow time scale \(\mapsto\) an infinite horizon position estimation policy.

- The combination of the latter will provide a solution to (22).
Slow time scale: adapting to user mobility (1)
We now tackle the infinite horizon positioning problem on the slow time scale.

- Solving the decision problem directly is too complex (belief-state monitoring [9]).
- A more practical approach ⇒ Abandon policy optimality for the sake of convergence speed.
- Solve a POMDP by exploiting its underlying Markov Decision process (MDP)[?].
- The agent’s confusion is ignored and most likely state (MLS) is assumed.

A belief space → A smaller state space (Reduced complexity)
Slow time scale: adapting to user mobility (1)

We now tackle the infinite horizon positioning problem on the slow time scale.

- Solving the decision problem directly is too complex (belief-state monitoring [9]).
- A more practical approach ⇒ Abandon policy optimality for the sake of convergence speed.
- Solve a POMDP by exploiting its underlying Markov Decision process (MDP)[?].
- The agent's confusion is ignored and most likely state (MLS) is assumed.

**Figure:** Positioning problem

A belief space $\rightarrow$ A smaller state space (Reduced complexity)
Slow time scale: adapting to user mobility (1)
We now tackle the infinite horizon positioning problem on the slow time scale.

- Solving the decision problem directly is too complex (belief-state monitoring [9]).
- A more practical approach ⇒ Abandon policy optimality for the sake of convergence speed.
- Solve a POMDP by exploiting its underlying Markov Decision process (MDP)[?].
- The agent’s confusion is ignored and most likely state (MLS) is assumed.

A belief space → A smaller state space (Reduced complexity)
Numerical Results

The gain attains 41.99 bit/s/Hz (DP) and 38.7 bit/s/Hz (AP) at the final stage of the optimization horizon $H$. 

Figure: Gain for different lower level algorithms
Take-away points

- Massive MIMO offers high spatial multiplexing gain but suffers from pilot contamination.
- Novel uplink training paradigms for TDD massive MIMO are required: e.g. Adapt training periodicity based on the real channel conditions of each user.
- Leverage the degree of freedom of CSI estimation periodicity: Planning.
- A two time scale learning problem: Learn an efficient uplink training strategy
  - Adapt to user mobility.
  - Improve the cumulative spectral efficiency with less training.
- Other methods that use the second order statistics of the users’ channels have been investigated in the literature.
- Leverage the evolution of the correlation between the wireless channel and the estimated CSI in order to improve SE.
Section 5

Multiple Access Techniques
Historical review

- Orthogonal Multiple Access (OMA) is widely used in Wireless networks (2G, 3G, 4G)
- TDMA, CDMA (WCDMA), OFDM/OFDMA
- avoid interference between user signals.
- improve the network performance and achieve Multiuser diversity gain (max weight scheduling, PF, etc): a resource is allocated to a user by taking into account the radio channel conditions of the users.
- Efficient resource allocation
- CSI/CQI feedback is required, additional control information: signaling overhead
- TDMA, OFDMA: synchronization of the devices
- Feedback delay, low user mobility
Multiple Access for 5G (1)

- OFDMA is adopted for eMBB and URLLC services.
- Flexible frame structure: Scalable TTI, Scalable subcarrier spacing ($15 \text{ kHz} \times 2^n$)
- Adapt to different contexts: Doppler Spread, outdoor, macro and small cell, etc.
- FFT size must scale so that processing complexity does not increase exponentially for large bandwidths.

**Figure: 5G Frame structure**
## Multiple Access for 5G (2)

<table>
<thead>
<tr>
<th>Carrier Spacing</th>
<th>1 ms</th>
<th>1 slot / 1 ms</th>
<th>1 slot / 0.5 ms</th>
<th>1 slot / 0.25 ms</th>
<th>1 slot / 0.125 ms</th>
<th>1 slot / 0.0625 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 KHz</td>
<td>1</td>
<td>1 slot 0</td>
<td>1 slot 0</td>
<td>1 slot 0</td>
<td>1 slot 0</td>
<td>1 slot 0</td>
</tr>
<tr>
<td>30 KHz</td>
<td>2</td>
<td>1 slot 0</td>
<td>1 slot 0, slot 1</td>
<td>1 slot 0, slot 1</td>
<td>1 slot 0, slot 1</td>
<td>1 slot 0, slot 1</td>
</tr>
<tr>
<td>60 KHz</td>
<td>4</td>
<td>1 slot 0</td>
<td>1 slot 0, slot 1, slot 2</td>
<td>1 slot 0, slot 1, slot 2</td>
<td>1 slot 0, slot 1, slot 2</td>
<td>1 slot 0, slot 1, slot 2</td>
</tr>
<tr>
<td>120 KHz</td>
<td>8</td>
<td>1 slot 0</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
</tr>
<tr>
<td>240 KHz</td>
<td>16</td>
<td>1 slot 0</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
<td>1 slot 0, slot 1, slot 2, slot 3</td>
</tr>
</tbody>
</table>

**Figure:** 5G Frame structure

---

Overview of 5G  Massive MIMO  Spatial division multiplexing based user scheduling  Coherence time based user grouping in massive MIMO  Multiple Access for 5G (3)

Figure: 5G Frame structure

Source: Ericsson
What about MTC traffic?

- Short packets, high density of devices
- Non-Orthogonal Multiple Access (NOMA) is a key technique.
- From Multi user information theory (OMA) is not optimal in general and gains may be obtained using (NOMA).
- Many papers have been published on NOMA and several variants have been proposed for 5G standardization by the 3GPP.
- Single Carrier and multicarrier NOMA
Basic Principle of Power domain NOMA (1)

- Consider a 2-user channel where User 1 has a strong power $P_1$ and User 2 has a weak power $P_2$.
- SIC receiver: The signal of User 1 can be detected in presence of the interference from User-2 signal, then it is subtracted from the received signal to detect the weak User-2 signal without interference.
- In the case of an AWGN channel of bandwidth $W = 1\, Hz$, the capacity in bits per channel use is:
  \[ R_1 = \log_2 \left( 1 + \frac{P_1}{P_2 + N_0} \right) \]  
  \[ R_2 = \log_2 \left( 1 + \frac{P_2}{N_0} \right) \]  
- The total capacity is $(P = P_1 + P_2)$:
  \[ R = R_1 + R_2 = \log_2 \left( \left( 1 + \frac{P_1}{P_2 + N_0} \right) \left( 1 + \frac{P_2}{N_0} \right) \right) = \log_2 \left( 1 + \frac{P}{N_0} \right) \]  
- The capacity is similar to that of a single-user channel with the same total power.
What about OMA?

- Consider an OFDMA scheme with 2 users. $P_1 = \alpha P$ and $P_2 = (1 - \alpha) P$, with $0 \leq \alpha \leq 1$.
- The signal power being uniformly distributed over the N carriers composing the OFDMA signal, we have $W_1 = \alpha W$ and $W_2 = (1 - \alpha) W$. The total capacity is:

$$R = R_1 + R_2 = \alpha \log_2 \left(1 + \frac{P}{N_0}\right) + (1 - \alpha) \log_2 \left(1 + \frac{P}{N_0}\right)$$  \hspace{1cm} (31)

- The capacity is identical to that of NOMA
- The difference appears when the signal of one user is attenuated (e.g. User-2 signal is attenuated by 6dB):

$$R_{\text{OFDMA}} = \alpha \log_2 \left(1 + \frac{P}{N_0}\right) + (1 - \alpha) \log_2 \left(1 + \frac{P/4}{N_0}\right)$$  \hspace{1cm} (32)

$$R_{\text{NOMA}} = \log_2 \left(1 + \frac{\alpha P}{(1 - \alpha)P/4 + N_0}\right) + \log_2 \left(1 + \frac{(1 - \alpha)P/4}{N_0}\right)$$

$$= \log_2 \left(1 + \frac{(1 + 3\alpha)P}{4N_0}\right).$$  \hspace{1cm} (33)

- $\alpha = 0.8$ and $P/N_0 = 15$, $R_{\text{OFDMA}} = 3.65$ and $R_{\text{NOMA}} = 3.78$. 


NOMA Uplink

Figure: Example of NOMA Uplink
Recent NOMA literature

- The power imbalance requires "user pairing" and/or different power allocation between the users. NP-hard in some cases.
- Iterative SIC receiver may be required, complexity issues.
- Several schemes have been proposed recently [61]-[67]
  - Scrambling Based: Use different scrambling sequences to distinguish different devices. Ex: RSMA, etc.
  - Spreading Based: Use different spreading codes to distinguish different devices. Ex: SCMA, PDMA, etc.
  - Interleaving based: Use different interleavers to distinguish different users. Ex: IDMA, IGMA, etc.

---

**Figure:** Example of SCMA [65]
Early NOMA literature

Although never mentioned in the recent NOMA literature, the foundation of NOMA actually dates back to the year 2000 when a series of papers introduced the concept of multiple access using two sets of orthogonal signal waveforms.

New NOMA scheme for 5G

- OFDMA for eMBB with high bit rate.
- How to transmit at low bit rate high amount of short packets using the same bandwidth of eMBB.
- Conflicting metrics!
- One possible solution: mapping between two waveforms OFDMA and MC-CDMA.


NOMA for 5G

Figure: An illustration of the combined OFDMA/MC-CDMA scheme, where the OFDMA signal set is used in full and the MC-CDMA signal set is used partially.
Basic Characteristics

- This scheme fully avoids the power imbalance requirement that is present in power-domain NOMA.
- The power imbalance that is required for reliable detection is an inherent property of the signal design.
- OFDMA and MC-CDMA can be assigned to users with different profiles and service requirements.
- Since it uses OFDMA as the primary signal set and MC-CDMA as a secondary signal set, this NOMA technique can be viewed as a convenient extension of OFDMA rather than a purely competing technology.
- SIC Receiver: interestingly, iterative SIC can provide significant gain
Overview of 5G Massive MIMO Spatial division multiplexing based user scheduling Coherence time based user grouping in massive MIMO Multiple Access

Receiver

Figure: Receiver structure
Figure: BER Performance of QPSK, N=256, overload= 20%
BER Performance

Figure: BER Performance of 16-QAM, N=256, overload= 20%
Density of the devices

- Results show that the proposed NOMA scheme provides a channel overload of 25% (one MTC user spread over 4 subcarriers).

- This implies that for a total bandwidth of 200MHz, and assuming a subcarrier spacing of 15KHz (which the standard spacing in 4G and 5G systems), the total number of subcarriers is 12000 (12 subcarriers per Resource Block and 100 RBs per 20MHz). The number of mMTC devices served simultaneously per a BS is then 3000.

- Assuming an MTC average activity of 1% and small cell deployment with Inter site distance of 40m, the density of connected devices without taking into account the inter cell interference is roughly 24M/Km2.

- Even when inter cell interference is taken into account, the aforementioned result shows that the proposed scheme is useful for mMTC.

- Improvement is required when the activity of the devices is higher.
Take-away points

- The foundation of NOMA actually goes back to the year 2000 when the concept of multiple access using two sets of orthogonal waveforms was introduced.
- This NOMA concept fully avoids the power imbalance issue and uses iterative interference cancellation.
- In this approach, NOMA appears as a natural extension of OFDMA rather than a purely competing technology.
- As M increases beyond a critical value, hard-decision detectors come to a performance limit, and soft-decision iterative interference cancellation becomes necessary to approach the performance of interference-free transmission.
- The channel overload factor with respect to OFDMA can be 20 to 25%.
- Iterative receiver: complexity issues! but improvement is possible.
- How to improve further the performance when several services use the same bandwidth?
- What about traffic patterns?
Section 6

Random Access for multiservice scenarios
- OMA requires signaling overhead to achieve multiuser diversity gain
- MTC devices have sparse activity and usually transmit short packets.
- Various QoS/QoE requirements (e.g. delay)
- Distributed Scheduling
- Possible 5G Scenarios: Underserved area or Megacities.
- Energy efficient solution with simple transmission/receiving schemes.
- Random Access solutions?
- Several strategies have been proposed for LTE and 5G [68]-[70].
Context

Figure: Example of System Model

---

21 Nour Kouzayha, Mona Jaber, and Zaher Dawy, "M2M Data Aggregation over Cellular Networks: Signaling-Delay Trade-offs", IEEE Globecom 2014 Workshop - Broadband Wireless Access
Carrier Sense Multiple Access (CSMA) is widely used as a Medium Access Control (MAC) in wireless networks due to its simplicity and distributed nature.

Throughput optimal CSMA has been proposed \(^{22}\).

New technologies emerged where prolonged battery life is crucial such as environment and industrial monitoring.

Combining throughput optimality and energy efficiency is therefore important.

Several solutions have been proposed in the literature e.g. WiFi, etc.

Power consumption is an important issue.

Heuristic solutions exist [71]-[75].

Optimality? Energy throughput trade-off?\(^{23}\)

---


\(^{23}\) A. Maatouk, M. Assaad and A. Ephremides, "Energy Efficient and Throughput Optimal CSMA Scheme", IEEE/ACM Transactions on Networking, revised, 2018
CSMA/CA - IEEE 802.11

- **802.11 sender**
  - Sensing: if channel idle for time period DIFS then transmit the frame
  - if channel busy:
    - start random backoff time
    - timer counts down while channel idle
    - transmit when the timer expires
    - if NACK, increases random backoff timer (binary exponential) and repeat.

- **Receiver**: if the frame is received correctly then return an ACK after SIFS time period.
CSMA/CA

Figure: Example of CSMA procedure

source: https://www.researchgate.net/publication/320344383_Cooperation_Tecniques_between_LTE_i_n_Unlicensed_Spectrum_and_Wi-Fi_towards_Fair_Spectral_Efficiency/figures?lo=1
Method 1b: CSMA/CA with RTS-CTS

- **CSMA/CA**: explicit channel reservation
  - **sender**: send RTS (20 bytes)
  - **receiver**: reply with CTS (16 bytes)

- **CTS** reserves channel for sender, notifying (possibly hidden) terminals

**Figure**: 5G Frame structure

---

25Source: https://slideplayer.com/slide/6673427/
System Model - I

Idealized CSMA conditions:

- No hidden nodes
- No sensing delay

These assumptions simplify the model and serve as a starting point. To reduce power consumption, links can be in one of the following states:

- **SLEEP** state: in this state, power consumption is minor and no sensing of the environment takes place
- **AWAKE** state: the adaptive CSMA sensing and back-off mechanisms take place

The decision to either wake-up/sleep is dictated by an appropriate timer.
System Model - II

When a link decides to sleep, it picks an exponentially distributed wake-up time with mean $1/W_k$ after which it wakes up. Once the link is awake, it picks an exponentially chosen sleep timer with mean $1/S_k$ after which it goes back to sleep. When the link is awake, the following takes place:

- An exponentially distributed back-off timer with mean $1/R_k$ is picked
- Continuous sensing of the channel takes place and whenever the channel is sensed idle, the back-off timer runs otherwise it is frozen. In both cases, the sleep timer keeps running
- Once the back-off timer runs out, the channel is captured by the link and the sleep timer is frozen
- Packets transmission time is assumed to be exponentially distributed with an average channel holding time $1/H_k$
Markov Chain

Let us consider the 2D continuous time stochastic process \( \{(A(t), X(t)) : t \geq 0\} \) where \( A(t) \in \{0, 1\}^K \) and \( X(t) \in \{0, 1\}^K \) denote the awake and transmission states of the network at time instant \( t \) respectively. By adopting our proposed scheme, \( \{(A(t), X(t)) : t \geq 0\} \) is a Markovian process.

![2D Markov Chain](image)

**Figure: 2D Markov Chain**

---

26 A. Maatouk, M. Assaad and A. Ephremides, "Energy Efficient and Throughput Optimal CSMA Scheme", IEEE/ACM Transactions on Networking, revised, 2018
Proposed CSMA scheme

Figure: Time-line of the proposed CSMA scheme
Stationary Distribution

We define the *transmission aggressiveness* and the *waking-up aggressiveness* as $r \in \mathbb{R}^K$ and $\rho \in \mathbb{R}^K$ where $r_k = \log\left(\frac{R_k}{H_k}\right)$ and $\rho_k = \log\left(\frac{W_k}{S_k}\right)$ respectively.

**Theorem**

The chain admits $\pi(a^j, x^i; r, \rho)$ as stationary distribution where:

$$
\pi(a^j, x^i; r, \rho) = \frac{\exp\left(\sum_{k=1}^{K} a_{jk}^i \rho_k\right) \exp\left(\sum_{k=1}^{K} x_{ik}^i r_k\right)}{C(r, \rho)}
$$

(34)

and $C(r, \rho)$ is a normalization factor:

$$
C(r, \rho) = \sum_{j=1}^{2^K} \exp\left(\sum_{k=1}^{K} a_{jk}^i \rho_k\right) \sum_{i=1}^{\left|I_j\right|} \exp\left(\sum_{k=1}^{K} x_{ik}^i r_k\right)
$$

(35)
Optimality analysis

A rate is said to be feasible if it can be written as a joint probability distribution $p$ over our Markov chain space:

$$
\lambda_k = \sum_{j=1}^{2^K} \sum_{i=1}^{|I_j|} p_{ij} a^i_k x^i_k
$$

(36)

Furthermore, we introduce a new parameter $f \in \mathbb{R}_+^K$, which we will refer to as the awake vector, as follows:

$$
f_k = \sum_{j=1}^{2^K} a^j_k \alpha_j = \mathbb{E}(a_k) \quad \text{and} \quad \alpha_j = \sum_{i=1}^{|I_j|} p_{ij}
$$

(37)

$f_k$ can be seen as the required awake duration of each link $k$ as dictated by the arrival rate vector’s joint probability distribution $p$. 
Theorem

For any arrival rate $\lambda \in \Lambda$, there exist $(r^*, \rho^*)$ such that $s_k(r^*, \rho^*) = \lambda_k \ \forall k$. Moreover, $\hat{f}_k(r^*, \rho^*) = f_k \ \forall k$.

where $s_k(r^*, \rho^*)$ is the throughput of link $k$ and $\hat{f}_k(r^*, \rho^*)$ is the awake duration of link $k$. The proof is based on the fact that for any feasible arrival rate $\lambda \in \Lambda$, our goal can be summarized as calibrating the parameters $(r, \rho)$ in a way to make our CSMA Markov chain’s stationary distribution as close as possible to $p$. This is equivalent to reducing the distance between these two distributions:

$$\min_{r, \rho} \ D(p \parallel \pi(r, \rho)) = \sum_{j=1}^{2^K} \sum_{i=1}^{\|I_j\|} p_{ij} \log \left( \frac{p_{ij}}{\pi(a^j, x^i; r, \rho)} \right)$$

Therefore, the scheme is throughput optimal. Moreover, we can achieve the desired awake duration by simply solving the optimization problem.
Regions - I

By rewriting $f_k$ in the following manner:

$$f_k = \lambda_k + \omega_k$$

where $0 < \omega_k < 1 - \lambda_k$, we can see that a new parameter is born. This parameter is referred to as the Power-Delay Tradeoff (PDT) parameter and is assigned to each link $k$. We can conclude that the network is characterized by two regions rather than one: the capacity region and the awake region. An example is presented in the next figures for the case of two interfering links.
Regions - II

Figure: Capacity region for the case of two interfering links

Figure: Awake region for the case of two interfering links

Figure: Network Regions
Implementation

Theorem

If \( \lambda \in \text{int}(\Lambda) \) and \( f \in \text{int}(\Theta(\lambda)) \), then the optimum is attained for a finite \((r^*, \rho^*)\)

The proof is based on manipulation of convergent real sub-sequences. We can therefore implement the algorithm in a distributed manner to reach the optimal parameters. The updates are consequently done as follow:

\[
\begin{align*}
    r_k(m+1) &= r_k(m) + \Delta_{1k}(m)(\lambda_k - s_k(m)) \\
    \rho_k(m+1) &= \rho_k(m) + \Delta_{2k}(m)(\lambda_k + \omega_k - \hat{f}_k(m))
\end{align*}
\]

where \( \Delta \) refers to the chosen step size.
Numerical Results - Scenario

We consider a realistic heterogeneous case where several groups of links exist in the network with each group having its own desired power delay tradeoff. The number of groups is chosen as 3 with 4 links in each group:

- **Group 1**: This group is made of links that are delay sensitive but can tolerate a high power consumption
- **Group 2**: This group is made of links that fall between the two extremes, they require a moderate power consumption without introducing a lot of delay
- **Group 3**: This group is made of links that can tolerate long delays however they are extremely power limited

We consider that the arrival for each link is $\lambda_k = 0.077 \, \forall k$. $\omega_1 = 0.8$, $\omega_2 = 0.4$ and $\omega_3 = 0.1$ (PDT) are assigned for Groups 1, 2 and 3 respectively. We take as a benchmark the adaptive CSMA to compare the performance.
The scheme provides decent power gain with respect to the adaptive CSMA.
Power Gain

The scheme provides decent power gain with respect to the adaptive CSMA.
Collided Packets
Low collision probabilities with high nodes density and high total arrival rate.
Transmission Aggressiveness

Low collision probabilities with high nodes density and high total arrival rate.

**Figure:** Evolution of the transmission Aggressiveness
Wake-up Aggressiveness

![Graph showing the evolution of the waking-up aggressiveness for different groups over time.](image)

**Figure:** Evolution of the waking-up Aggressiveness
Take-away points

► We proposed a new CSMA scheme that combines both throughput optimality and energy efficiency
► A new parameter is introduced that has the interpretation of being a power delay tradeoff.
► We highlighted the gain in terms of power consumption with respect to the adaptive CSMA
► Suitable for the context of multiple services
► what about density? e.g. mMTC
► Comparison to other random access schemes, Grant-free Access, etc.
► Battery life?
Section 7

Transmission and Control of devices in dense networks
Motivation

- Power consumption and density are two key issues for mMTC.
- The allocation policy must minimize collisions from simultaneously transmitting machines and must exploit the channel quality between the machines and the receiver.
- Complex stochastic optimization problem, that can be cast as a Markov Decision Process (MDP). Decentralized solution?
- The problem of power control in large scale networks has been investigated in the past using game theory and mean field game, e.g. [76, 77, 78].
- Numerical solutions are usually challenging
- traffic pattern should be included: Efficient resource allocation.
- Queue- aware solutions: max weight [79, 80] or MDP [81, 82, 83].
- MDP shortcoming: high complexity (Bellman equation involves solving an exponentially large system of non-linear equations).
- Mean Field?\(^{27}\)

Network Scenario

Figure: Example of network scenario
System Model (1)

- $N$ transmitters communicating with a Base Station (BS).
- We consider the channel state of transmitter $n$, i.e., $h_n(t)$, to take values in the set \( \{c_1, \ldots, c_K\} \).
- The channel is further assumed to evolve from one time-slot to another according to a Markovian model with transition matrix $B^n = (b^n_{ij})_{i,j \in \{1,\ldots,K\}}$, for all $n$.
- We further note that $b^n_{ij} = \mathbb{P}(h_n(t+1) = c_i | h_n(t) = c_j)$.
- The SINR of user $n$ is given by
  \[
  \text{SINR}_n(h(t), p(t)) = \frac{h_n(t)p_n(t)}{\sum_{k \neq n} \alpha_k h_k(t)p_k(t) + N_0},
  \]
  \]
- In order to ensure a minimum quality of service it is required that
  \[
  \log_2(1 + \text{SINR}_n(t)) \geq \bar{\theta} \implies \text{SINR}_n(t) \geq 2^{\bar{\theta}} - 1 = \theta. \quad (38)
  \]
- Therefore, the achievable data rate of user $n$ is given by
  \[
  R_n(h(t), p(t)) = 1_{\{\text{SINR}_n(t) \geq \theta\}}, \quad (39)
  \]
Network Scenario

Figure: Example of network scenario (User Side)
System Model (2)

- Bursty traffic: $A_n(t)$ random data arrivals $A_n(t)$ (i.i.d. over slots).
- We assume that in each time slot there will be at most one packet arrival, i.e., $P(A_n(t) = 1) = \rho$ and $P(A_n(t) = 0) = 1 - \rho$ with $\rho > 0$.
- Let $Q_n^\phi(t)$ be the queue length under a power allocation policy $\phi$.

$$Q_n^\phi(t + 1) = \max\{Q_n^\phi(t) - R_n(h(t), p(t)), 0\} + A_n(t). \quad (40)$$

- The $N$ queue dynamics are coupled together due to the interference term in the expression of the SINR.
- Let $X_n^\phi(t) = (h_n(t), Q_n^\phi(t))$, be the state of transmitter $n$. The stochastic process $X_n^\phi(t)$ is a controlled Markov chain. The objective of the present work is then to minimize

$$\mathcal{L}^\phi = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{n=1}^{N} \mathbb{E} \left[ p_n(t) + \lambda Q_n^\phi(t) \right]$$
Mean Field Approach (1)

- The state of user \( n \) at time \( t \) is denoted as \( X_{n}^{\phi}(t) \) and equals \((c, q)\), where \( c \in \{c_1, \ldots, c_K\} \) and \( q \in \{0, \ldots, Q_{\text{max}}\} \).

- The users are distinguishable only through their state. This means that the behavior of the system only depends on the proportion of users in every state.

- Let \( M^N(t) \) be the empirical measure of the collection of users, it is a \( S \)-dimensional vector with the \( n \)-th component given by

  \[
  M^N_i(t) = (M^N_{i_1}(t), \ldots, M^N_{i_K}(t)),
  \]

  where \( M^N_i(t) = (M^N_{i,0}(t), \ldots, M^N_{i,Q_{\text{max}}}(t)) \), and

  \[
  M^N_{i,j}(t) = \frac{1}{N} \sum_{n=1}^{N} 1_{\{X_{n}^{\phi}(t)=(c_i,j)\}},
  \]

  for all \( i \in \{1, \ldots, K\} \) and all \( j \in \{1, \ldots, Q_{\text{max}}\} \).

- The value of \( M^N_{i,j}(t) \) is to be interpreted as the proportion of transmitter/users in channel state \( c_i \) and queue length \( j \).

- We then have that, the set of possible values for \( M^N \) is the set of probability measures on

  \[
  S = \{(c, q) : c \in \{c_1, \ldots, c_K\}, q \in \{1, \ldots, Q_{\text{max}}\}\}.
  \]
Mean Field Approach (2)

- The mean field approach allows us to move from a stochastic optimal control problem to a deterministic one.

- Let us denote by $DM^N(t)$ the expected drift of $M^N(t)$, that is,

$$DM^\phi := \mathbb{E}(M^N(t + 1) - M^N(t)|M^N(t)).$$

- $s_i$ is the state that corresponds to the $i^{th}$ entry in $M^N$ and $\nu_{i,i}(m)$ is the probability that a user in state $s_i \in S$ at time slot $t$.

- The transitions to state $s_j \in S$ at time slot $t + 1$ given that $M^N(t) = m$, is,

$$\nu_{i,j}(m) := g_i(m)\gamma_{i,j}^1 + (1 - g_i(m))\gamma_{i,j}^0,$$

where $g_i(m)$ is the fraction of users in state $s_i \in S$ whose SINR$_n(t) \geq \theta$, with $n$ a user in state $s_i$. $\gamma_{i,j}$ represents the transition probabilities from state $s_i$ to state $s_j$, when the SINR$_n(t) \geq \theta$ for all users $n$ in state $s_i$. 
Mean Field Approach (3)

We then have \( D M^N(t) \bigg|_{M^N(t) = m} = \sum_i \sum_j \nu_{i,j}(m) \tilde{e}_{ij} = U^\phi(m)m \), where

\[
\tilde{e}_{ij} = (0, \ldots, 0, -1, 0, \ldots, 0, 1, 0, \ldots),
\]
i.e., the \( K \cdot (Q_{\text{max}} + 1) \) dimensional vector with a \(-1\) entry in the \( i\)th position and the entry at \( j\)th position equal to \( 1 \) and \( I \) is the identity matrix. We further have \( \tilde{e}_{ii} = 0 \). Also note that

\[
U^\phi_{i,j}(m) = \begin{cases} 
- \sum_{r \neq i} \nu_{i,r}(m) & \text{if } i = j, \\
\nu_{j,i}(m) & \text{if } i \neq j.
\end{cases}
\]

We can now define \( m(t + 1) - m(t) = U^\phi_{i,j}(m(t))m(t) \).

Assuming that \( \phi \) is such that all users in same state \( s \in S \) are allocated same power and that \( \mu_n = \mu_n' \) if user \( n \) and \( n' \) are both in the same state \( s \in S \), we can equivalently write

\[
\mathcal{L}^\phi,N \approx \limsup_{T \to \infty} \frac{N}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{K \cdot Q_{\text{max}} + 1} \mathbb{E}(p_i(t) + m_i(t)\sigma(i)), \tag{41}
\]

where \( \sigma(\cdot) \) is a mapping between \( i \in \{1, K \cdot (Q_{\text{max}} + 1)\} \) and the queue-length. Namely, if \( i = z \ast K + j \) then \( \sigma(i) = j \) for all \( z \in \{0, \ldots, Q_{\text{max}} + 1\} \).
Mean Field Approach (4)

Proposition

A user in state $i$ achieves $\text{SINR}_i(t) = \frac{p_i(t)h_i}{\sum_{j=1}^{K \cdot (Q_{\text{max}} + 1)} p_j(t)h_jm_j(t) + N_0}$.

The problem is therefore to find the power allocation policy $\phi$ such that we

$$\text{minimize} \int_0^{\infty} \left( \sum_{i=1}^{K \cdot (Q_{\text{max}} + 1)} p_i(t) + m_i(t)\sigma(i) - E^* \right) dt,$$

where $E^*$ is the optimal equilibrium cost, subject to

$$dm(t) = U^\phi(m(t))m(t)dt, \text{ and } p_i(t) \leq p_{\text{max}}.$$
Threshold Policy

Theorem

Let \( N_0^0 \) and \( N_0^1 \) be given by Equation (42). The optimal solution is:

- If \( N_0 \geq N_0^0 \), then \( s_4(t) = 0 \) if \( m_4(t) \leq m_4^1 \) and \( s_4(t) = 1 \) otherwise.
- If \( N_0 \leq N_0^1 \), then \( s_4(t) = 0 \) if \( m_4(t) \leq m_4^0 \) and \( s_4(t) = 1 \) otherwise.
- If \( N_0 \in (N_0^1, N_0^0) \), then \( s_4(t) = 0 \) if \( m_4(t) \leq m_4^* \) and \( s_4(t) = 1 \) otherwise. We note that \( m_4^1 = \beta_1 \rho / (\rho + \beta_1 (1 - \text{rho})) \), \( m_4^0 = \beta_1 \) and \( m_4^* \) solution of \( dE^*(\bar{s}_4) / d\bar{s}_4 \).

\[
N_0^0 = \frac{\beta_1 (1 - \rho)(1 - \theta \beta_1)}{\rho \theta}, \quad \text{(42)}
\]

\[
N_0^1 = \frac{\beta_1 (1 - \rho)\rho(\rho + \beta_1 - \beta_1 \rho(1 + \theta))^2 / (\beta_1 + \rho - \beta_1 \rho)^2}{\theta(2\beta_1 (1 - \rho)\rho(1 - \theta \beta_1) + \rho^2 (1 - \theta \beta_1) + \beta_1^2 (1 - \rho)^2)}, \quad \text{(43)}
\]

- The optimal solution is of threshold type. That is, it suffices to compare the fraction of users that have one packet to transmit and are in a good channel state with respect to \( N_0 \).
Take-away points

- Density could be seen as an opportunity.
- The transmission and Control can be of threshold type, which allows to simplify the control of the network.
- The achieved QoS (outage, reliability, etc.) has to be studied.
- Interaction with NOMA and other services (multiple access schemes) requires particular attention.
- Although Mean Field can help reducing the complexity, obtained solutions are simple only under some particular assumptions.
- Critical MTC applications may require other tools.
- Other tools: multi-agent systems, Machine Learning,...
- Future work: Energy harvesting
- Impact of Mobility? V2X?
Section 8

Conclusion
Conclusion

- Massive MIMO: enhancement of spectral and energy efficiencies
  - CSI acquisition
  - User grouping can help to increase the number of served users.
  - Adapt to user mobility
- Non-Orthogonal Multiple Access (NOMA) is suitable for MTC
  - Receiver complexity requires a particular attention especially for high number of devices.
- Random Access (collision management, sleep mode, etc) can be used at the MAC level to improve the network performance and increase the battery life.
- Mean field approach can also be used to simplify the resource allocation and network control strategies.
- Other topics: D2D, V2X, FD, etc.
Thank You!

Questions?
References I


References II


"NGMN alliance 5G white paper," https://www.ngmn.org/5g-whitepaper/5g-white-paper.html, 2015.


References III


References IV


References V


References VI


Tamas Szalka and Peter Fulop and Sandor Imre, *General mobility modeling and location prediction based on markovian approach constructor framework*, Periodica Polytechnica Electrical Engineering (Archives), 55, 2011.

References VII


Yang Li, Young-Han Nam, Boon Loong Ng, Jianzhong Zhang, *A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse*, Global Communications Conference (GLOBECOM), 2012 IEEE.


Nokia, Alcatel-Lucent Shanghai Bell, Performance of Interleave Division Multiple Access (IDMA) in combination with OFDM family waveforms, R1?165021, 3GPP TSG RAN WG1 Meeting, Nanjing, China, May 2016.
References VIII

ZTE, "Contestion-based non-orthogonal multiple access for UL mMtc", R1-164269, 3GPP TSG RAN WG1 Meeting, Nanjing, China, 23rd - 27th May, 2016.


References IX


F. Meriaux and S. Lasaulce. Mean-field games and green power control. In IEEE International Conference on Network Games, Control and Optimization (NetGCooP), 2011.


