Machine-type Communications: from massive connectivity to ultra-reliable low latency communication

Hirley Alves and Jimmy Nielsen

T5: TUTORIALS

28 August, ISWCS 2018, Lisbon

UNIVERSITY OF OULU

CWC Oulu CENTRE FOR WIRELESS COMMUNICATIONS University of Oulu

AALBORG UNIVERSITY

Machine-type Wireless Communications: from massive connectivity to ultra-reliable low latency communication

Hirley Alves and Jimmy Nielsen

UNIVERSITY OF OULU

CWC Oulu CENTRE FOR WIRELESS COMMUNICATIONS University of Oulu

AALBORG UNIVERSITY

Hirley Alves

2017 Adj. Prof. University of Oulu, Finland

2015 DSc

Research interests

- wireless and cooperative communications
- wireless full-duplex communications
- PHY-security
- 5G/6G, IoT, MTC & URLLC

UNIVERSITY OF OULU

CWC Oulu CENTRE FOR WIRELESS COMMUNICATIONS University of Oulu

Jimmy Nielsen

Assoc. Prof. Aalborg University, Denmark

Research interests

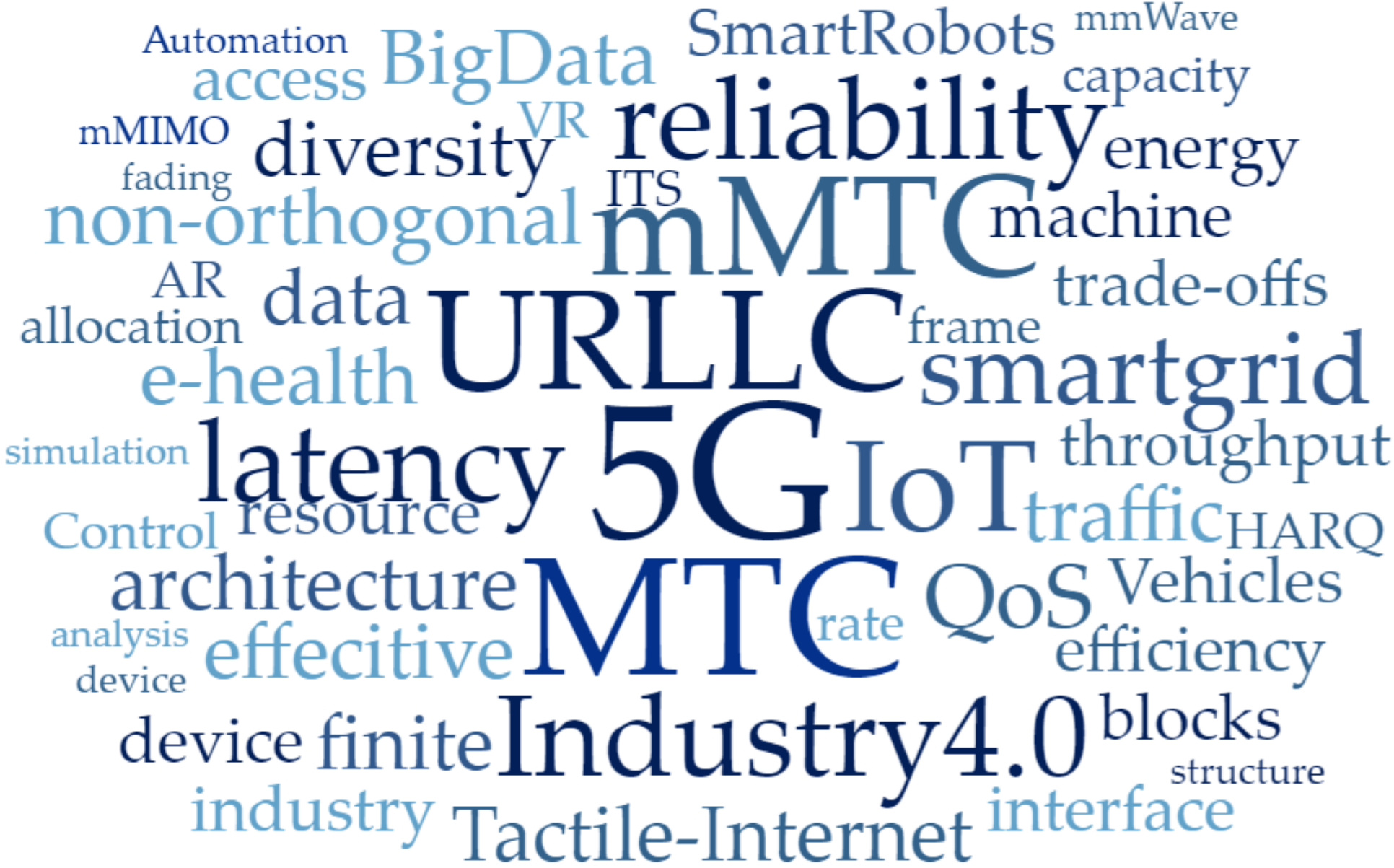
- URLLC and mMTC in the context of smart grid
- IoT
- 5G system

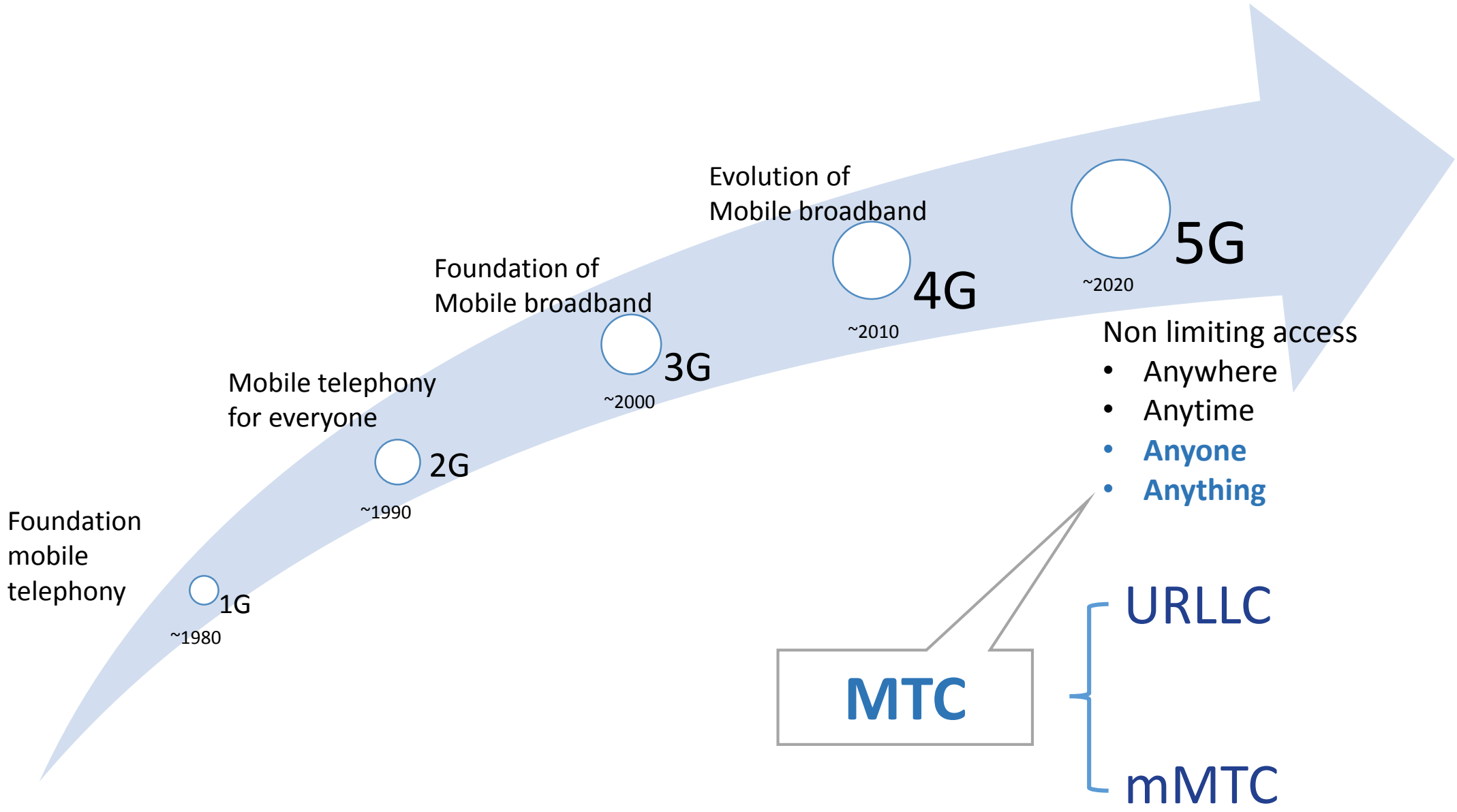
AALBORG UNIVERSITY

What is Machine-type Wireless Communications?

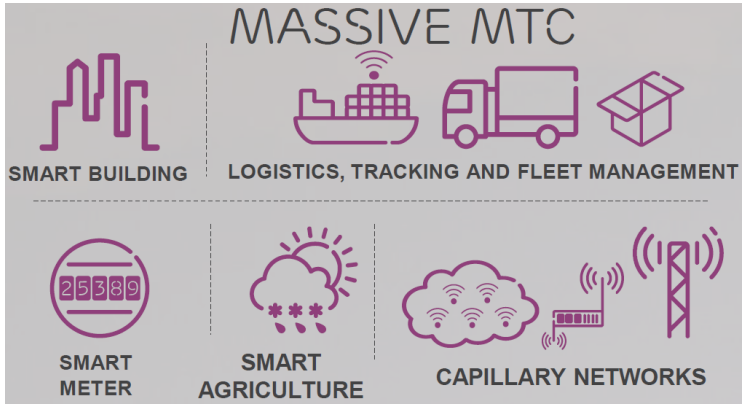
What is Machine-type Wireless Communications?

- IoT
- M2M, MTC, URLLC?
- V2V, V2X?
- nb-IoT, LPWANs,…
- ….
- ….

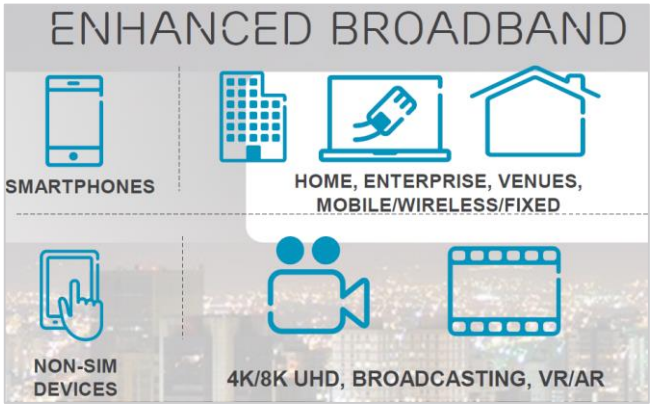

Automation
BigData
SmartRobots
mmWave
access
capacity
reliability
VR
diversity
mMIMO
energy
fading
ITS
non-orthogonal
mMTC
machine
AR
data
trade-offs
allocation
URLLC
frame
smartgrid
e-health
5G
IoT
throughput
simulation
latency
traffic
HARQ
resource
Control
architecture
MTC
QoS
Vehicles
analysis
rate
effecitive
efficiency
device
Industry4.0
device
finite
blocks
structure
industry
Tactile-Internet
interface

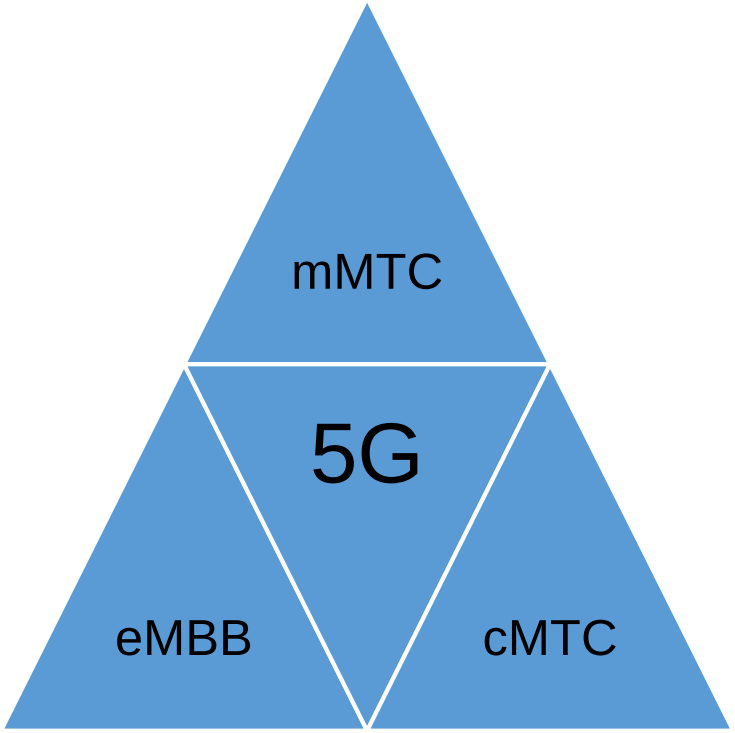

Foundation mobile telephony
1G
~1980
Mobile telephony for everyone
2G
~1990
Foundation of Mobile broadband
3G
~2000
Evolution of Mobile broadband
4G
~2010
5G
~2020
Non limiting access
• Anywhere
• Anytime
• Anyone
• Anything
MTC
URLLC
mMTC

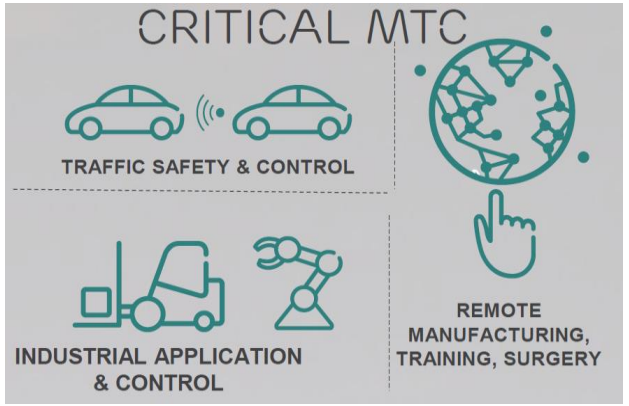
Low Cost
Low Energy
Low Data Volume
Large Numbers



Ultra Reliable,
Low Latency,
High availability

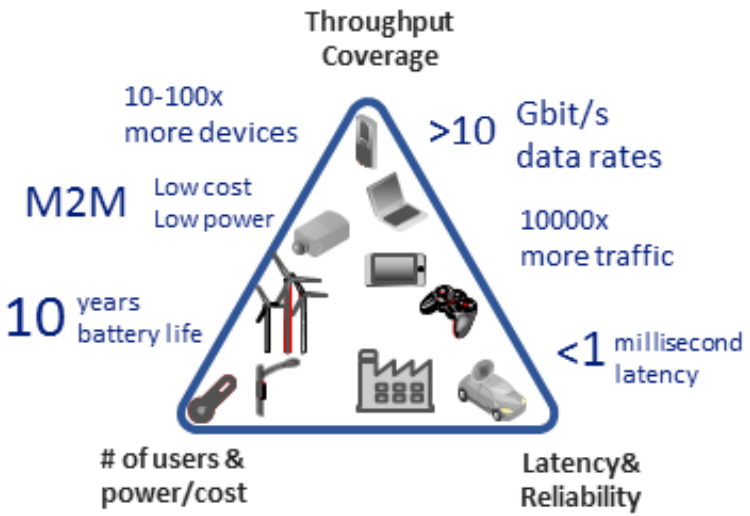




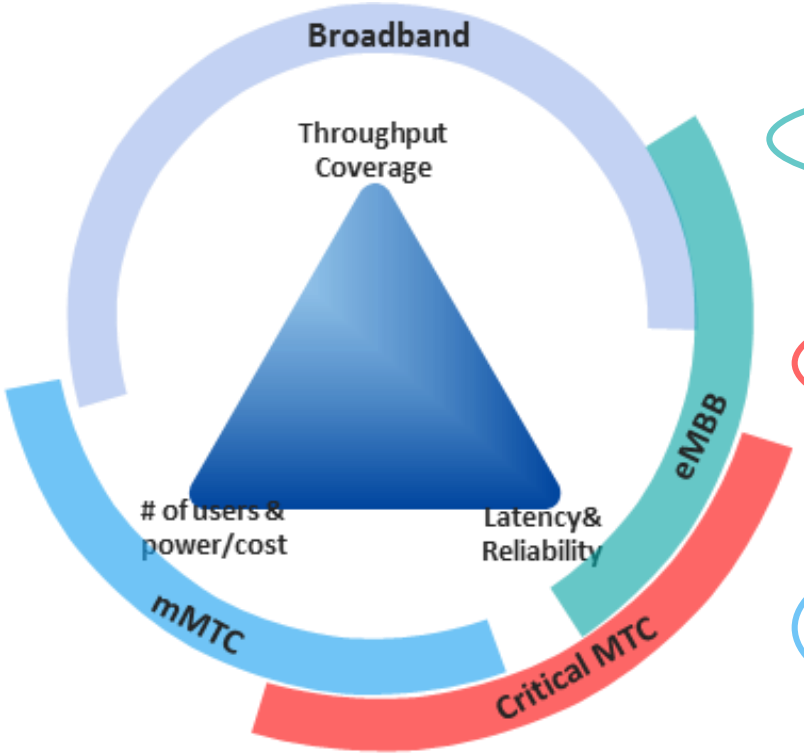


High Throughput
Low Latency

Figures from: H. Ji, *et al.* "Introduction to Ultra Reliable and Low Latency Communications in 5G", arXiv:1704.05565v1
O. Yilmaz, Ultra-Reliable and Low-Latency 5G Communication, EuCNC'16



5G Cornerstones

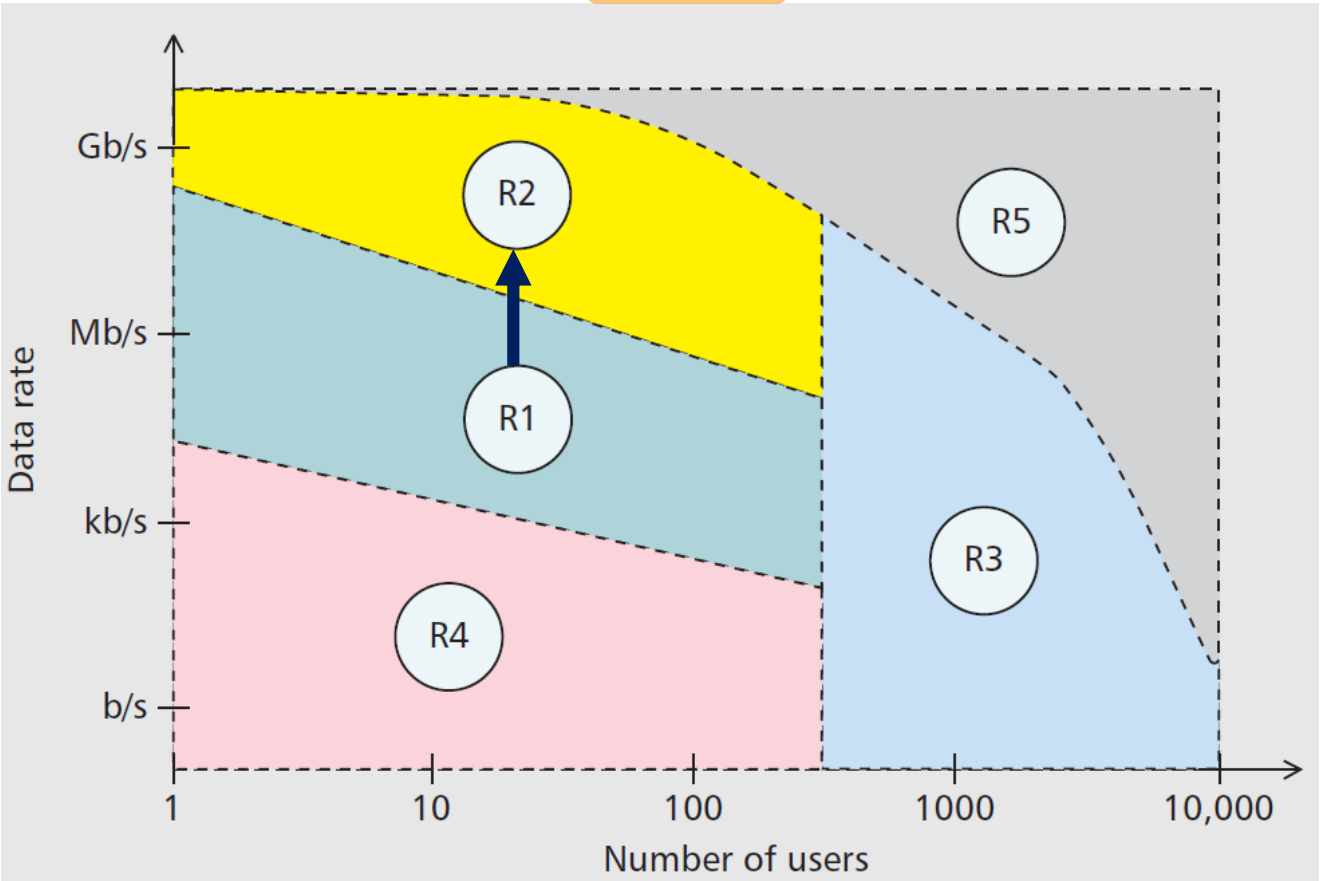


eMBB: Virtual and augmented reality
Throughput: 4-28 Gbps
Latency: < 7ms

cMTC: Factory automation/control
Latency: < 1ms
Reliability: > 99.999999%

mMBB: factory monitoring, smart grid/city
Density: .1-10 devices/m^3
Latency: 4-10 ms
Reliability: up to 99.999%

~ 4G to ‘5G’



- **R1 ~4G**
- **R2 ~eMBB**
- **R3 ~mMTC**
- **R4 ~URLLC (cMTC)**
- **R5 – not feasible**

R4 -> URLLC ?
What is missing then?

Figure from F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," in *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74-80, February 2014.

Factory Automation
≤ 1 ms

Motion Control
≤ 1 ms

Remote Control
5-100 ms

Intelligent Transportation Systems
5 ms

Smart Grid
3-5 ms

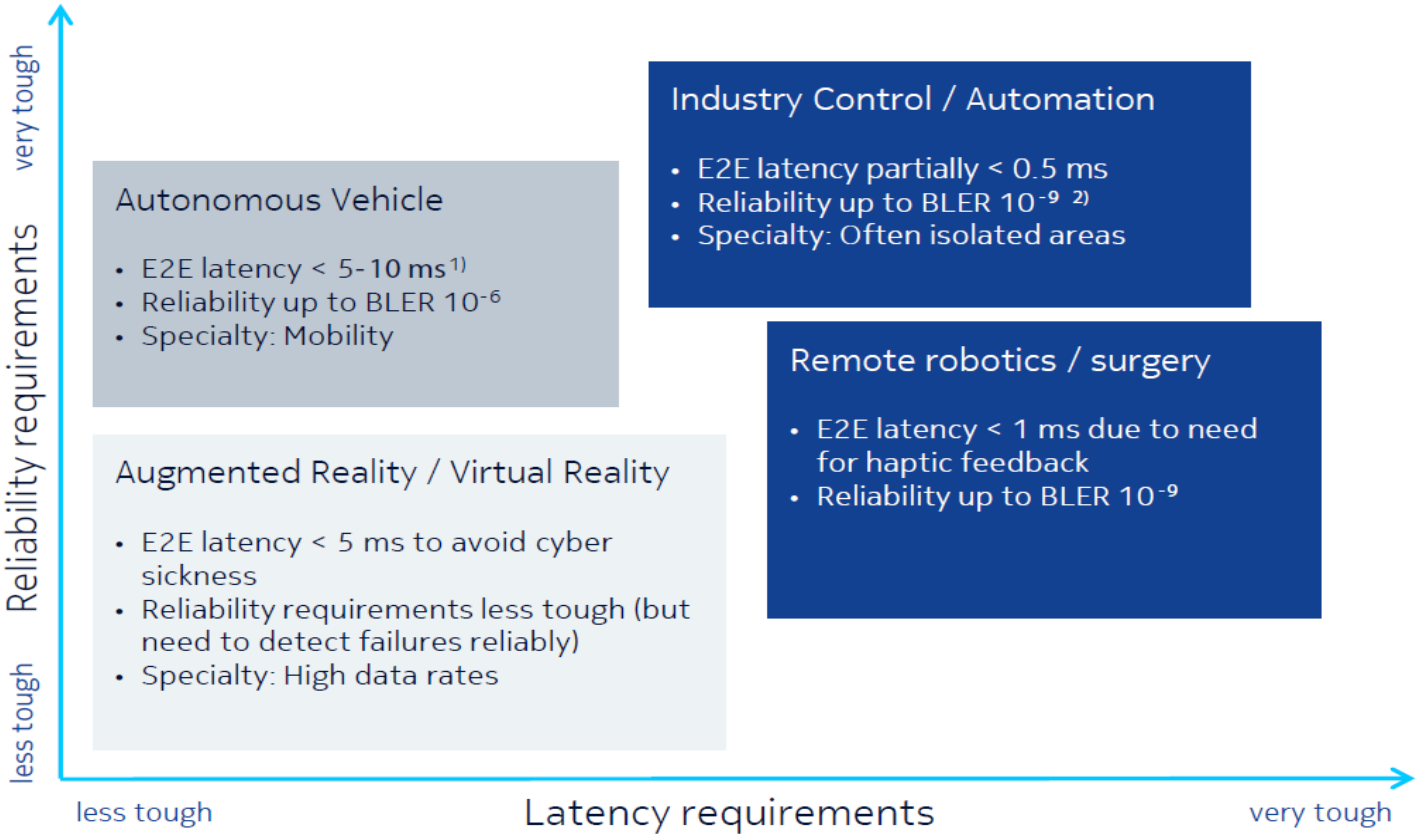
Tactile Internet
1 ms

Process Automation
100 ms

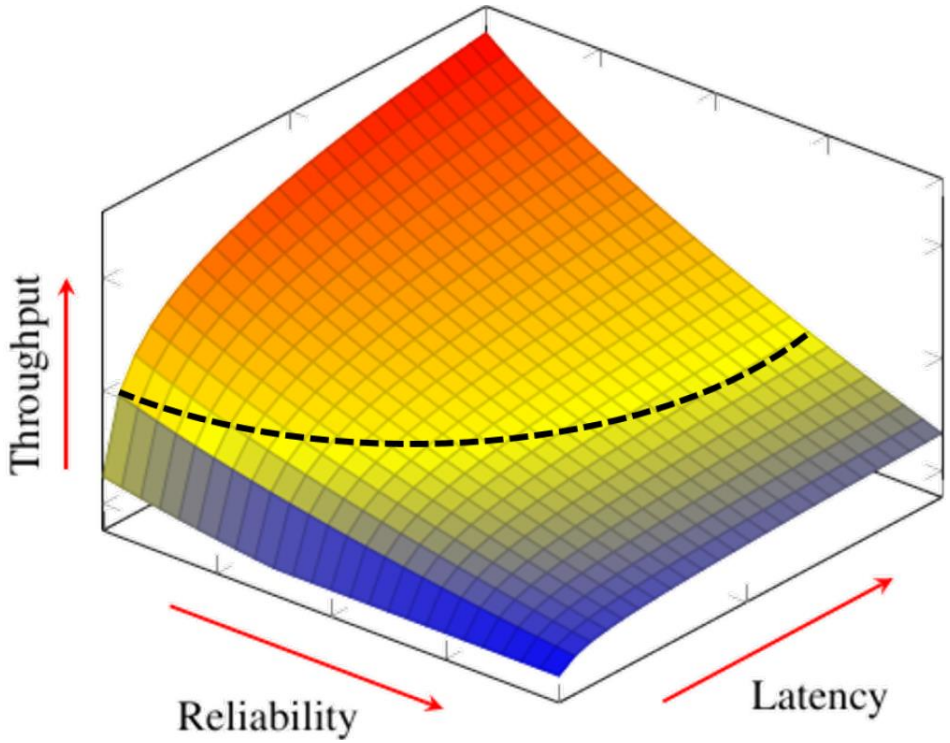
Automated Guided Vehicle
15-20 ms

Numbers are examples, requirements vary within one application area

Figures from: H. Ji, *et al.* "Introduction to Ultra Reliable and Low Latency Communications in 5G", arXiv:1704.05565v1
O. Yilmaz, Ultra-Reliable and Low-Latency 5G Communication, EuCNC'16



Figures from: "5G for Mission Critical Communication Achieve ultra-reliability and virtual zero latency", Nokia White Paper, 2016



Figures from: H. Ji, *et al.* "Introduction to Ultra Reliable and Low Latency Communications in 5G", arXiv:1704.05565v1

MTC over Cellular Networks

MTC over Cellular Networks

Pros

+ Coverage
+ Roaming
+ Interoperability
+ QoS guarantees
+ Service level agreements/platforms
+ ….

Cons

- Identification
- Coverage (indoor)
- Access to core networks
- Generated traffic
- Congestion
- Massive # of devices
- Complexity
- Power consumption
- Energy Efficiency
- Cost
- …..

Requirements	HTC over cellular	MTC over cellular
Uplink	Uplink is usually more lightly loaded and power-constrained	For many MTC applications, the main bottleneck; high signaling overhead and extreme power constraints
Downlink	The main bottleneck for high data rate services, since most traffic comes from the core network	Needs to be able to deep sleep, but wake up on command for network-initiated communication
Subscriber load	Relatively few (< 100) simultaneous devices per cell	Many (>> 100) simultaneous devices per cell with traffic uploading that can be event-triggered, periodic, or continuous
Device types	Relatively homogeneous, smart phones and data consumption devices like tablets	Extremely heterogeneous device landscape that includes environmental sensors, utility meters, wearable devices, and many unforeseen applications
Delay requirements	Defined service classes by 3GPP, vary between real-time conversational and best effort data	Very diverse delay requirements, ranging from emergency/ time critical to very delay tolerant applications
Energy requirements	Flexible energy requirements due to the ability to recharge daily	Many ultra-low energy applications that require extreme power consumption measures
Signaling requirements	Signaling protocol overhead is not a concern and the design provides reliable mobility and connection management mechanisms	Application-dependent signaling protocols, with extremely efficient overhead signaling and contention resolution
Architectural requirements	Well-understood hierarchical cellular architecture with standardized interfaces between access and core network elements	Wide area coverage may require integration of data aggregators with multihop relaying; relaxed requirements for handover and roaming support

Z. Dawy, *et al*, "Toward Massive Machine Type Cellular Communications," in *IEEE Wireless Communications*, vol. 24, no. 1, pp. 120-128, February 2017.

Massive MTC: requirements & characteristics

Massive MTC: requirements & characteristics

Requirements & characteristics

- Small packets
- Large number of users/cell
- Uplink dominant
- Low data rates
- Mixed traffic models
  - Periodic
  - Event based
- Low complexity
- Low energy consumption

Challenges

- Control signaling
  - Lower overhead stablishing connection, and recovering from idle mode
- Access
  - No reservation
- Multi service integration
  - Coexistence of several services with heterogeneous requirements
- Energy Efficiency
  - Lower power consumption
  - Event based transmission

Carsten Bockelmann *et al.* "Towards Massive Connectivity Support for Scalable mMTC Communications in 5G networks",arXiv:1804.01701v1

Massive MTC: requirements & characteristics

Requirements & characteristics

- Small packets
- Large number of users/cell
- Uplink dominant
- Low data rates
- Mixed traffic models
  - Periodic
  - Event based
- Low complexity
- Low energy consumption

Challenges

- Control signaling
  - **Data aggregation**
- Access
  - **Non-orthogonal multiple access**
- Energy Efficiency
  - Lower power consumption
  - Event based transmission
- Multi service integration
  - Coexistence of several services with heterogeneous requirements

Carsten Bockelmann *et al.* "Towards Massive Connectivity Support for Scalable mMTC Communications in 5G networks",arXiv:1804.01701v1

Non-orthogonal Multiple Access

Mohammad Shehab (E. Dosti)

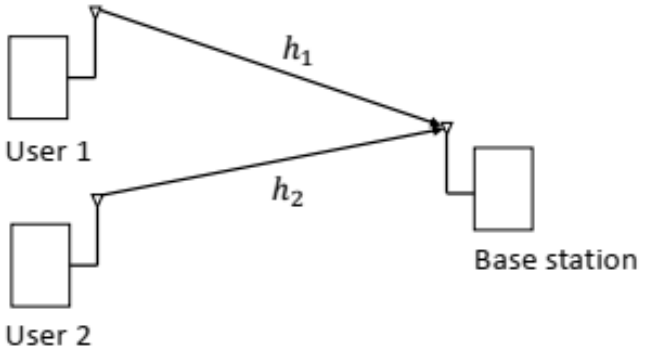
Non-orthogonal Multiple Access

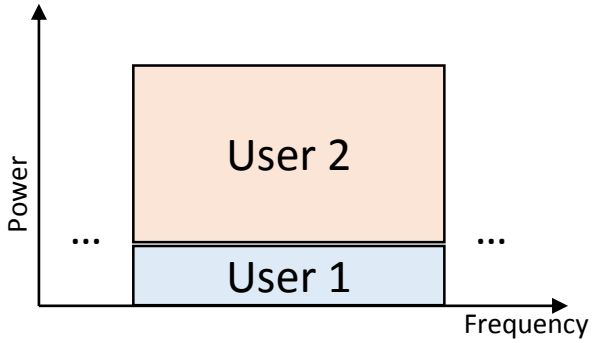
Non-orthogonal Multiple Access

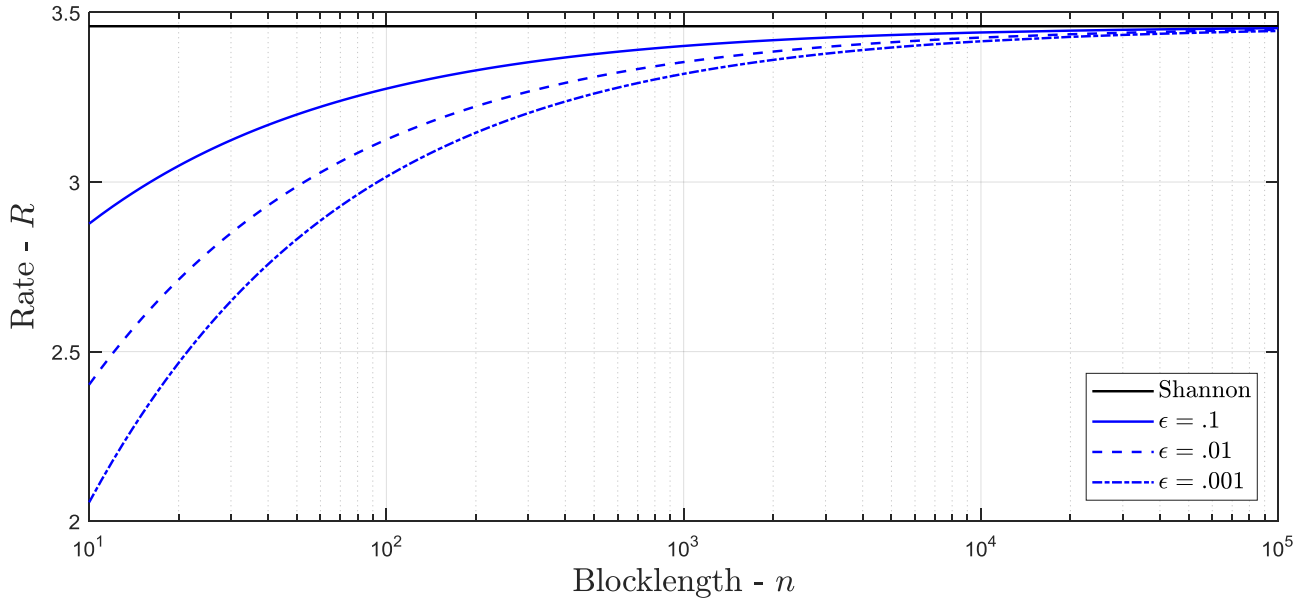
- NOMA
  - Power domain multiplexing
  - Large number of served users
  - Short packets

$$R_f \approx \log_2(\mathbf{1} + \boldsymbol{SNR}) - \sqrt{\frac{V(SNR)}{n}} Q^{-1}(\epsilon) \log_2 e$$

$$V(x) = x\,\frac{2+x}{(1+x)^2}$$ **Channel dispersion**







E. Dosti, M. Shehab, H. Alves and M. Latva-aho, "On the Performance of Non-orthogonal Multiple Access in the Finite Blocklength Regime" conditionally accepted.

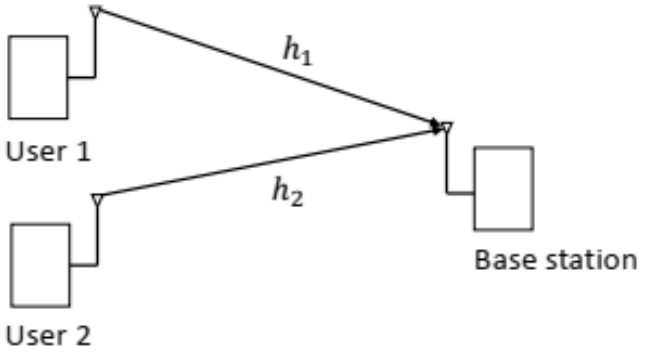
Non-orthogonal Multiple Access

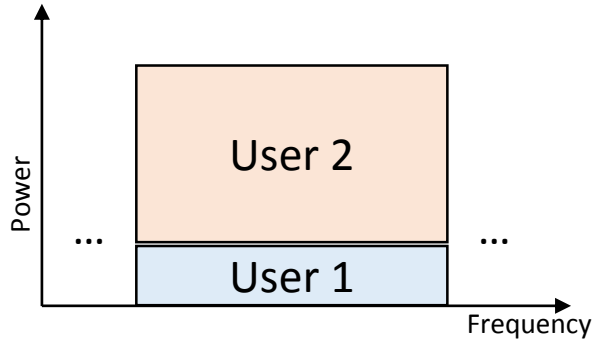
Non-orthogonal Multiple Access

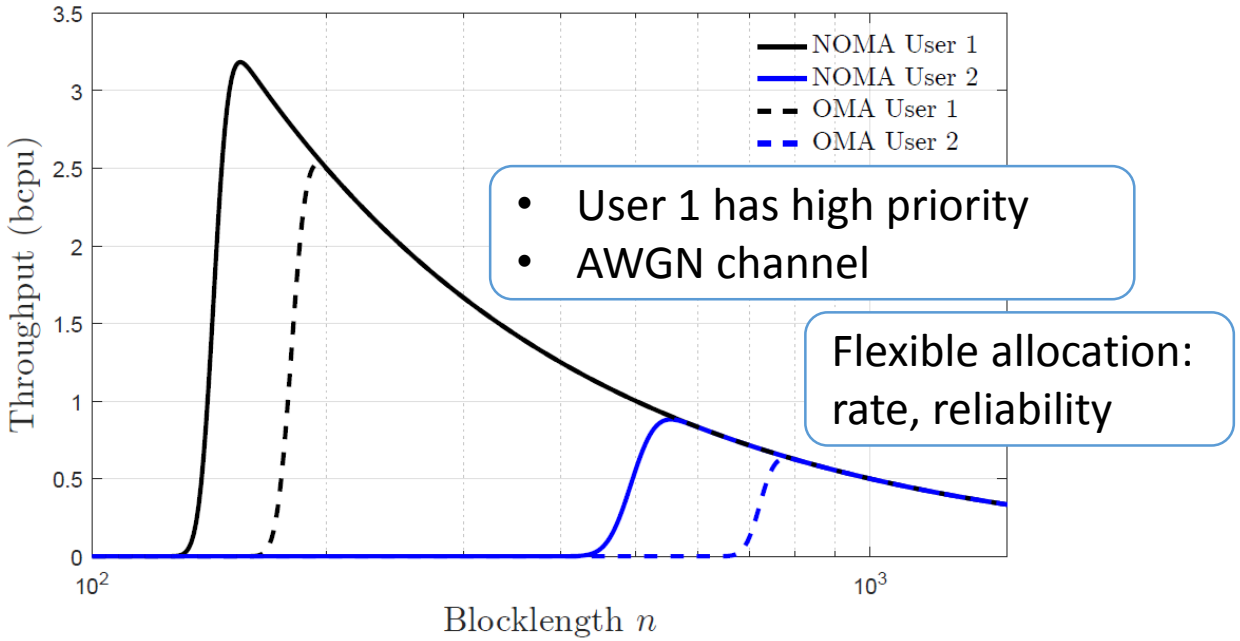
- NOMA
  - Power domain multiplexing
  - Large number of served users
  - Short packets

$$R_f \approx \log_2(\mathbf{1} + \boldsymbol{SNR}) - \sqrt{\frac{V(SNR)}{n}} Q^{-1}(\epsilon) \log_2 e$$

$$V(x) = x\,\frac{2+x}{(1+x)^2}$$ **Channel dispersion**







- User 1 has high priority
- AWGN channel

Flexible allocation: rate, reliability

Figure 3: Throughput as a function of the number of channel uses, considering $k = 500$ bits, $P_1 = P_2 = 10$ dB.

E. Dosti, M. Shehab, H. Alves and M. Latva-aho, "On the Performance of Non-orthogonal Multiple Access in the Finite Blocklength Regime" conditionally accepted.

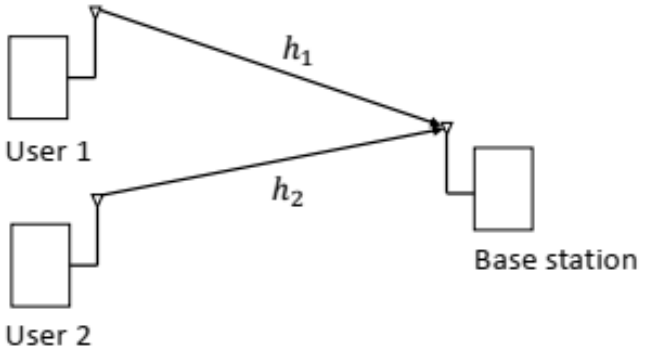
Non-orthogonal Multiple Access

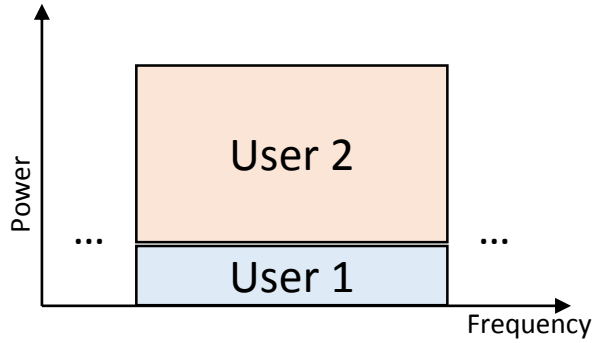
Non-orthogonal Multiple Access

- NOMA
  - Power domain multiplexing
  - Large number of served users
  - Short packets
  - Fading channels
  - ARQ

- User 1 has high priority
- Large gains for small packets
- Even User 2 experiences better performance

Drawback: Delay





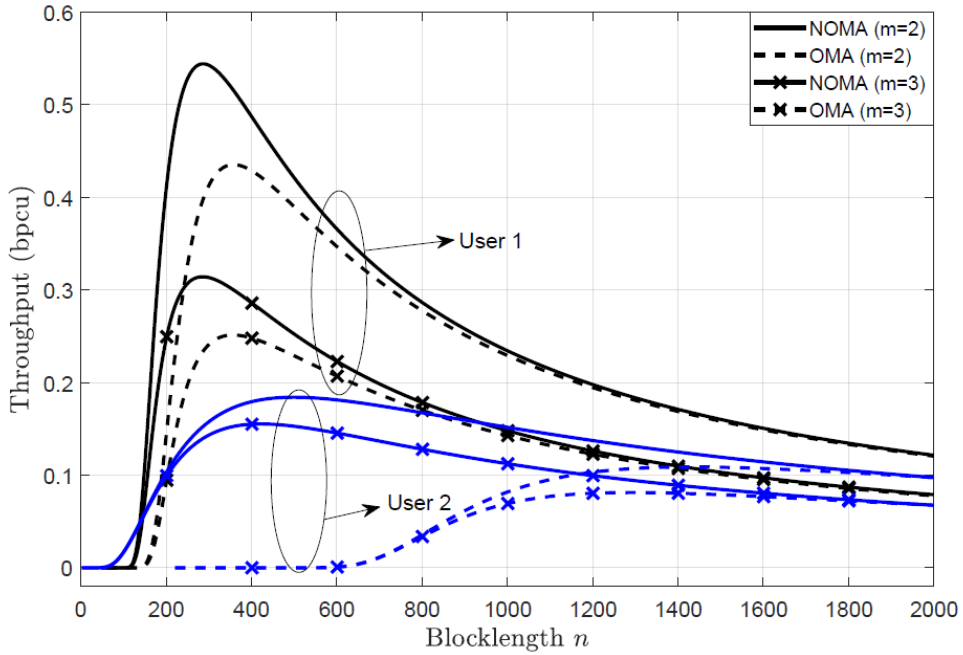


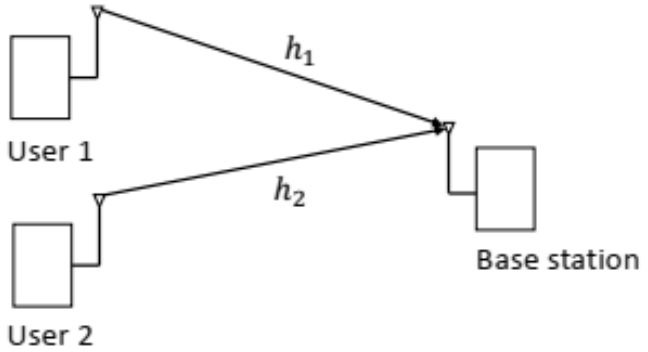
Figure 8: Throughput of ARQ in NOMA and OMA schemes as a function of blocklength n for $P_1 = P_2 = 10$ dB and $k = 500$.

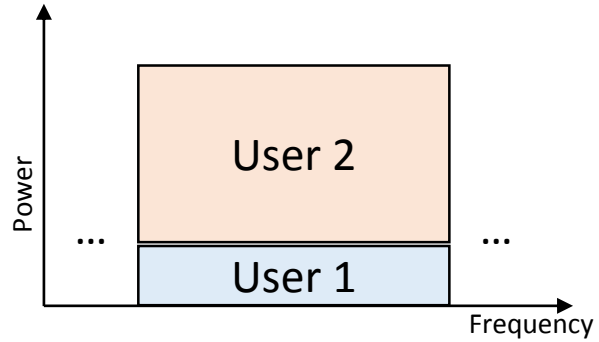
E. Dosti, M. Shehab, H. Alves and M. Latva-aho, "On the Performance of Non-orthogonal Multiple Access in the Finite Blocklength Regime" conditionally accepted.

Non-orthogonal Multiple Access

Non-orthogonal Multiple Access

- NOMA
  - Power domain multiplexing
  - Large number of served users
  - Short packets
  - ACK is not granted





- Latency increases with use of resources
- NOMA > OMA
- Larger gains for User2

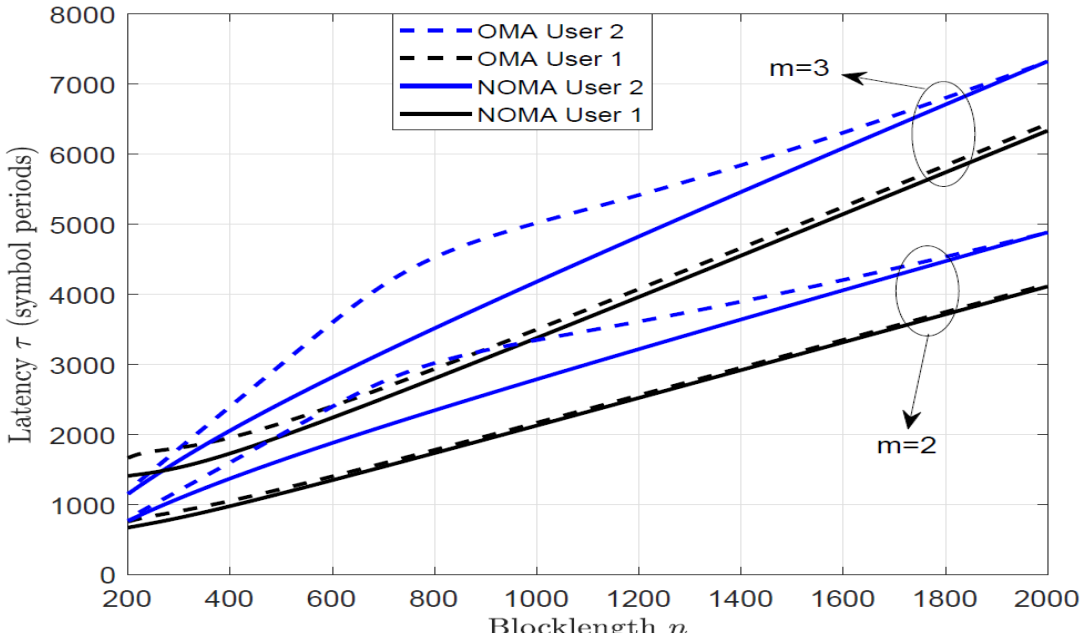


Figure 10: Latency of ARQ in NOMA and OMA schemes as a function of blocklength n for $P_1 = P_2 = 10$ dB and $k = 500$.

E. Dosti, M. Shehab, H. Alves and M. Latva-aho, "On the Performance of Non-orthogonal Multiple Access in the Finite Blocklength Regime" conditionally accepted.

Data Aggregation and Non-orthogonal Multiple Access

Onel López

Data Aggregation and Non-orthogonal Multiple Access

- NOMA: network level
- Aggregators forming a HPPP.
- MTDs uniformly distributed around the aggregator
- K ~ Poiss(m) MTDs per cluster.
- N orthogonal channels per cluster.
- L MTDs per channel (L=2).
- Rayleigh fading
- Perfect CSI at the aggregators.
- Imperfect SIC.
- Full inversion power control.

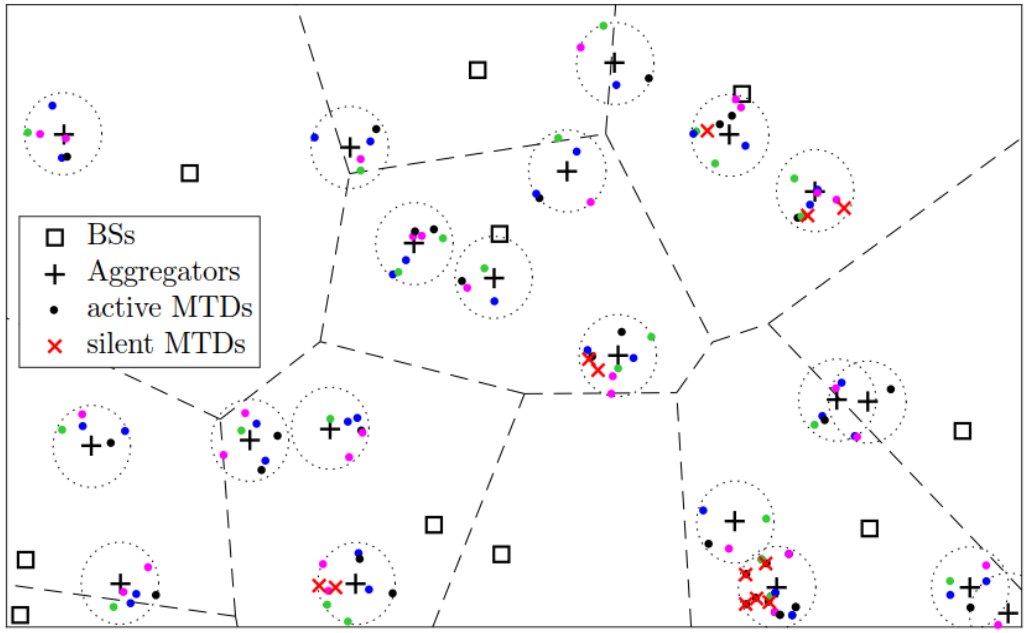
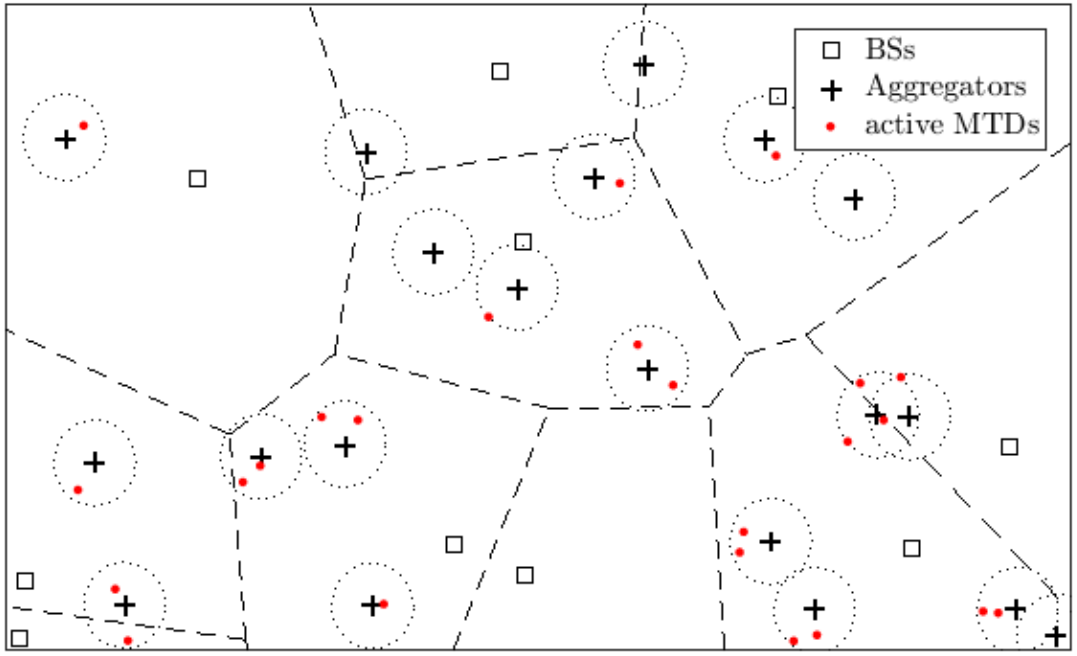


Fig. 1. Snapshot of the system model with $\bar{m} = 6$, $L = 2$ and $N = 4$. MTDs with the same color are using the same channel across the entire network.

O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4750-4765, July 2018. arXiv:1708.07691

Data Aggregation and Non-orthogonal Multiple Access

- NOMA: network level
- Aggregators forming a HPPP.
- MTDs uniformly distributed arou the aggregator
- K ~ Poiss(m) MTDs per cluster.
- N orthogonal channels per cluste
- L MTDs per channel (L=2).
- Rayleigh fading
- Perfect CSI at the aggregators.
- Imperfect SIC.
- Full inversion power control.



Snapshot of the active MTDs in a given channel.

O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4750-4765, July 2018. arXiv:1708.07691

Data Aggregation and Non-orthogonal Multiple Access: Scheduling

Random Resource Scheduling (RRS)

- N out of the K MTDs requiring transmissions are independently and randomly chosen and matched, one-to-one, with the N channels.
- If $K \leq N$, all MTDs get channel resources.
- If $K \geq N$, the channel allocation is executed again by allowing the remaining MTDs to share channels with the already served MTDs.
- Repeat until all the MTDs are allocated or the maximum number of MTDs per channel, L, is reached for all the channels.

CSI is only required at the aggregators when decoding the arriving information and not for resource scheduling.

Channel-aware Resource Scheduling (CRS)

- The MTD with better SIR will be preferentially assigned with the available channel resources.
- If $K \leq N$ all the get channel resources.
- If $K \geq N$, best N MTDs while allocating them randomly in the N channels.
- Remaining MTDs can be still allocated sharing those same resources, i.e., users N + 1,...,K go to the second round for allocation.
- Repeat until all the MTDs are allocated or the maximum number of MTDs per channel, L, is reached.

- **CSI** for decoding multiple user data over the same orthogonal channel with SIC
- CRS strongly relies on the CSI for resource scheduling.

O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4750-4765, July 2018. arXiv:1708.07691

Data Aggregation and Non-orthogonal Multiple Access: Scheduling

Random Resource Scheduling (RRS)

- N out of the K MTDs requiring transmissions are independently and randomly chosen and matched, one-to-one, with the N channels.
- If K ≤ N, all MTDs get channel resources.
- If K ≥ N, the channel allocation is executed again by allowing the remaining MTDs to share channels with the already served MTDs.
- Repeat until all the MTDs are allocated or the maximum number of MTDs per channel, L, is reached for all the channels.

$$\mathrm{SIR}_{1,1}^{r} = \frac{h}{I_r},$$

$$\mathrm{SIR}_{1,2}^{r} = \frac{\max(h', h'')}{I_r + \min(h', h'')},$$

$$\mathrm{SIR}_{2,2}^{r} = \frac{\min(h', h'')}{I_r + \mu \max(h', h'')},$$

Channel-aware Resource Scheduling (CRS)

- The MTD with better SIR will be preferentially assigned with the available channel resources.
- If K ≤ N all the get channel resources.
- If K ≥ N, best N MTDs while allocating them randomly in the N channels.
- Remaining MTDs can be still allocated sharing those same resources, i.e., users N + 1,...,K go to the second round for allocation.
- Repeat until all the MTDs are allocated or the maximum number of MTDs per channel, L, is reached.

$$\mathrm{SIR}_{1,1}^{c\ (i)} = \frac{h_i}{I_c},$$

$$\mathrm{SIR}_{1,2}^{c\ (i)} = \frac{a_i h_i}{I_c + b_i h_{i+N}},$$

$$\mathrm{SIR}_{2,2}^{c\ (i)} = \frac{b_i h_{i+N}}{I_c + \mu a_i h_i}.$$

O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4750-4765, July 2018. arXiv:1708.07691

Data Aggregation and Non-orthogonal Multiple Access: Scheduling

$a_i + b_i = \delta$ Fair coexistence between OMA and NOMA

Theorem 4. *A proper approximate choice for a_i and b_i in order to attain a similar reliability for both MTDs sharing the sharing channel when $u = 2$ is given by*

$$a_i = \frac{\delta\left(1+\frac{1}{\theta}\right)\left(\psi(K+1)-\psi(i+N)\right)}{\left(1+\mu+\frac{2}{\theta}\right)\psi(K+1)-\left(\mu+\frac{1}{\theta}\right)\psi(i)-\left(1+\frac{1}{\theta}\right)\psi(i+N)}, \quad (31)$$

$$b_i = \delta - a_i. \quad (32)$$

Theorem 6. *The required δ, δ^*, for a fair coexistence between OMA and NOMA setups is approximated by the solution of*

$$\xi^{\delta^{\frac{2}{\alpha}}-1} + \xi^{2^{\frac{\alpha-2}{\alpha}}\delta^{\frac{2}{\alpha}}-1} = 2, \quad (35)$$

where $\xi = \exp(-\chi c_2 s^{\frac{2}{\alpha}})$, and it is bounded by

$$2^{\frac{2-\alpha}{\alpha}} \leq \delta^* \leq 1. \quad (36)$$

Proof. See Appendix G

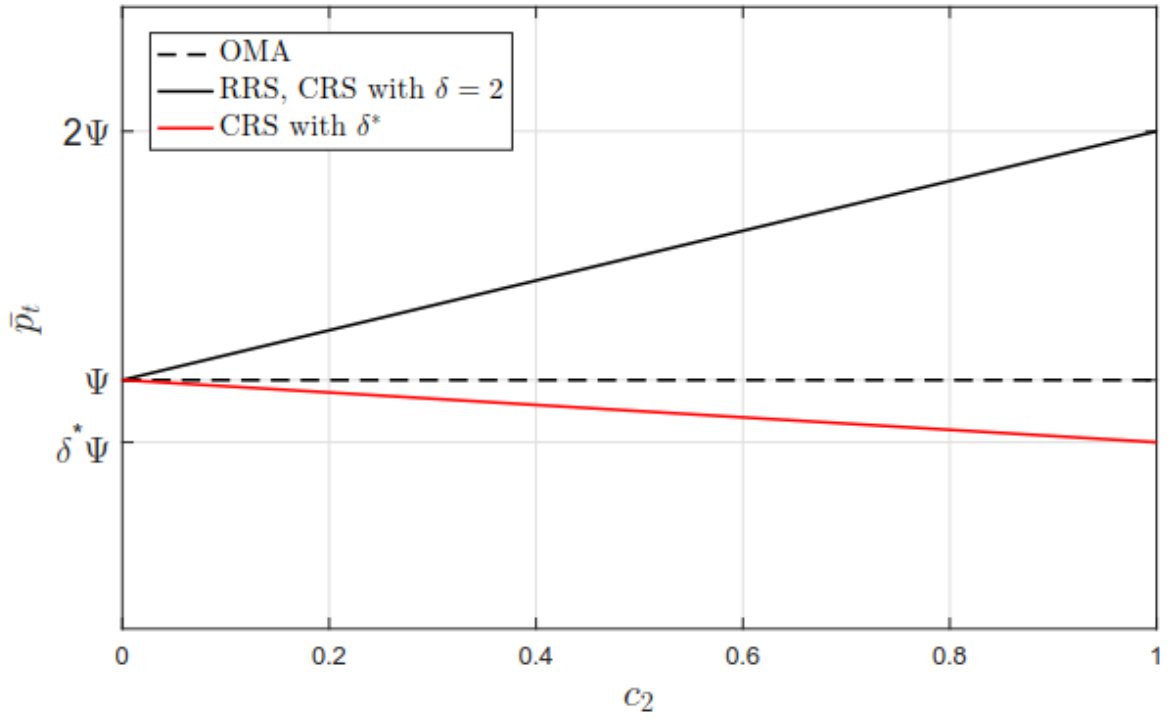
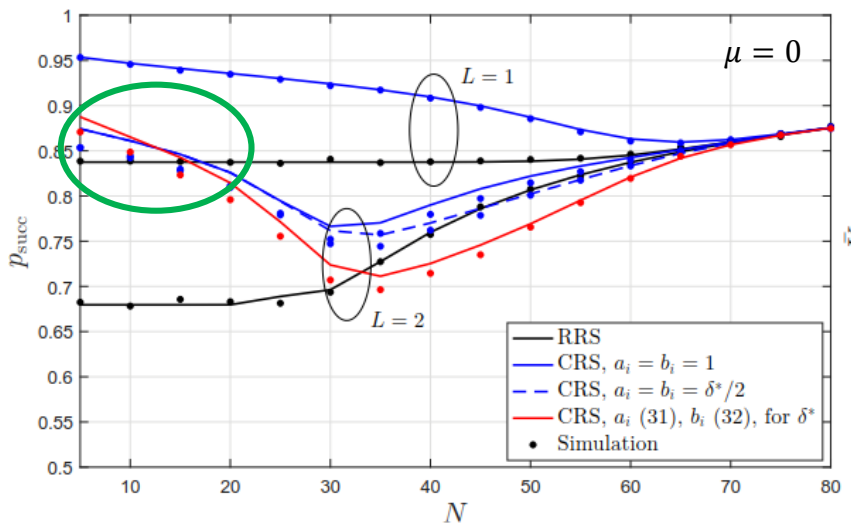
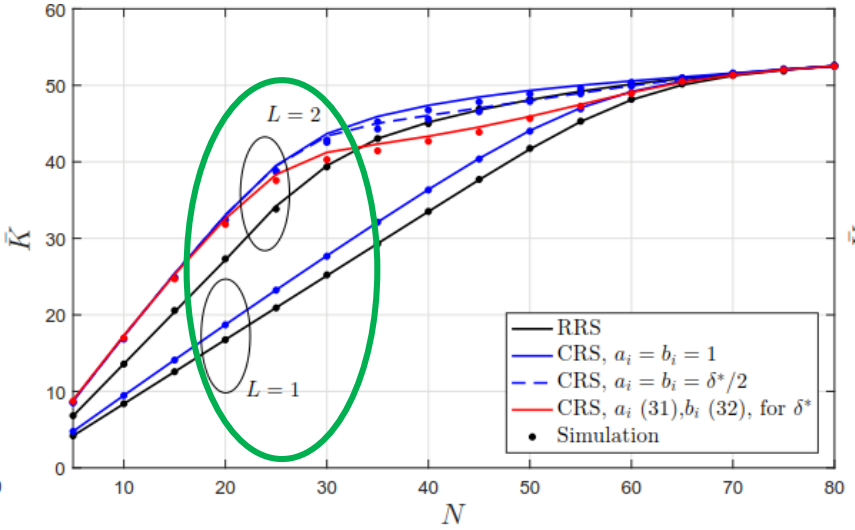


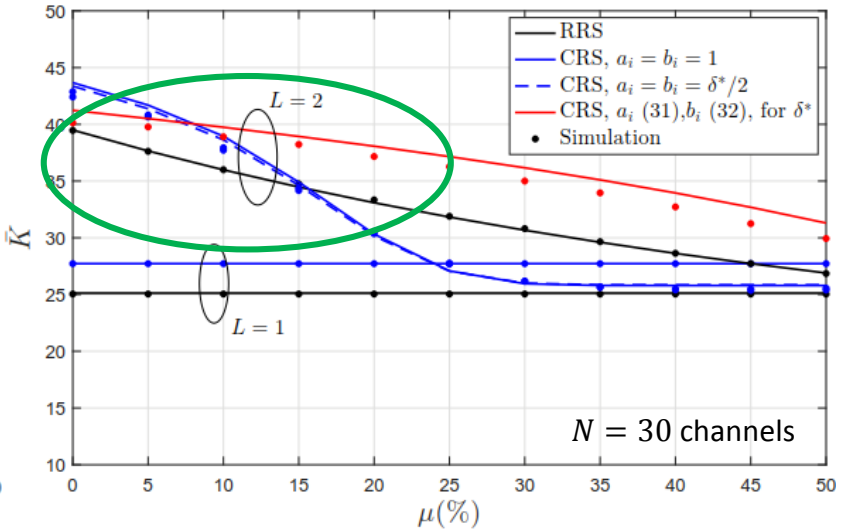
Fig. 4. Average transmit power per orthogonal channel with $c_0 = 0$.

O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4750-4765, July 2018. arXiv:1708.07691

Data Aggregation and Non-orthogonal Multiple Access: Scheduling







θ (SIR Threshold)	1
α	3.6
$\bar{m}$ (MTDs)	60
λ_a	$39.81/km^2$
R_a (aggregator radius)	40 m

- Power constraints on the MTDs sharing the same channel - fair coexistence with OMA
- Power control coefficients both MTDs can perform with similar reliability
- Lower average power consumption / orthogonal channel and / MTD,
- Hybrid scheme with CRS outperforms the OMA setup
- NOMA > OMA for some network configurations
  - Intra-cluster interference

O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli and M. Latva-aho, "Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4750-4765, July 2018. arXiv:1708.07691

Energy Efficient Statistical QoS Provisioning for MTC Networks

Mohammad Shehab

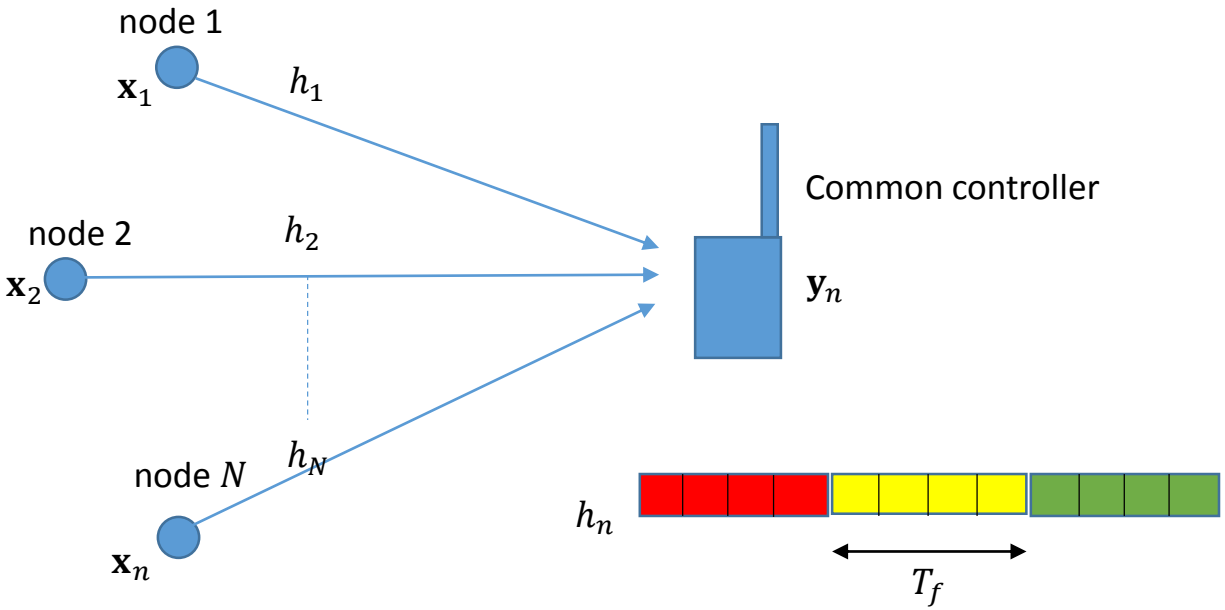
Effective Capacity with Short Messages

N nodes communicate with a common controller

$$\rho_i = \frac{\rho}{1+\rho\sum_s |h_s|^2} \approx \frac{\rho}{1+\rho\,(N-1)}$$

Nakagami-m block fading channel with block length T_f

$$f_z(z) = \frac{m^m z^{m-1}}{\Gamma(m)} e^{-mz}$$

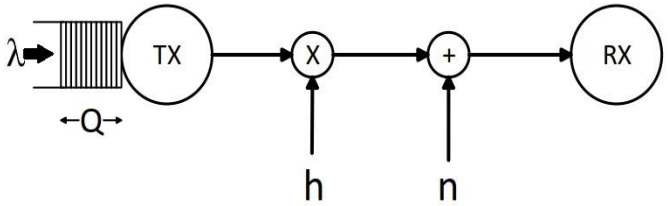


$$r = \log_2(1+\rho_i|h|^2) - \sqrt{\frac{1}{T_f}\left(1-\frac{1}{(1+\rho_i|h|^2)^2}\right)} Q^{-1}(\epsilon)\log_2(e) \quad \epsilon \in [0,1]$$

M. Shehab, H. Alves, M. Latva-aho, "Finite Blocklength Performance of Multi-Node Networks in Nakagami-m channels", accepted in *EURASIP JWCN* (July 2017).
M. Shehab, E. Dosti, H. Alves, M. Latva-aho, "On the Effective Capacity of MTC Networks in the Finite Blocklength Regime", *EUCNC 2017* (nominated for the best student paper award).

Effective Capacity with Short Messages

Effective capacity (EC) indicates the capability of communication nodes to exchange data with maximum rate and under a given QoS constraint.



$$C_E(\gamma,\theta) = -\lim_{t\to\infty}\frac{1}{\theta t}\log \mathbb{E}\{e^{-\theta S[t]}\}$$

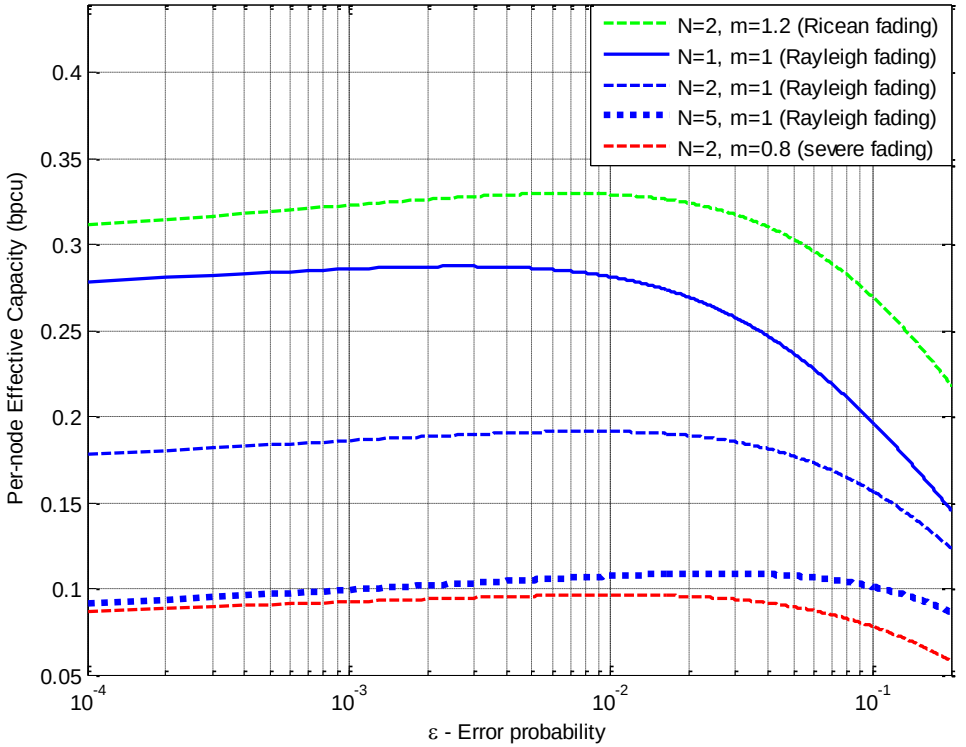
$$S[t] \triangleq \sum_{k=1}^{t} P[k]$$

- EC in bits per channel use (bpcu) is given by

$$EC(\rho_i,\theta,\epsilon) = -\frac{1}{T_f\theta}\ln\left(E_{z=|h|^2}\left[\epsilon+(1-\epsilon)e^{-T_f\theta r}\right]\right)$$

- The delay exponent θ determines the system's tolerance to certain delay bound according to

$$P_{out\_delay} = Pr(delay \geq D_{max}) \approx e^{-\theta.EC.D_{max}}$$



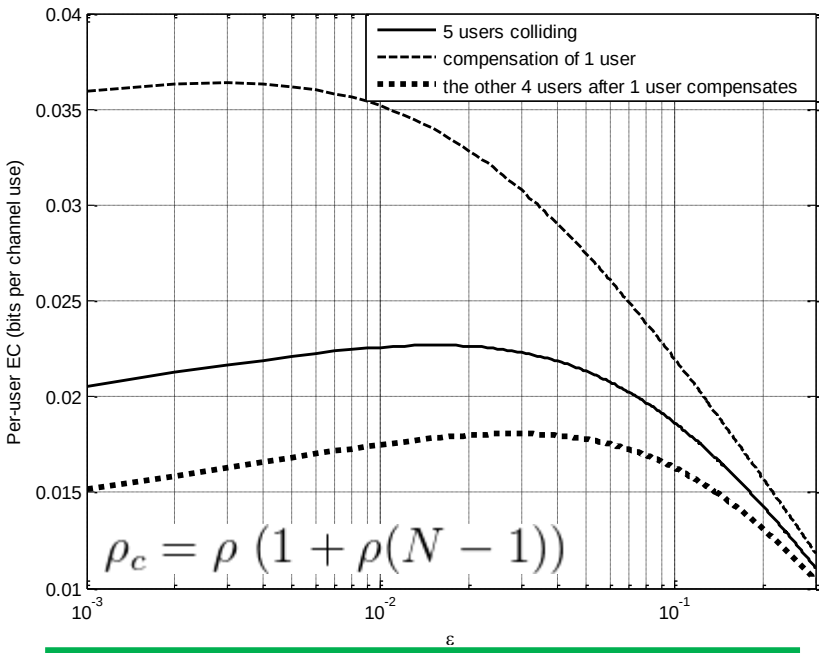
EC with $T_f = 1000, \theta = 0.01$ and $\rho = 2$

Increasing the number of nodes degrades the per-node EC due to interference.

M. Shehab, H. Alves, M. Latva-aho, "Finite Blocklength Performance of Multi-Node Networks in Nakagami-m channels", accepted in *EURASIP JWCN* (July 2017).
M. Shehab, E. Dosti, H. Alves, M. Latva-aho, "On the Effective Capacity of MTC Networks in the Finite Blocklength Regime", *EUCNC 2017* (nominated for the best student paper award).

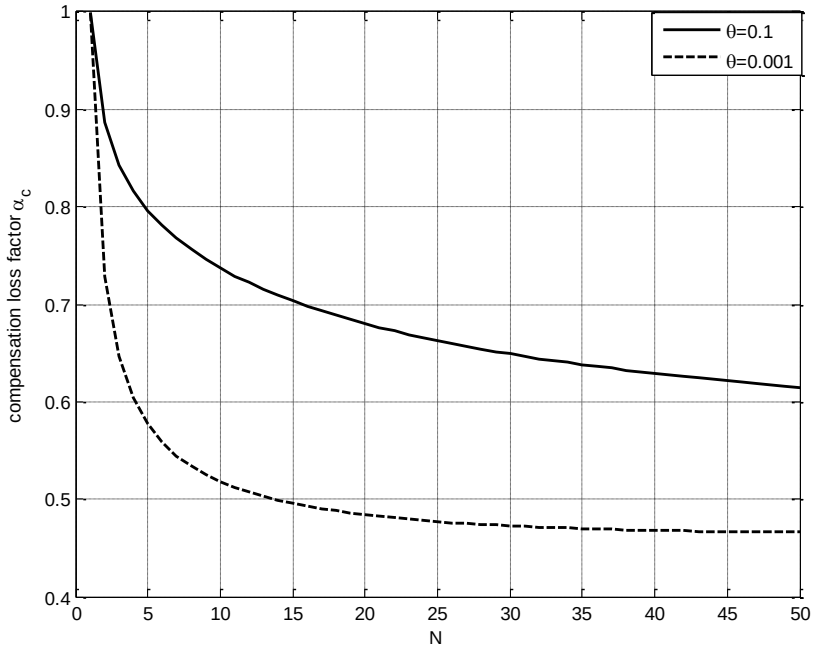
Effective Capacity with Short Messages

Compensation via power control



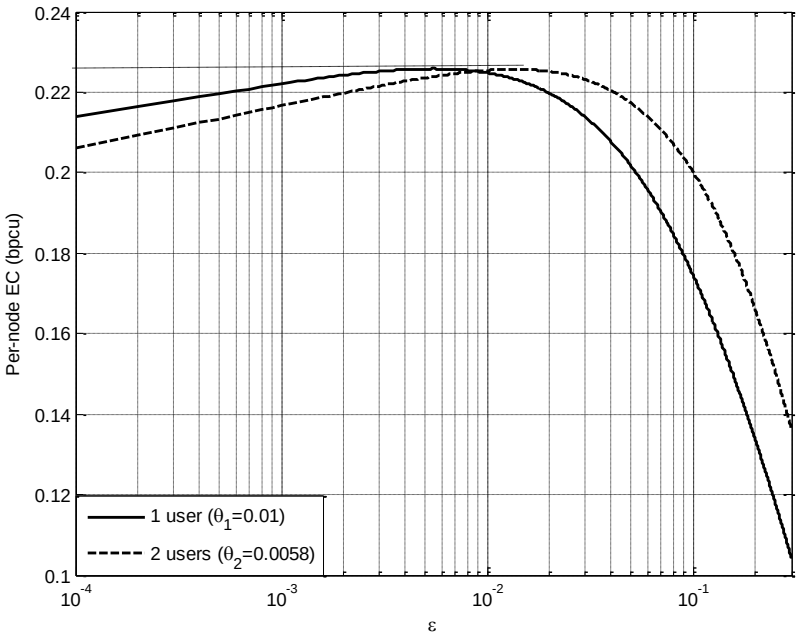
Benefits one user – degrades others

This causes more interference to other nodes degrading their EC



$$\alpha_c = \frac{EC_{Ry}(\rho_s, \theta, \epsilon_s^*)}{EC_{Ry}(\rho_i, \theta, \epsilon_i^*)}$$

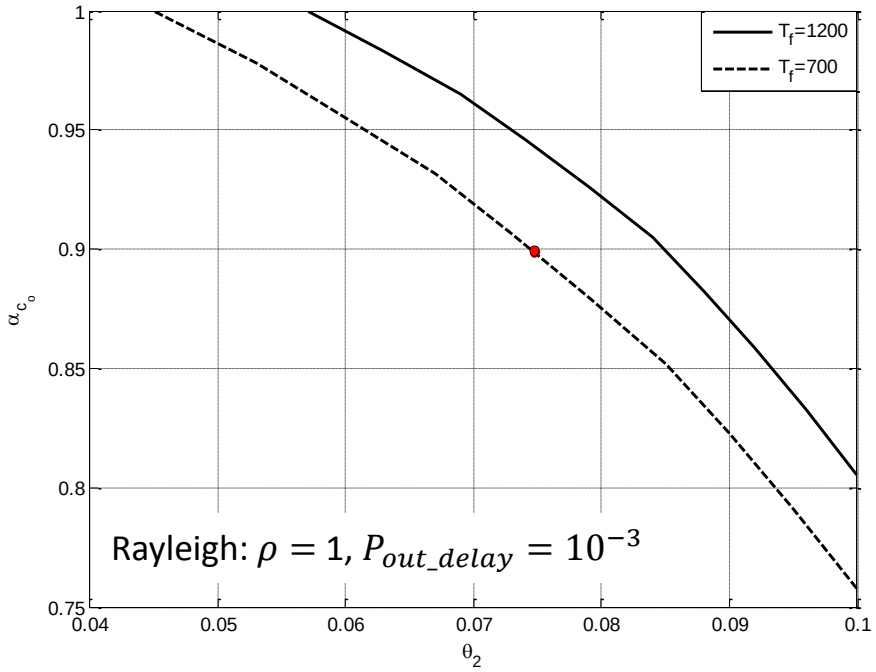
Power control is not convenient less stringent delay constraints (lower θ)



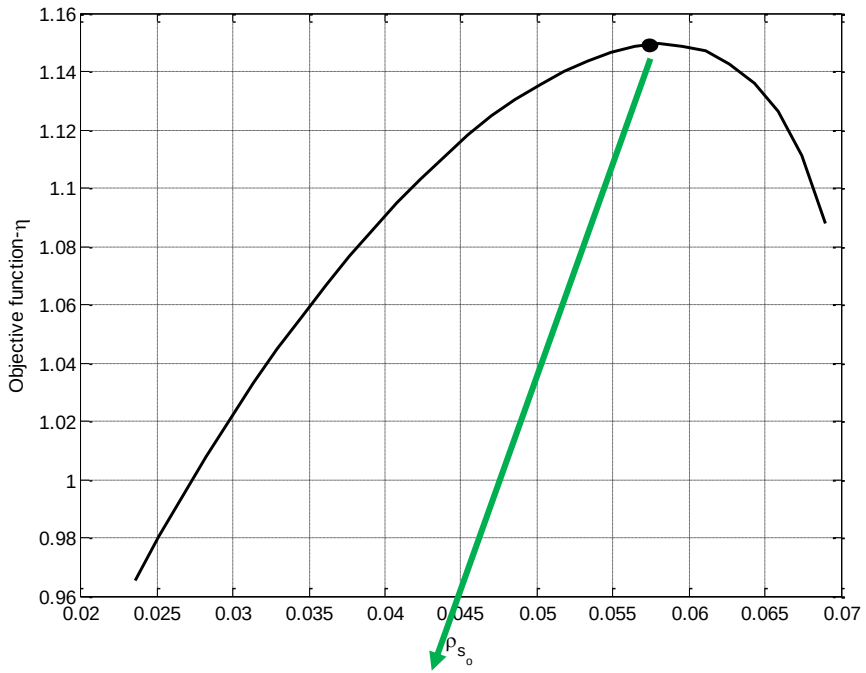
$$EC_{Ry}(\rho, \theta_1, \epsilon^*) = EC_{Ry}(\rho_i, \theta_2, \epsilon_2^*)$$

M. Shehab, H. Alves, M. Latva-aho, "Finite Blocklength Performance of Multi-Node Networks in Nakagami-m channels", accepted in *EURASIP JWCN* (July 2017).
M. Shehab, E. Dosti, H. Alves, M. Latva-aho, "On the Effective Capacity of MTC Networks in the Finite Blocklength Regime", *EUCNC 2017* (nominated for the best student paper award).

Effective Capacity with Short Messages



Rayleigh: $T_f = 1000, \theta = 0.1$ and $\rho = 2$



$$\eta_{max} = \max_{\theta_2 \geq 0} \ \eta_\alpha \alpha_{c_o} + \eta_\theta \theta_2$$
$$s.t \ \ \rho_s \leq \rho_{s_o} \leq \rho_i$$

η_α: compensation loss priority factor

η_θ: delay priority factor

ρ_{s_o}: SINR of other nodes (set s)

- Operational point:
  $\alpha_{c_o} = 0.9, \theta_2 = 0.075$

Before compensation : $D_{max} = 2500\ sps$

After compensation : $D_{max} = 2500\ sps$

- 0.9 loss in EC of other nodes.
- Nearly no loss in delay bound as restoring EC compensates for the decrease in θ.

- For $\eta_\alpha = 1$ and $\eta_\theta = 4$ (means delay constraint is of high priority),
  Optimum OP: $\rho_{s_o} = 0.057$, $\alpha_{c_o} = 0.94$ (6% loss of EC of other nodes),
  $\theta_2 = 0.053$
  - SNR of compensating user is raised to 8.

M. Shehab, H. Alves, M. Latva-aho, "Finite Blocklength Performance of Multi-Node Networks in Nakagami-m channels", accepted in *EURASIP JWCN* (July 2017).

M. Shehab, E. Dosti, H. Alves, M. Latva-aho, "On the Effective Capacity of MTC Networks in the Finite Blocklength Regime", *EUCNC 2017* (nominated for the best student paper award).

Effective Energy Efficiency with Short Messages

Linear model

Rayleigh fading

Effective Energy Efficiency (EEE) is $\eta_{ee} = \dfrac{-\frac{1}{n\theta}\log\left(\mathbb{E}_Z\left[\epsilon + (1-\epsilon)e^{-n\theta r}\right]\right)}{\zeta\rho + P_c}$

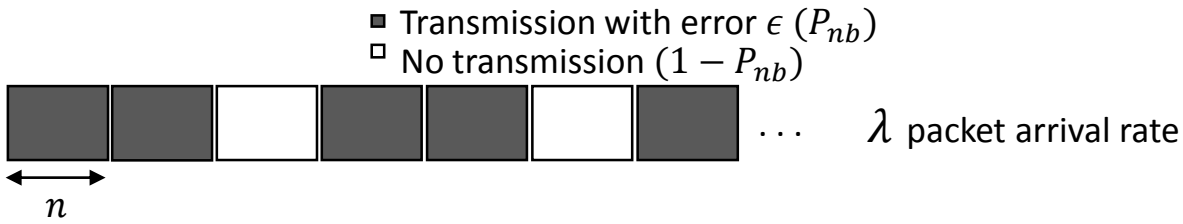
$$\eta_{ee}(\rho,\theta,\epsilon) \approx -\frac{\log\left[\epsilon + (1-\epsilon)\,\mathcal{J}\right]}{n\theta\left(\zeta\rho + P_c\right)}$$

ξ inverse drain efficiency of the transmit amplifier

P_c the hardware power dissipated in circuit.

$$\mathcal{J} = e^{\frac{1}{\rho}}\rho^{\alpha}\left[\left(\frac{\beta^2}{2}+\beta+1\right)\Gamma\left(\alpha+1,\frac{1}{\rho}\right) - \left(\frac{\beta^2}{2}+\beta\right)\frac{\Gamma\left(\alpha-1,\frac{1}{\rho}\right)}{\rho^2}\right],$$

$$\alpha = \frac{-\theta n}{\log 2}, \beta = \theta\sqrt{n}Q^{-1}(\epsilon)\log_2 e, \text{ and } \gamma = \sqrt{(1-\frac{1}{(1+\rho z)^2}}).$$



λ packet arrival rate

$$\eta_{ee} = \frac{-\frac{1}{n\theta}\log\left[\epsilon + (1-\epsilon)\,\mathcal{J}\right]}{\frac{\lambda}{\mathbb{E}[r]}\zeta\rho + P_c}$$

$$\max_{\rho\geq 0,\theta\geq 0} \eta_{ee} = \frac{-\frac{1}{n\theta}\log\left[\epsilon + (1-\epsilon)\,\mathcal{J}\right]}{P_{nb}\zeta\rho + P_c},$$

$$s.t \quad C_e(\rho,\theta,\epsilon) \geq \lambda$$

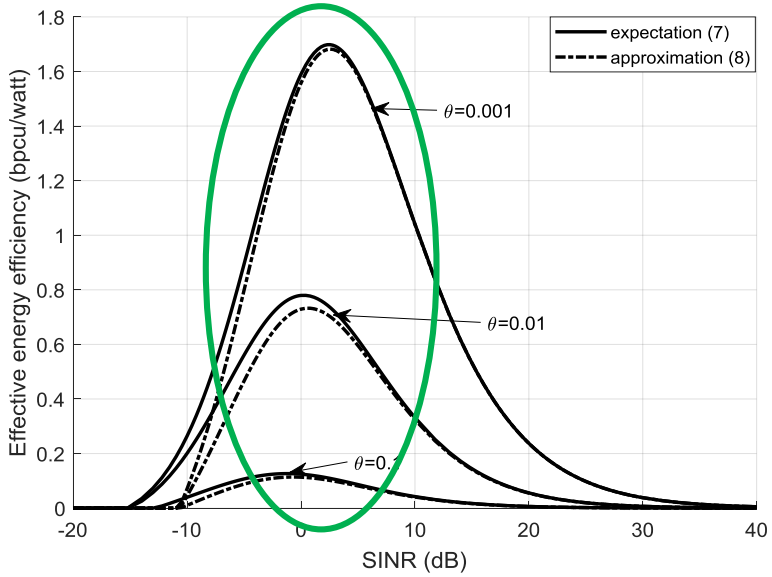
$$P_{nb}e^{-\theta\lambda\delta} \leq \Lambda$$

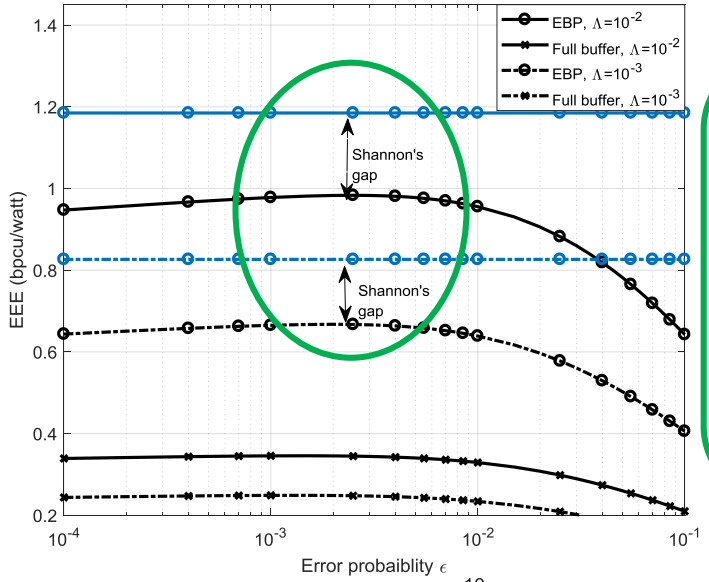
$$\rho \leq \rho_{max}$$

$$\epsilon \leq \epsilon_t$$

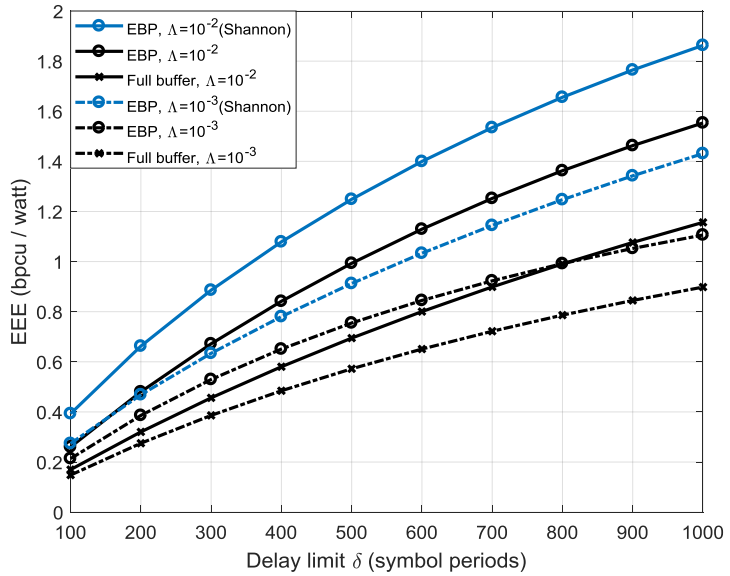
M. Shehab, E. Dosti, H. Alves, M. Latva-aho, "On the Eective Energy Eciency of Ultrareliable Networks in the Finite Blocklength Regime", ISWCS'2017, pp.1-6, Bologna, Italy, Aug. 2017. ISWCS'2017 Best Student Paper Award

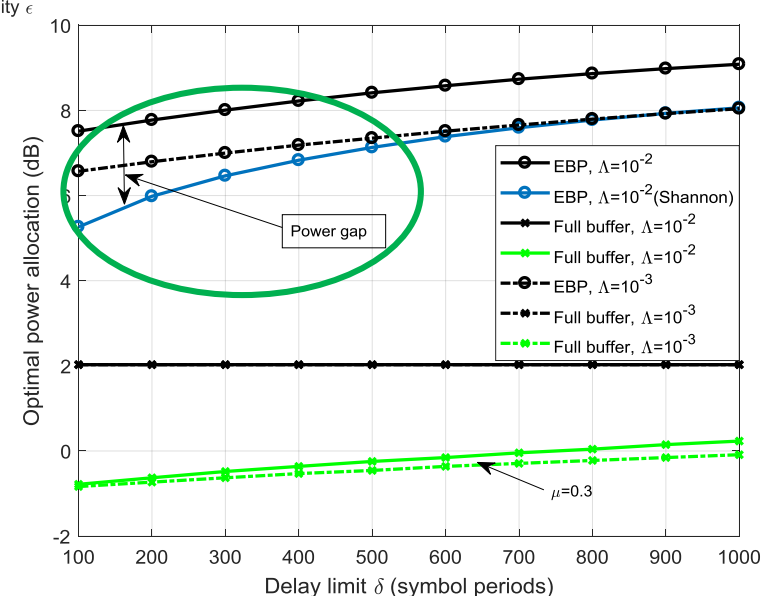
Effective Energy Efficiency with Short Messages





- EEE is concave in ϵ and quasi-concave in SNR
- EEE increases when extending the delay δ and relaxing the delay outage probability Λ.
- Shannon's model underestimates the optimum power allocation.
- The optimum power decays when the arrival rate declines





$n = 500, , \epsilon = 10^{-3}, P_c = 0.2, \zeta = 0.2, \rho_{max} = 10$ dB and $\lambda = 1$.

M. Shehab, E. Dosti, H. Alves, M. Latva-aho, "On the Effective Energy Effciency of Ultrareliable Networks in the Finite Blocklength Regime", ISWCS'2017, Bologna, Italy, Aug. 2017. ISWCS'2017 Best Student Paper Award

What about LPWANs?

Arliones Hoeller (UFSC, Brazil)

A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, “Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity,” in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

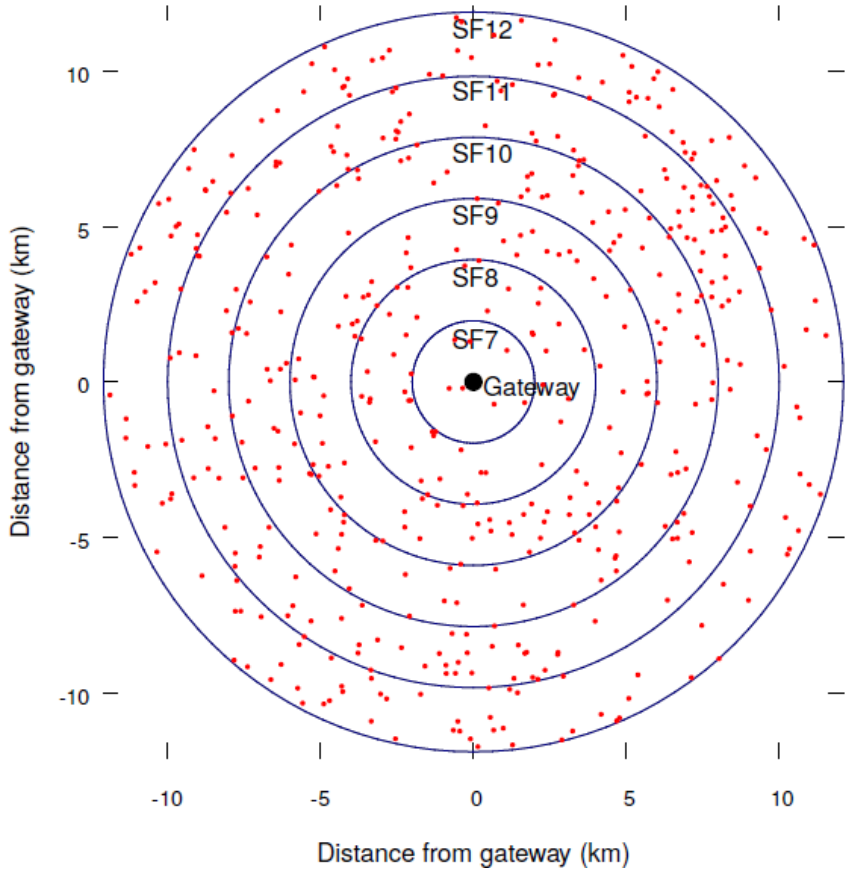
What is LPWAN?

- **Coverage < 1 km**
  - IEEE 802.15.4,
  - IEEE P802.11ah,
  - Bluetooth/LE
  - Telensa
- **Coverage > 1km**
  - LoRaWAN
  - Sigfox
  - Ingenu

- Low Power Wide Area Networks
  - Coverage of large area
  - Limited Power/Energy
    - Battery constrained
  - Short payloads/messages
  - (Bi)Directional TX – Uplink/Downlink
  - Robustness to interference
  - Security
  - Capacity - #of users

A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, “Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity,” in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Enhanced Reliability of LORAWAN

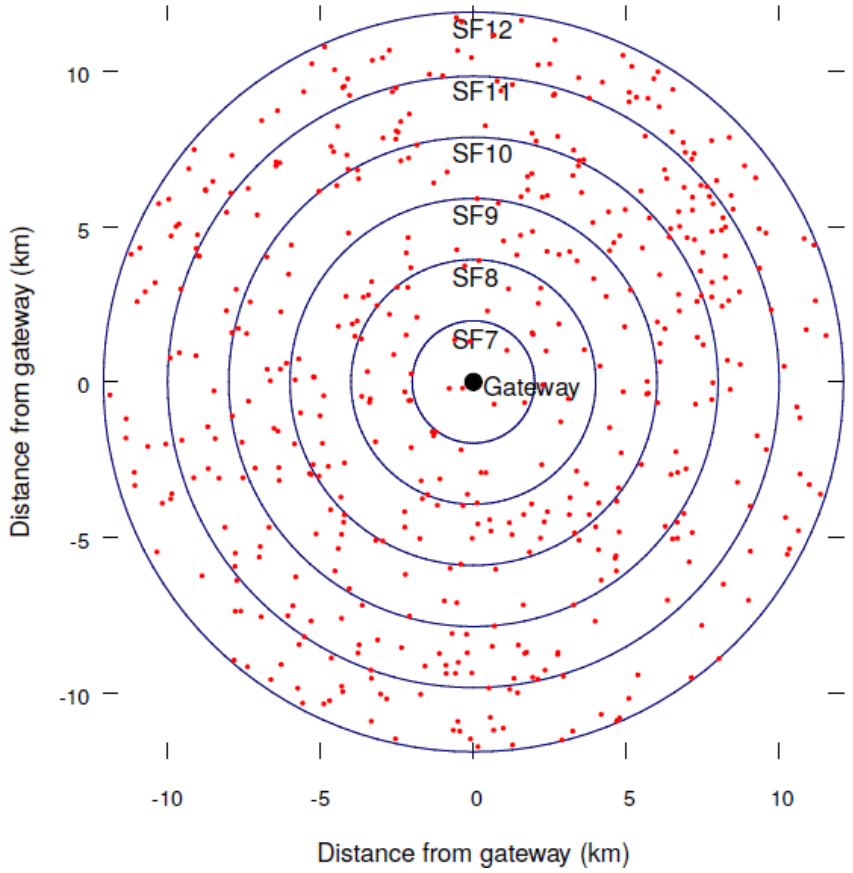


- Nodes are uniformly distributed around a gateway.
- ALOHA-like transmissions duty cycle
- Spreading Factor assigned to nodes according to their distance from the gateway increasing every 2km.
- All nodes transmit with the same power
- Model captures:
  - Interference at the gateway
  - From same SF

$$H_1 = \mathbb{P}[\text{SNR} \geq q_{\mathbb{S}} | d_1]$$
$$= \mathbb{P}\left[|h_1|^2 \geq \frac{\mathcal{N} q_{\mathbb{S}}}{\mathcal{P}_1 g(d_1)}\right] = \exp\left(-\frac{\mathcal{N} q_{\mathbb{S}}}{\mathcal{P}_1 g(d_1)}\right)$$

A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, "Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity," in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Enhanced Reliability of LORAWAN

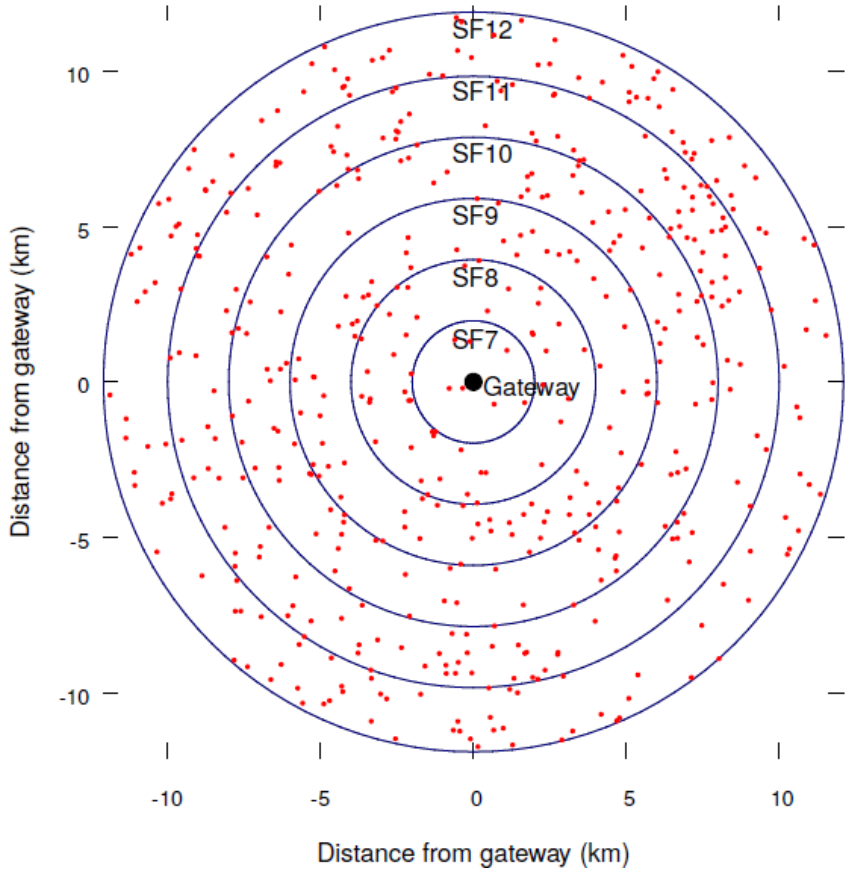


- Nodes are uniformly distributed around a gateway.
- ALOHA-like transmissions duty cycle
- Spreading Factor assigned to nodes according to their distance from the gateway increasing every 2km.
- All nodes transmit with the same power
- Model captures:
  - Interference at the gateway
  - From same SF

$$
\begin{aligned}
Q_1 &= \mathbb{P}\left[\frac{|h_1|^2 g(d_1)}{|h_{k^*}|^2 g(d_{k^*})} \geq 4 \;\middle|\; d_1 \right] \\
&= \mathbb{E}_{|h_1|^2}\left[\mathbb{P}\left[X_{k^*} < \frac{|h_1|^2 g(d_1)}{4} \;\middle|\; |h_1|^2, d_1 \right] \right] \\
&= \mathbb{E}_{|h_1|^2}\left[F_{X_{k^*}}\left(\frac{|h_1|^2 g(d_1)}{4} \right) \right] \\
&= \int_0^\infty e^{-z} F_{X_{k^*}}\left(\frac{z g(d_1)}{4} \right) \mathrm{d}z
\end{aligned}
$$

A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, “Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity,” in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Enhanced Reliability of LORAWAN



- Nodes are uniformly distributed around a gateway.
- ALOHA-like transmissions duty cycle
- Spreading Factor assigned to nodes according to their distance from the gateway increasing every 2km.
- All nodes transmit with the same power
- Model captures:
  - Interference at the gateway
  - From same SF

Message Replication

$$H_{1,M} = 1 - (1 - H_1)^M$$

$$Q_{1,M} = 1 - (1 - Q_1)^M$$

ARQ
M repetitions

Multiple Antennas

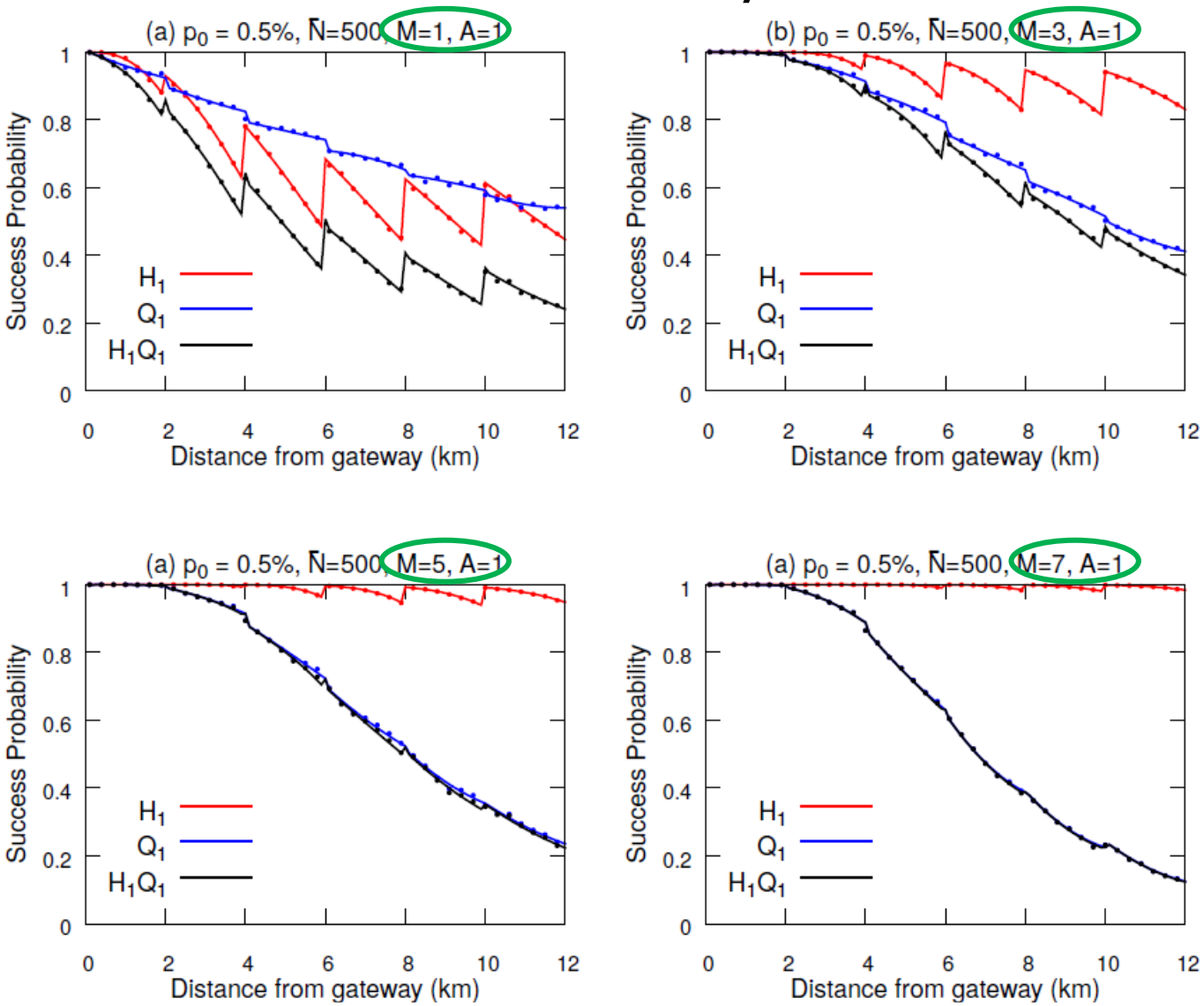
$$H_{1,A} = 1 - (1 - H_1)^A$$

$$Q_{1,A} = \mathbb{P}\left[\max_{i=1,\cdots,A} \mathrm{SIR}_i^* < 4 \middle| d_1\right]$$

Mutiple Gateway
Multiple Connectivity

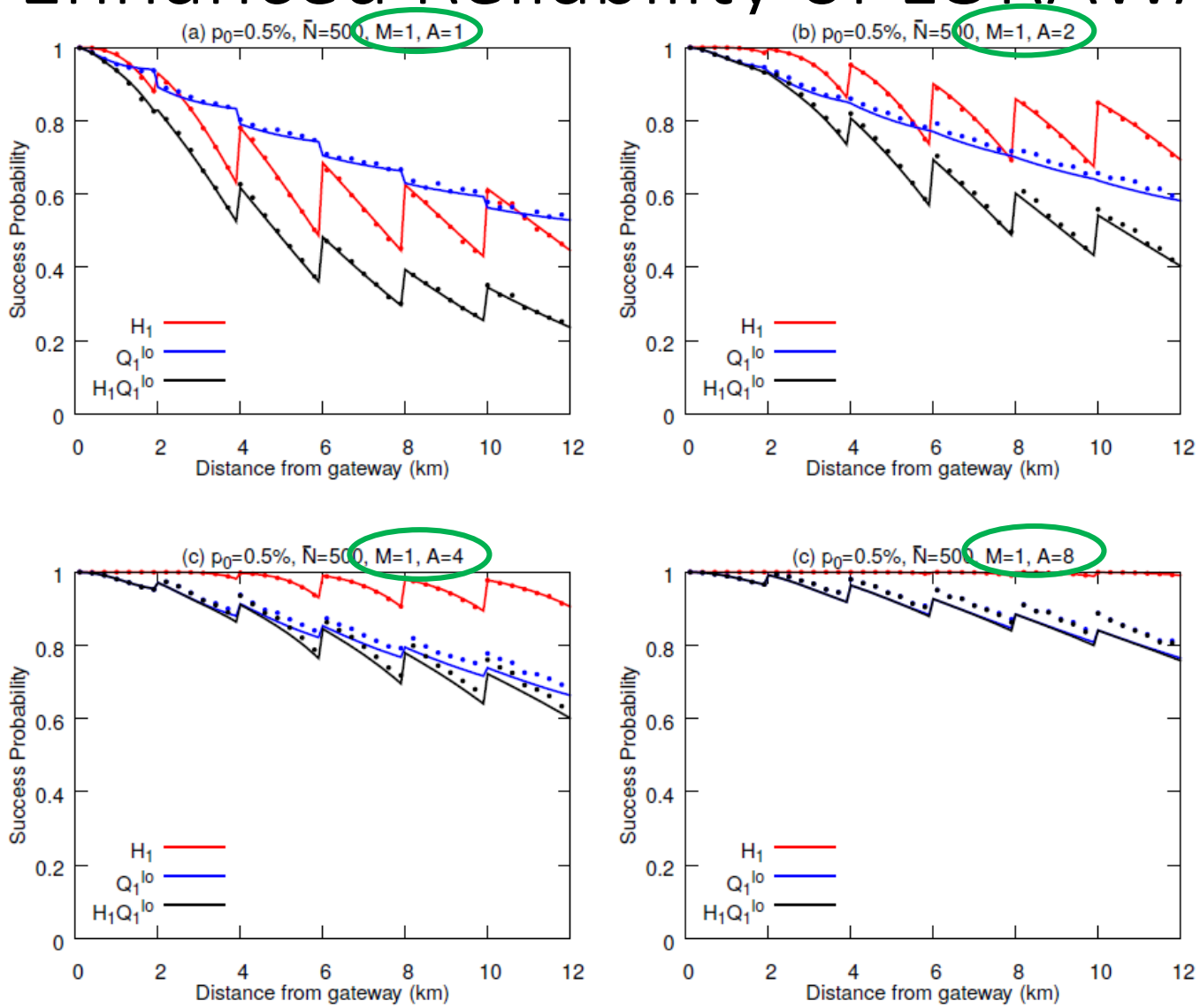
A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, “Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity,” in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Enhanced Reliability of LORAWAN



A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, "Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity," in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Enhanced Reliability of LORAWAN



A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, "Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity," in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Enhanced Reliability of LORAWAN

Optimum M* for different configurations of network density and number of Antennas.

		$\bar{N} = 500$		$\bar{N} = 1000$		$\bar{N} = 1500$	
p_0	**A**	M^*	$\rho_c[H_1 Q_1]$	M^*	$\rho_c[H_1 Q_1]$	M^*	$\rho_c[H_1 Q_1]$
0.1%	**1**	8	99.7%	5	91.0%	4	79.1%
	2	4	100.0%	5	96.6%	4	89.2%
	4	3	100.0%	5	99.5%	3	95.8%
	8	2	100.0%	3	100.0%	4	99.4%
0.5%	**1**	3	59.2%	2	33.0%	2	20.5%
	2	3	73.3%	2	47.1%	1	33.3%
	4	2	85.6%	1	61.6%	1	49.1%
	8	2	94.0%	1	76.5%	1	64.2%

A. Hoeller, R. D. Souza, O. L. A. Lopez, H. Alves, M. d. N. Neto and G. Brante, "Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity," in IEEE Access, vol. 6, pp. 32820-32829, 2018. [Online] Available (Open Access)

Thanks!

Let’s go for some coffee and be back for part 2:URLLC!

ISCWS'18

T5: Machine-Type Communications: *from massive connectivity to URLLC*

part 2

ASSOC. PROF. JIMMY J. NIELSEN (JJN@ES.AAU.DK)

CONNECTIVITY SECTION, AALBORG UNIVERSITY, DENMARK

This work has partly been performed in the framework of the horizon 2020 project ONE-5G (ICT- 760809) receiving funds from the european union. The author would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this work are those of the authors and do not necessarily represent the project.

ONE5G

AALBORG UNIVERSITY
DENMARK

What is URLLC?

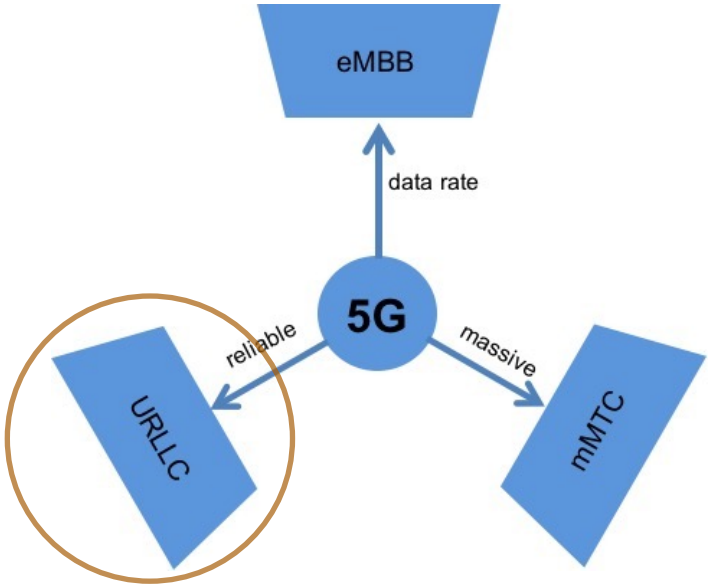
Ultra-Reliable Low Latency Communication

A key feature of 5G is support for URLLC.

Two parts – with different main focii:

- Ultra-Reliable Communications (URC)
- Low Latency Communications (LLC)

Typically more difficult to achieve simultaneously than satisfying just one at a time.



URLLC will enable new use cases with:

- packet error rate of 10^{-5} down to 10^{-9}
- end-to-end latency of few ms to fraction of ms

Latency CDF

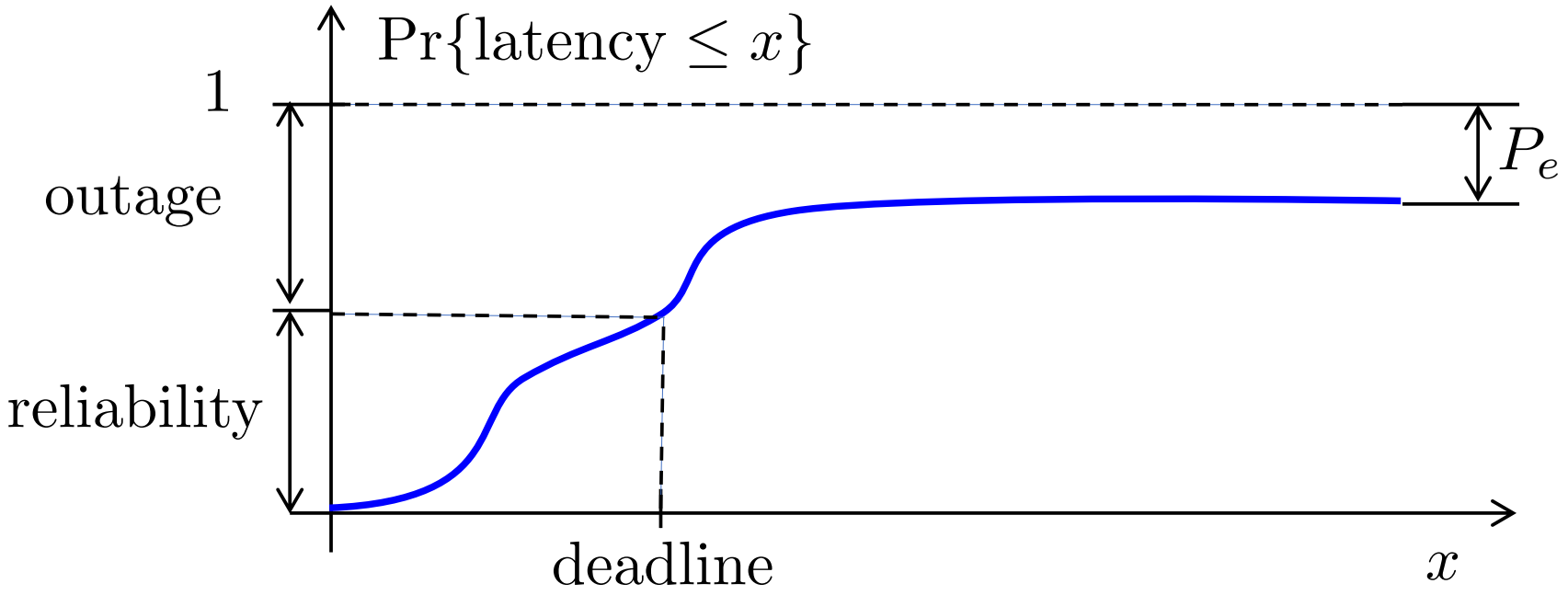
Latency vs. reliability

Application deadline → experienced outage

P_e is outage due to lost packets (fading, collisions w.o. reTX), infrastructure failures, etc.

URC: push up

LLC: push left



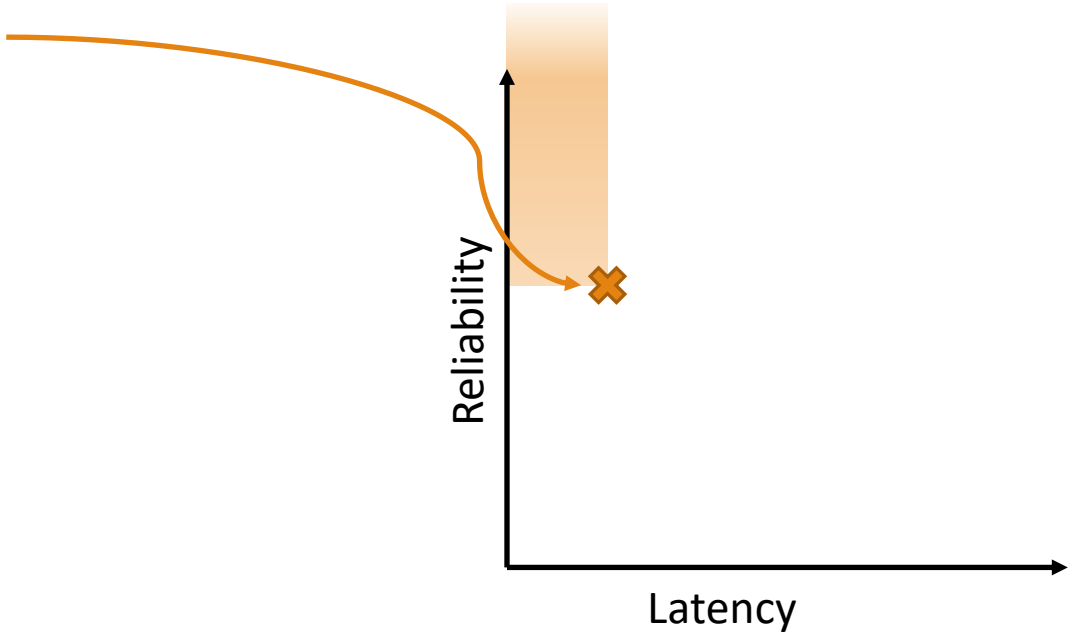
Courtesy: Erik Ström

URLLC design target in 3GPP

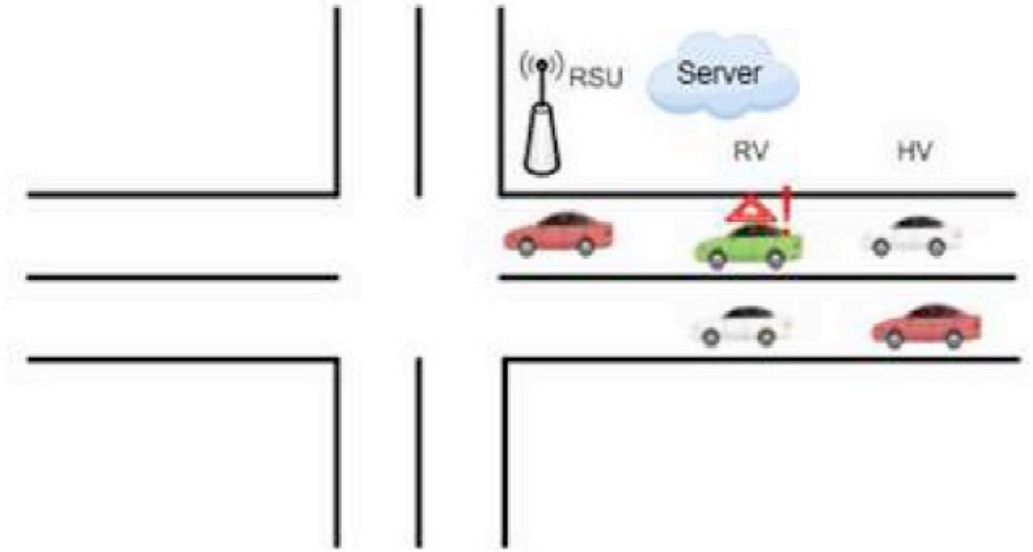
3GPP study item for Next Generation Radio Access Technologies (TR 38.913):

general reliability requirement:

- 32 bytes within 1 ms at BLER = 10^{-5}
  - (user plane latency)
- Just a single point, but
  - R>1-10^{-5} also fulfills requirements
  - L<1 ms also fulfills requirements



V2X use case

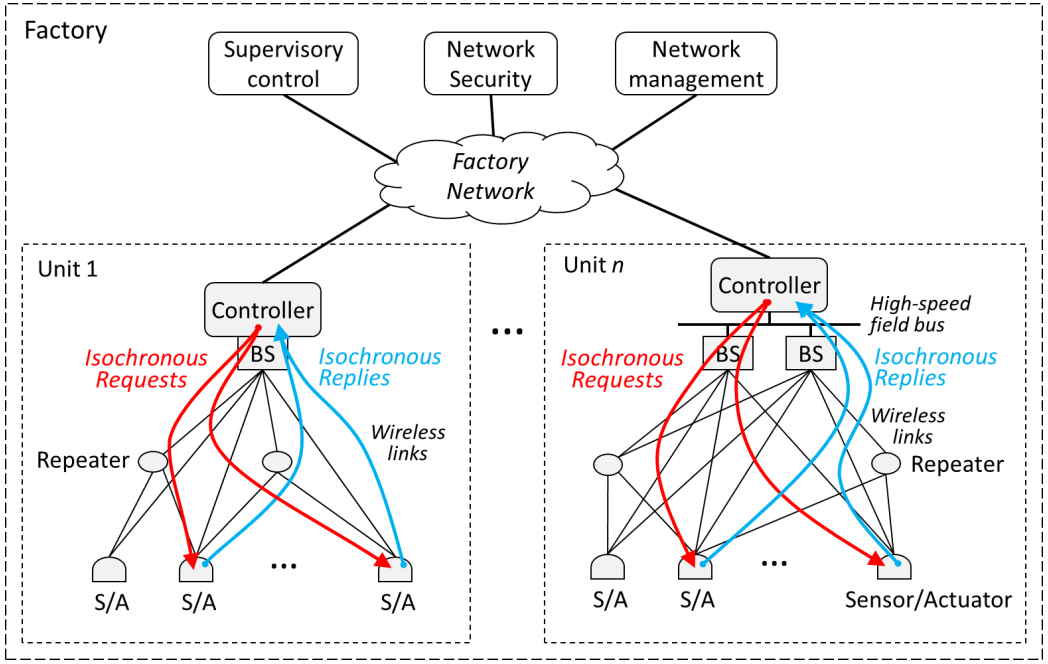


1. Assisted driving aided by roadside infrastructure
   - RSU improves coverage and enables low-latency.
   - Car-to-car communication.
2. Cooperative driving between nearby vehicles
   - No RSUs
   - communication is purely car-to-car based, with the aid of the network infrastructure (wherever available).
3. Tele-operated driving
   - Cellular URLLC for control data transmission in downlink
   - Reliable low-latency video (plus other sensor data) transmission in the uplink.

Service KPIs	Service 1	Service 2	Service 3	Comments
U-plane maximum UL/DL radio latency (ms)	0.5 ms	0.1 ms	2 ms	Taken as 1/10th of the end-to-end maximum latency. Radio protocol layer in which it is measured should be specified.
U-plane maximum E2E latency (ms)	5 ms	1 ms	20 ms	Taken from [22.886].
C-plane maximum UL/DL radio latency (ms)	10 ms	2 ms	10 ms	Max. time for C-plane state transition to “connected state”. Taken from [38.913], reduced for Service #2.
U-plane maximum DL/UL radio packet loss (%)	0.001%	0.001%	0.001% or lower	Taken as (100 - reliability)%
U-plane reliability	99.999%	99.999%	99.999 % or higher, up to 250 km/h.	Probability that IP packets are correctly received within the latency time. Taken from [22.886].

ONE5G D2.1

Industry 4.0 use case



UE KPIs	KPIs' Targets	Comments
Reliability	URLLC: 99.999% for one transmission of a packet of length 32 bytes with a user plane latency of 1ms	[38.913] The foreseen reliability is inadequate for most of the representative services of this use case.
U-Plane average latency (ms)	URLLC: 0.5 ms	[38.913]

Example: Motion control

- Controllers periodically issue control-commands to actuators, typically machines with moving parts, like machine tools, printing machines, paper mills and textile machines.
- The communications in this service are assumed to be isochronous. The cycle* times are of the order of milliseconds, putting extreme requirements on the communications in terms of latency.
- Controlled processes may incur risks to the factory personnel or overall production, which puts extreme requirements on reliability and availability of communications (>1-10^{-6}).

*Cycle time is the time from execution of the command until the feedback from the actuator is received, which includes all processing and latencies on the air interface and actuation times.

ONE5G D2.1, 3GPP TR 38.913

Smart grid use case

Wide Area Situational Awareness (WASA)

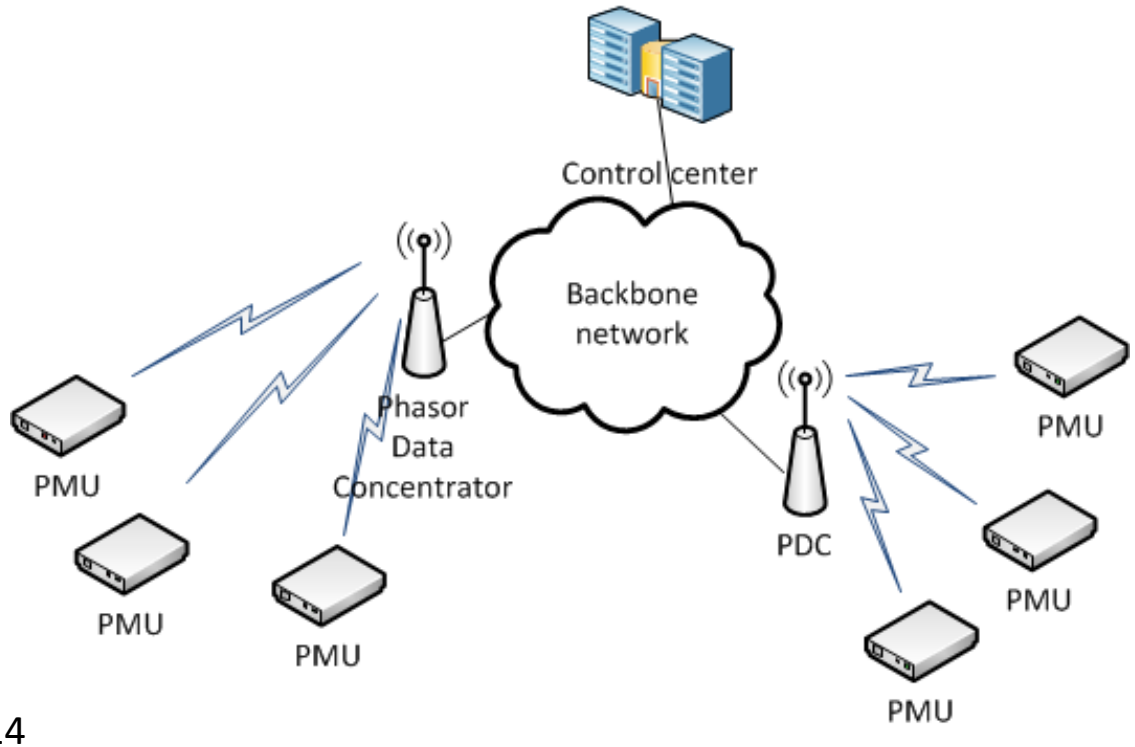
- Hundreds of PMUs should be deployed in the distribution grid in order to obtain the high resolution image of the grid
- IEEE C37.118 defines report (frame) structure, reporting frequencies and delays
- 50-100 Hz reporting frequency per PMU

Traffic type:

- Periodic, frequent traffic

Requirements:

- Latency: 20 – 200 ms
- Data rates: 600-1500 kbps
- Reliability: 99.999 – 99.9999%
- Security: High

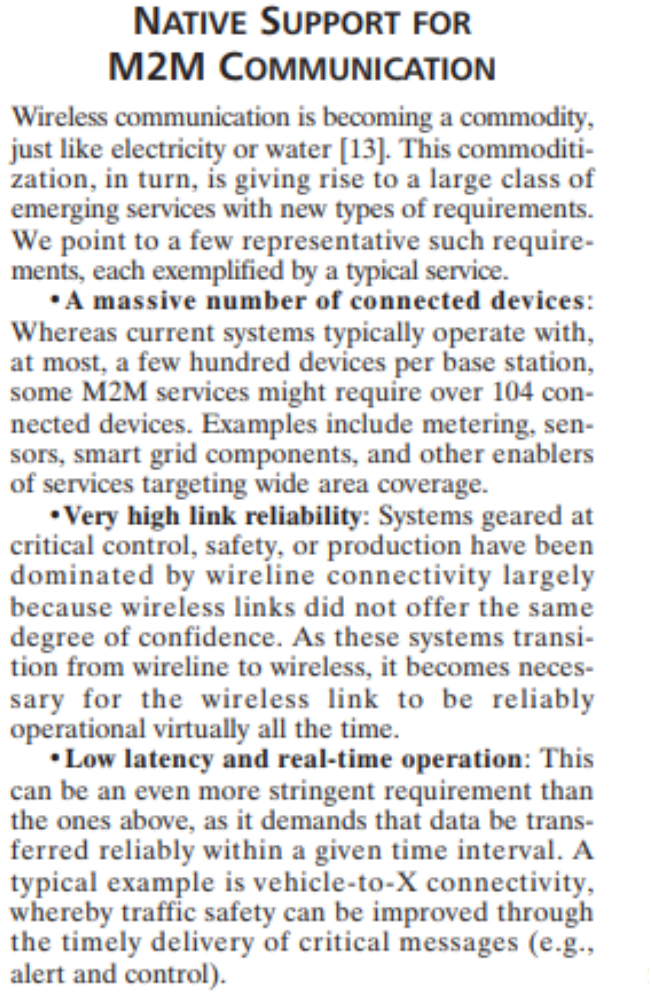


IEEE, “IEEE vision for smart grid communications: 2030 beyond”, 2014

State of the art

URLLC requirements for 5G were outlined in, e.g.:

- F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," in *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74-80, February 2014.
- J. G. Andrews *et al*., "What Will 5G Be?," in *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065-1082, June 2014.

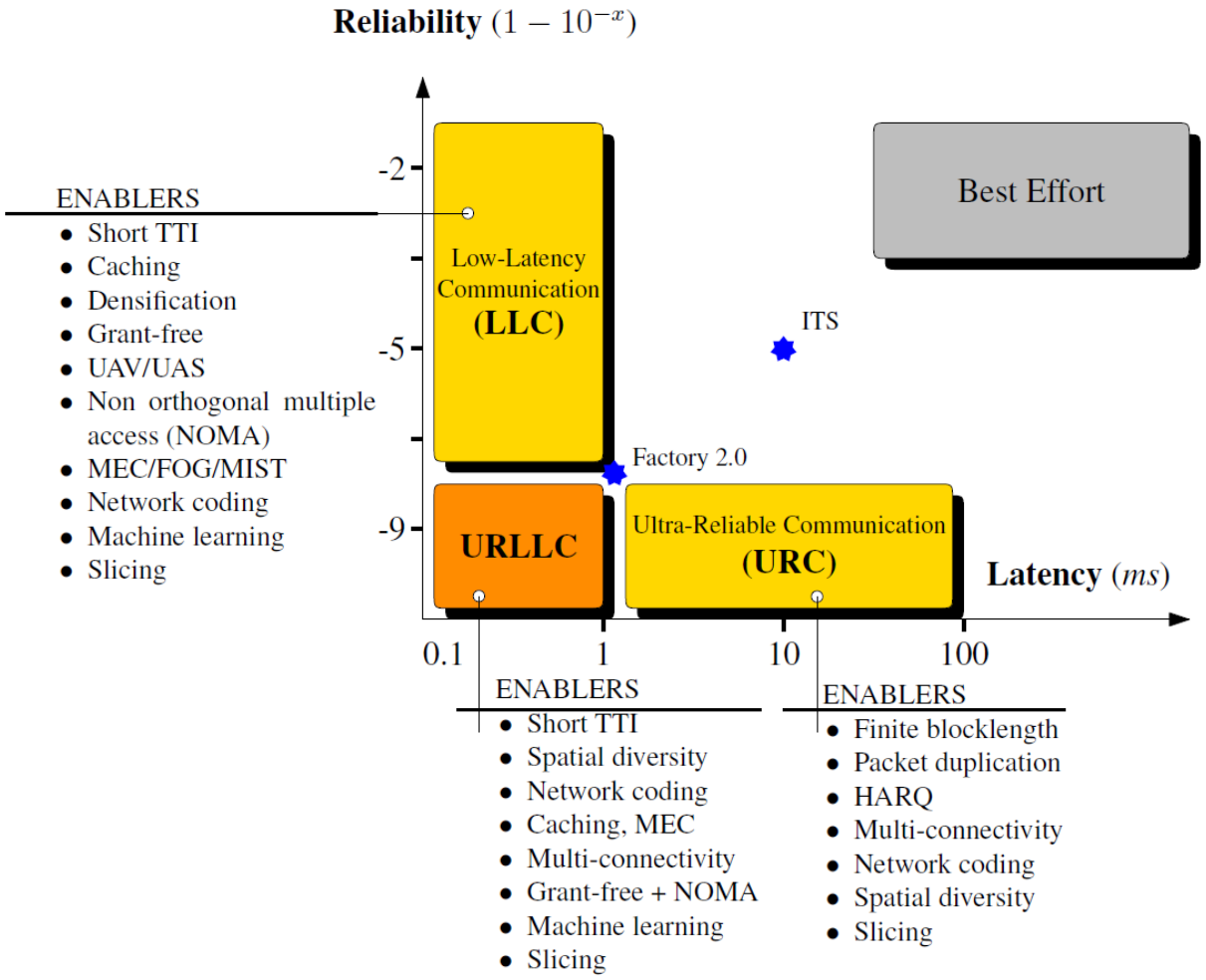


NATIVE SUPPORT FOR M2M COMMUNICATION

Wireless communication is becoming a commodity, just like electricity or water [13]. This commoditization, in turn, is giving rise to a large class of emerging services with new types of requirements. We point to a few representative such requirements, each exemplified by a typical service.

- **A massive number of connected devices**: Whereas current systems typically operate with, at most, a few hundred devices per base station, some M2M services might require over 104 connected devices. Examples include metering, sensors, smart grid components, and other enablers of services targeting wide area coverage.
- **Very high link reliability**: Systems geared at critical control, safety, or production have been dominated by wireline connectivity largely because wireless links did not offer the same degree of confidence. As these systems transition from wireline to wireless, it becomes necessary for the wireless link to be reliably operational virtually all the time.
- **Low latency and real-time operation**: This can be an even more stringent requirement than the ones above, as it demands that data be transferred reliably within a given time interval. A typical example is vehicle-to-X connectivity, whereby traffic safety can be improved through the timely delivery of critical messages (e.g., alert and control).

URLLC, URC, LLC enablers



Latency distinctions:

- Uplink/downlink transmission
- End-to-end latency
  - Over-the-air, queueing, processing
- User plane latency
  - Assuming UE in RRC_active, time to deliver packet
- Control plane latency
  - From idle state to RRC_active

Note: URLLC may start from 5 nines reliability.

Bennis, M., Debbah, M. and Poor, H.V., 2018. Ultra-reliable and low-latency wireless communication: Tail, risk and scale. *arXiv preprint arXiv:1801.01270*.

Selected URLLC enablers

Introduction to:

1. How is short TTI and URLLC transmissions achieved in 5G?
2. Grant free uplink access via semi-persistent scheduling
3. Massive MIMO
4. Mobile Edge Computing
5. Multi-Connectivity
6. Network Slicing

Short TTI in 5G

Short TTI is achieved through larger subcarrier spacing

- Leads to reduction of symbol time

Denoted as ”Numerology scaling”

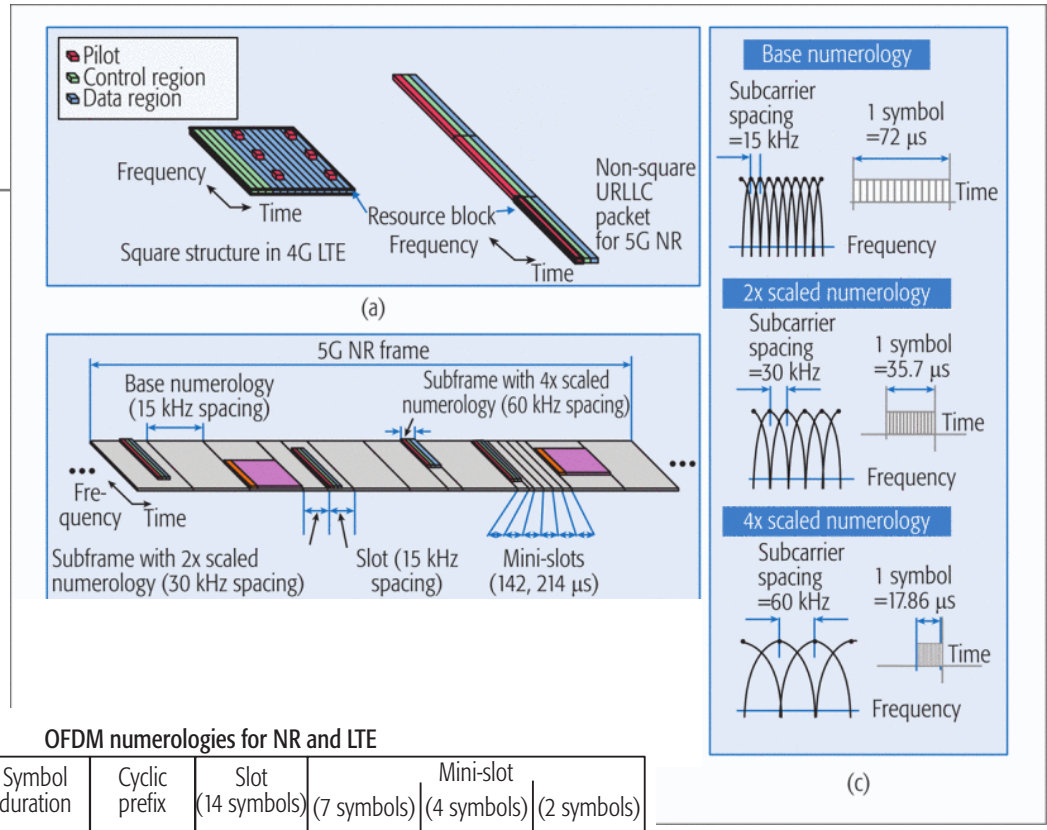
Slot lengths:

- 1000 (LTE), 500, 250, 125 μs

Mini-slots for URLLC/LLC transmissions:

- 7, 4 or 2 symbols.
- → Low fraction of latency budget for 1 ms

- (slot structures on next slide)



OFDM numerologies for NR and LTE

	Sub-carrier spacing	Symbol duration	Cyclic prefix	Slot (14 symbols)	Mini-slot (7 symbols)	Mini-slot (4 symbols)	Mini-slot (2 symbols)
NR and LTE	15 kHz	66.67 µs	4.76 µs	1000 µs	500 µs	286 µs	143 µs
NR	30 kHz	33.33 µs	2.38 µs	500 µs	250 µs	143 µs	71 µs
NR	60 kHz	16.67 µs	1.19 µs	250 µs	125 µs	71 µs	36 µs
NR	120 kHz	8.33 µs	0.59 µs	125 µs	63 µs	36 µs	18 µs

Note: 60 kHz sub-carrier spacing is optional in Rel-15

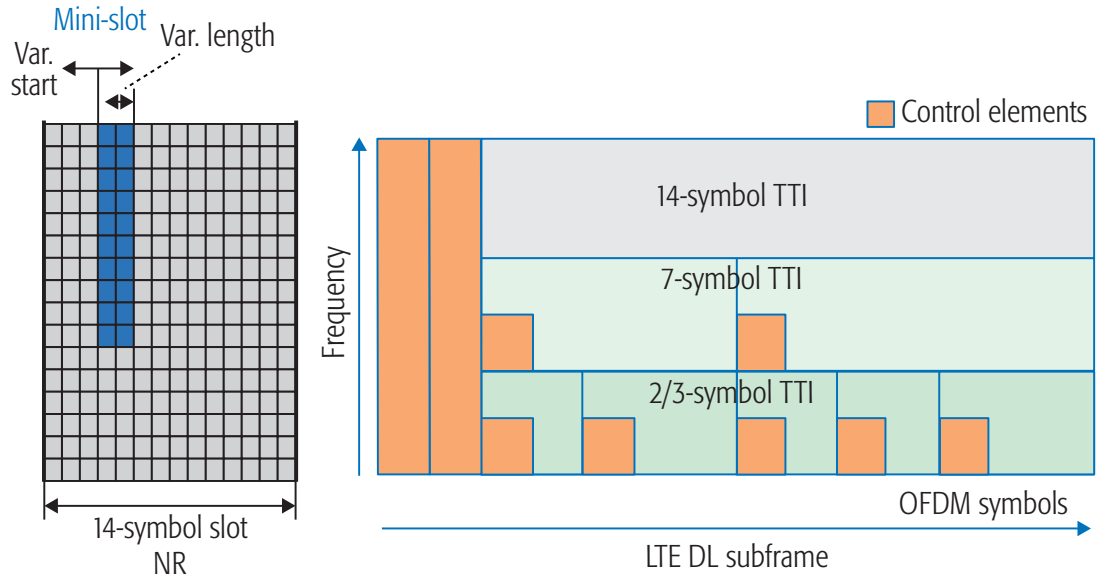
- H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee and B. Shim, "Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects," in IEEE Wireless Communications, vol. 25, no. 3, pp. 124-130, JUNE 2018.
- Sachs, J., Wikstrom, G., Dudda, T., Baldemair, R. and Kittichokechai, K., 2018. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. IEEE Network, 32(2), pp.24-31.

Short TTI in 5G

Different slot lengths are supported:

- Normal slot (14 symbols)
- Mini-slots (2-3 or 7 symbols)

Additional control signaling is required for mini-slots → increased overhead.



Note:

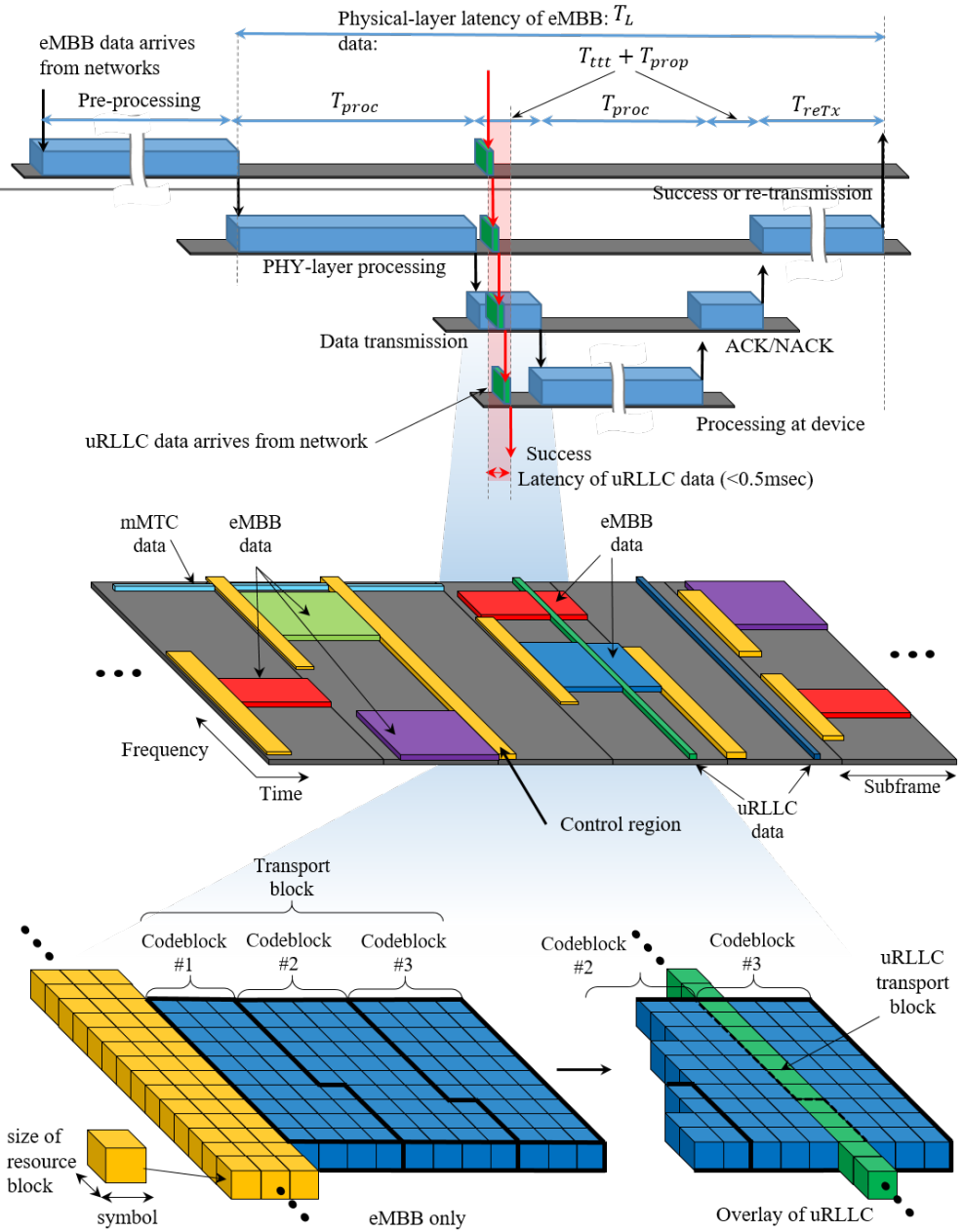
- *Strictly speaking, NR slot corresponds to LTE subframe*

Sachs, J., Wikstrom, G., Dudda, T., Baldemair, R. and Kittichokechai, K., 2018. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. IEEE Network, 32(2), pp.24-31.

5G URLLC downlink

Puncturing/pre-emptive scheduling

- When URLLC packet arrives, it is immediately sent in mini-slot(s), regardless of ongoing eMBB transmission
- URLLC RBs span over time rather than frequency, due to low-latency.
- URLLC punctures/pre-empts eMBB.
- URLLC transport block size is different than eMBB transport block size



Ji, H., Park, S., Yeo, J., Kim, Y., Lee, J., & Shim, B. (2018). Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects. IEEE Wireless Communications, 25, 124-130.

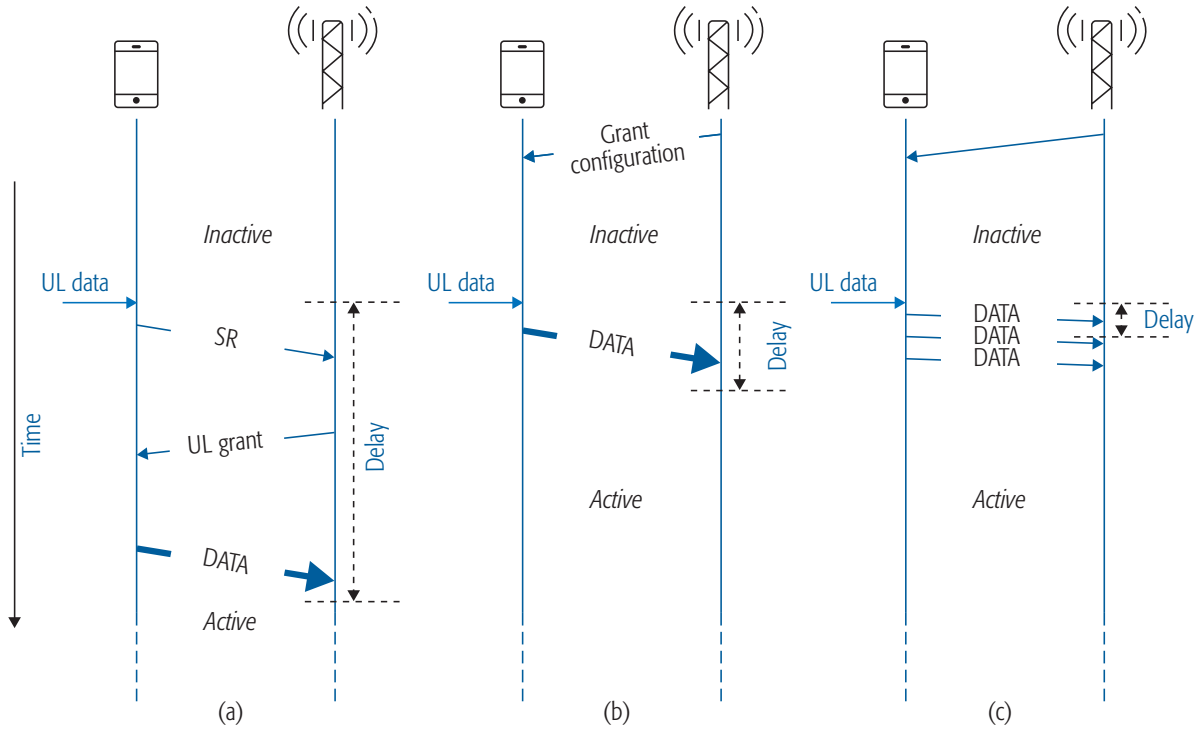
5G NR uplink access

Service Request (SR)

- UE sends service request
- BS replies with UL grant
- Too time consuming for URLLC

Semi-persistent scheduling (SPS)

- Both NR and LTE specify SPS modes, where periodically occuring minislots slots are assigned to UEs.
  - Often URLLC applications are periodic
- To save resources, overlap assignments.
- Short TTI allows reTX within latency budget.



Sachs, J., Wikstrom, G., Dudda, T., Baldemair, R. and Kittichokechai, K., 2018. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. IEEE Network, 32(2), pp.24-31.

Transmission latency

Worst-case latency for NR and LTE for

- Subcarrier spacing
- Slot length
- TDD/FDD mode
- Downlink/Uplink
- Semi-Persistent Scheduling / Service Request

Retx delay is added per reTX.

Most configurations support 1 ms latency

With reTX, latency budget becomes tight.

FDD		Downlink (ms)	Uplink (ms)		Retx delay (ms)
			SPS-based	SR-based	
NR	30 kHz, 14 s mini-slot	1.7	1.7	3.2	1 .5
	30 kHz, 7 s mini-slot	0.86	0.86	1.6	0.75
	30 kHz, 4 s mini-slot	0.54	0.54	0.96	0.43
	30 kHz, 2 s mini-slot	0.39	0.39	0.75	0.36
	120 kHz, 14 s slot	0.46	0.46	0.83	0.38
	120 kHz, 7 s mini-slot	0.33	0.33	0.64	0.31
LTE	15 kHz, 14 s TTI	4.0	4.0	10	6.0
	15 kHz, 7 s sTTI	2.0	2.0	6.0	4.0
	15 kHz, 2 s sTTI	1.0	0.86	2.3	1.4
TDD (DUDU pattern)		Downlink (ms)	Uplink (ms)		Retx delay (ms)
			SPS-based	SR-based	
NR	30 kHz, 14 s slot	2.2	2.2	4.1	2.0
	30 kHz, 7 s slot	1.1	1.1	2.1	1.0
	30 kHz, 4 s mini-slot	0.68	0.68	1.3	0.57
	120 kHz, 14 s slot	0.58	0.58	1.1	0.5
	120 kHz, 7 s mini-slot	0.39	0.39	0.64	0.25

TABLE 1. Worst case RAN transmission latencies for different 5G URLLC configurations (note that average latencies can be lower).

Sachs, J., Wikstrom, G., Dudda, T., Baldemair, R. and Kittichokechai, K., 2018. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. IEEE Network, 32(2), pp.24-31.

Massive MIMO

Large number of antenna elements, e.g. 128.

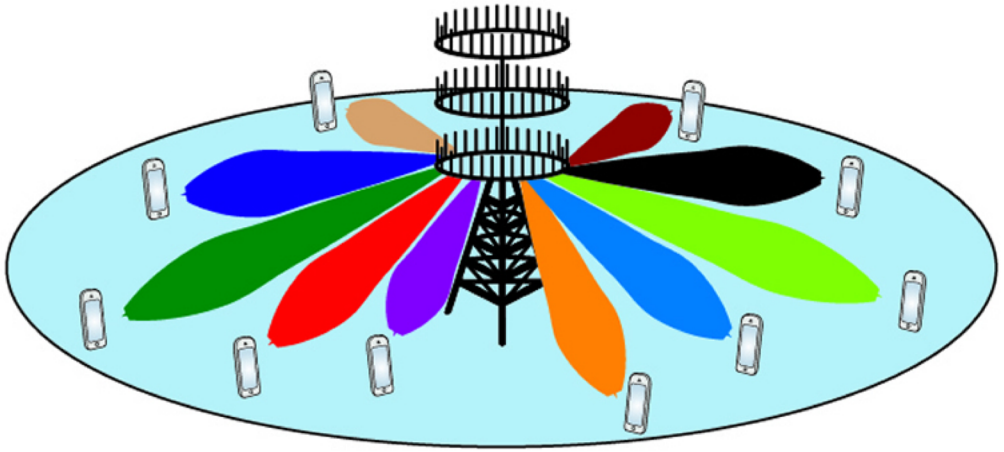
- Multiplexing of many users

Channel hardening **eliminates fast fading effects**

- Large scale propagation and shadowing

CSI can be obtained using TDD, pilot estimation, and exploiting channel reciprocity

Results in **ultra-reliable** link

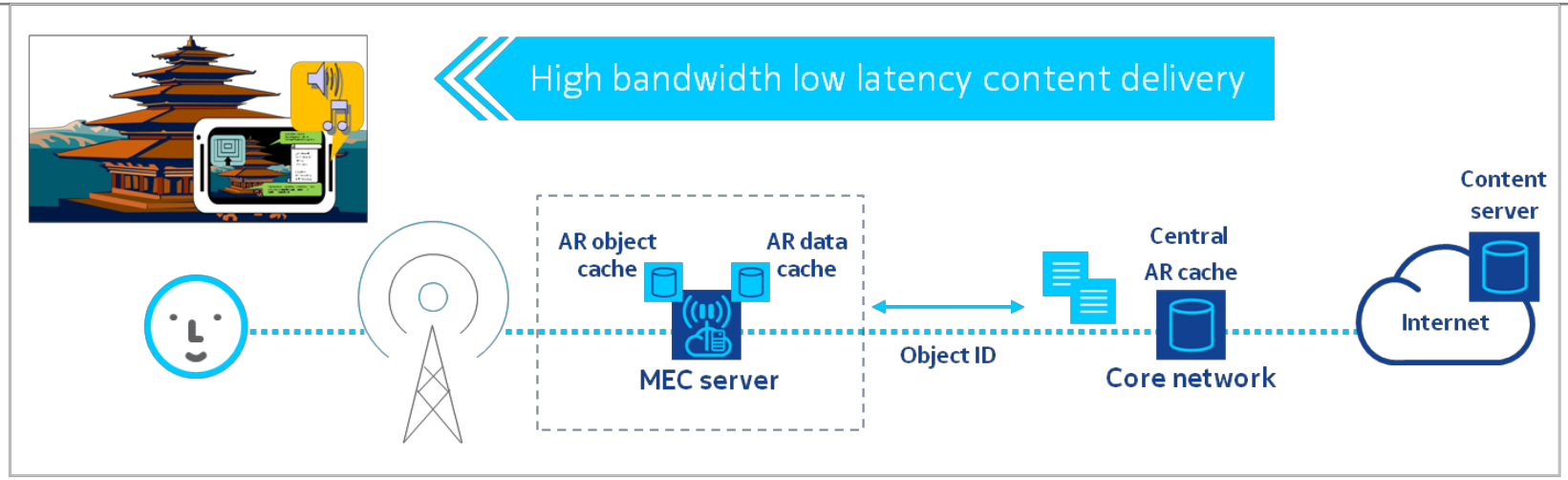


https://5g.co.uk/guides/what-is-massive-mimo-technology/

Simplified illustruation of TDD. 5G has flexible frame structure.

https://5g.ieee.org/tech-focus/march-2017/massive-mimo-for-5g

Mobile Edge Computing



Feature of Cloud RAN, Enabled through SDN and NFV technologies

MEC is geographically close to user:

- Computing resources
- Caching
- → Low latency and less traffic through core

MEC framework and architecture defined by ETSI MEC ISG standardization group

- Hu, Y.C., Patel, M., Sabella, D., Sprecher, N. and Young, V., 2015. Mobile edge computing—A key technology towards 5G. *ETSI white paper*, *11*(11), pp.1-16.
- D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach and F. Giust, "Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things," in IEEE Consumer Electronics Magazine, vol. 5, no. 4, pp. 84-91, Oct. 2016.

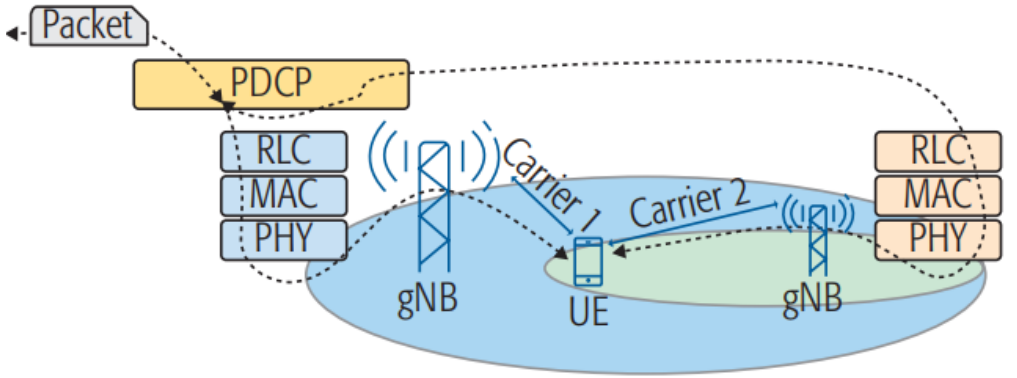
Multi-Connectivity

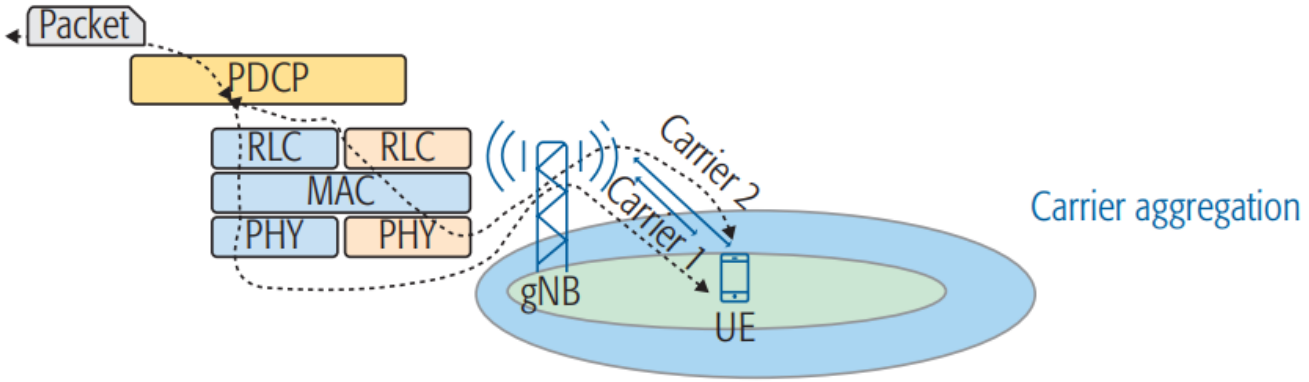
Currently, two options are suported:

- Dual connectivity:
  - Since release-12
  - Two BSs used for transmission to UE
  - Aggregation point is Packet Data Convergence Protocol (PDCP)
- Carrier Aggregation (CA):
  - Since release-10
  - One base station transmits on multiple carriers
  - Aggregation point in MAC, centralized scheduling according to channels
  - But requires tight integration of radio protocol stack

Packet Duplication for reliability boosting

- Introduced in release-15
- Both DC and CA duplicate packet in PDCP
- → Independent paths





Sachs, J., Wikstrom, G., Dudda, T., Baldemair, R. and Kittichokechai, K., 2018. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. *IEEE Network*, *32*(2), pp.24-31.

Network slicing in 5G

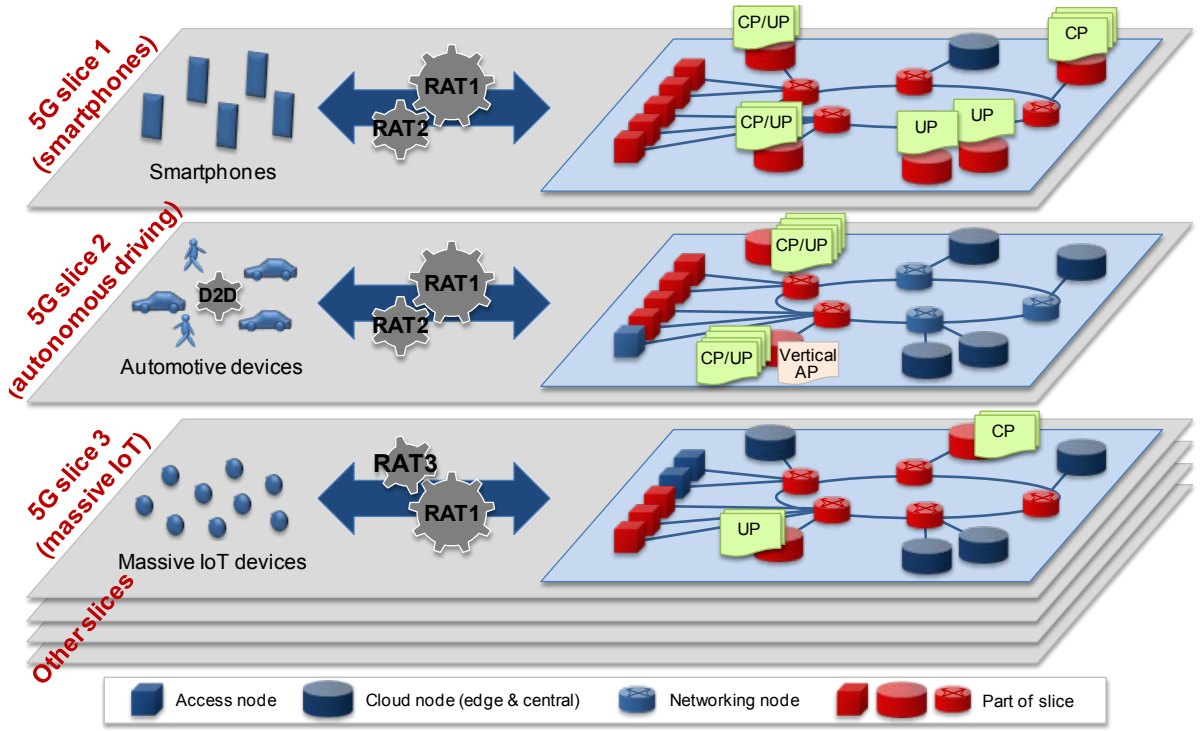
A network slice can cover many elements of the network:

- software modules running on cloud nodes
- specific configurations of the transport network supporting flexible location of functions,
- dedicated radio configuration or even a specific RAT,
- configuration of the 5G device

Wireless slicing refers to the allocation of wireless resources to different service types.

- Dedicated/orthogonal, or
- Shared/non-orthogonal (several flavours)

Also in 3GPP specs, e.g. TS 38.300



Alliance, N.G.M.N., 2015. 5G white paper. *Next generation mobile networks, white paper*, pp.1-125.

Examples of recent advances

Pre-emptive scheduling for (UR)LLC:

- Null Space Based Preemptive Scheduling For Joint URLLC and eMBB Traffic in 5G Networks
- Wireless Network Slicing for eMBB, URLLC, and mMTC

Multi-connectivity for URC:

- Novel Duplication Status Report for Multi-Connectivity Applications
- Optimized Interface Diversity for Ultra-Reliable Low Latency Communication

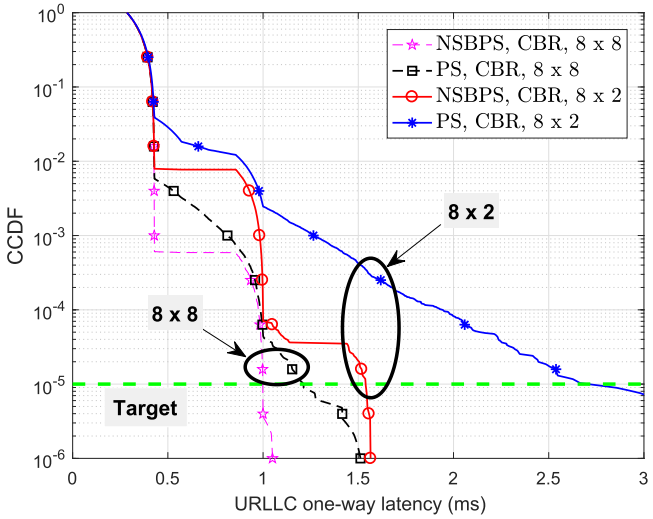
AALBORG UNIVERSITY
DENMARK

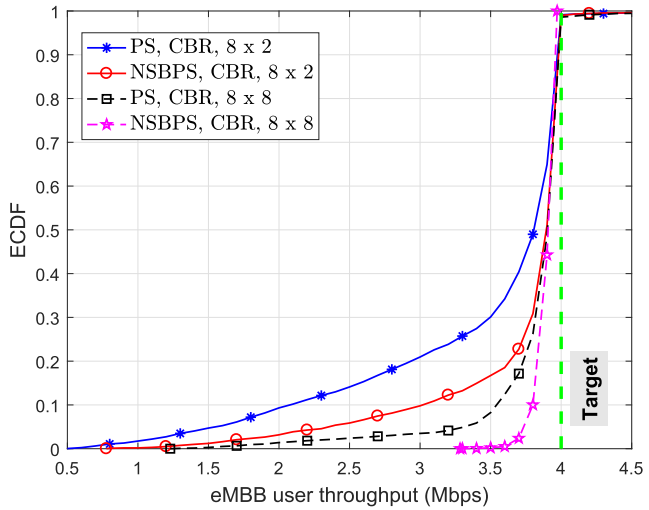
Null Space Based Preemptive Scheduling For Joint URLLC and eMBB Traffic in 5G Networks

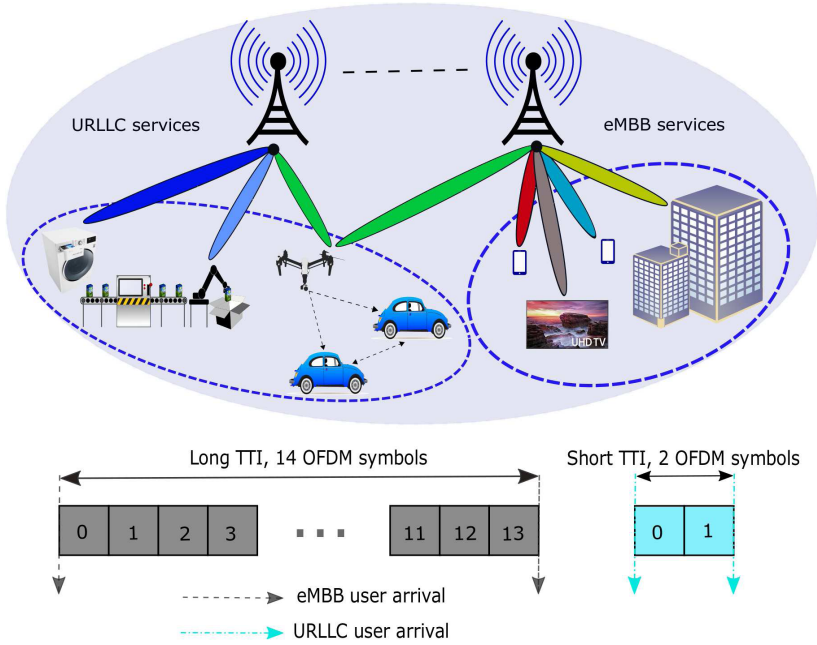
Exploits spatial degrees of freedom (MIMO) to simultaneously schedule URLLC and eMBB.

NSBPS scheduler:

- Arriving URLLC transmissions are paired with spatially orthogonal eMBB transmission
- Comparison to simple punctured scheduler (PS)







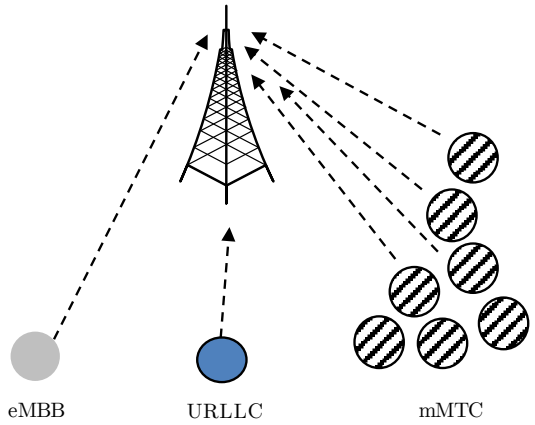
Esswie, A.A. and Pedersen, K.I., 2018. Null Space Based Preemptive Scheduling For Joint URLLC and eMBB Traffic in 5G Networks. *arXiv preprint arXiv:1806.04727*.

Wireless Network Slicing for eMBB, URLLC, and mMTC

Problem statement

Considers orthogonal (a, reserved) and non-orthogonal (b, puncturing) slicing:

- eMBB + mMTC
- **eMBB + URLLC**

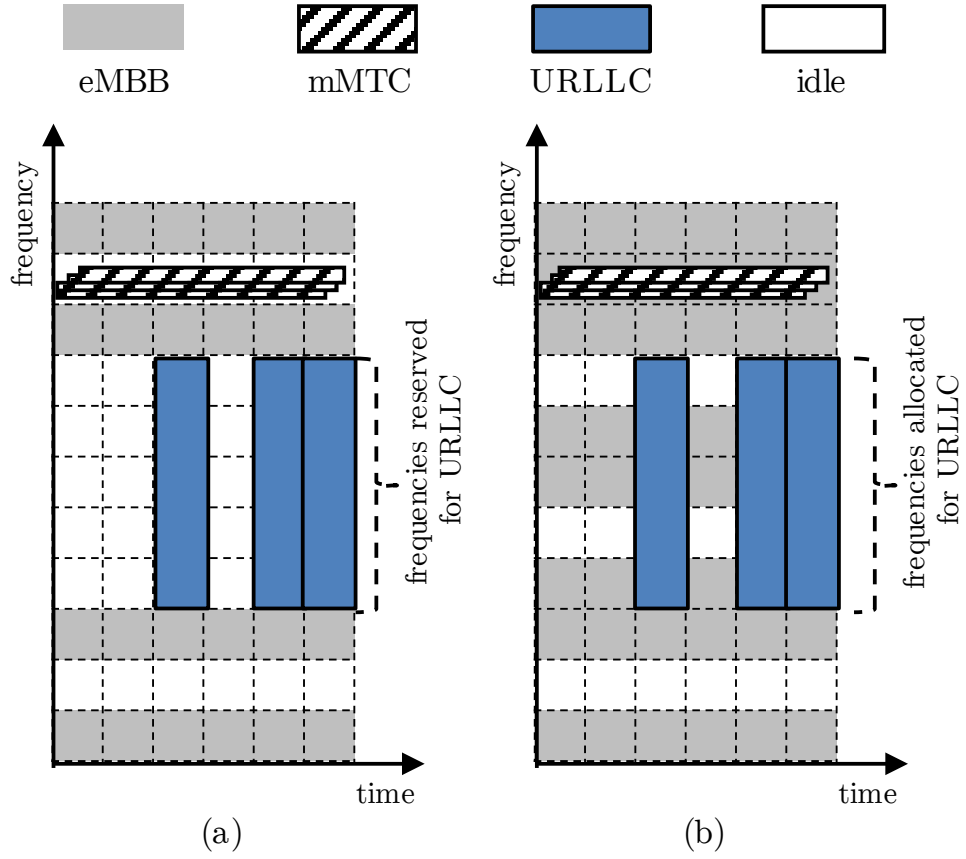


Research question:

Given reliability constraints:

- $\varepsilon_{\text{URLLC}} = 10^{-5}$
- $\varepsilon_{\text{eMBB}} = 10^{-3}$
- and scenario parameters such as SNR, what are achievable rates?

This work presents an information theoretic model to determine optimal slicing strategy, in different situations.



Popovski, P., Trillingsgaard, K.F., Simeone, O. and Durisi, G., 2018. 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View. *arXiv preprint arXiv:1804.05057*.

Wireless Network Slicing for eMBB, URLLC, and mMTC

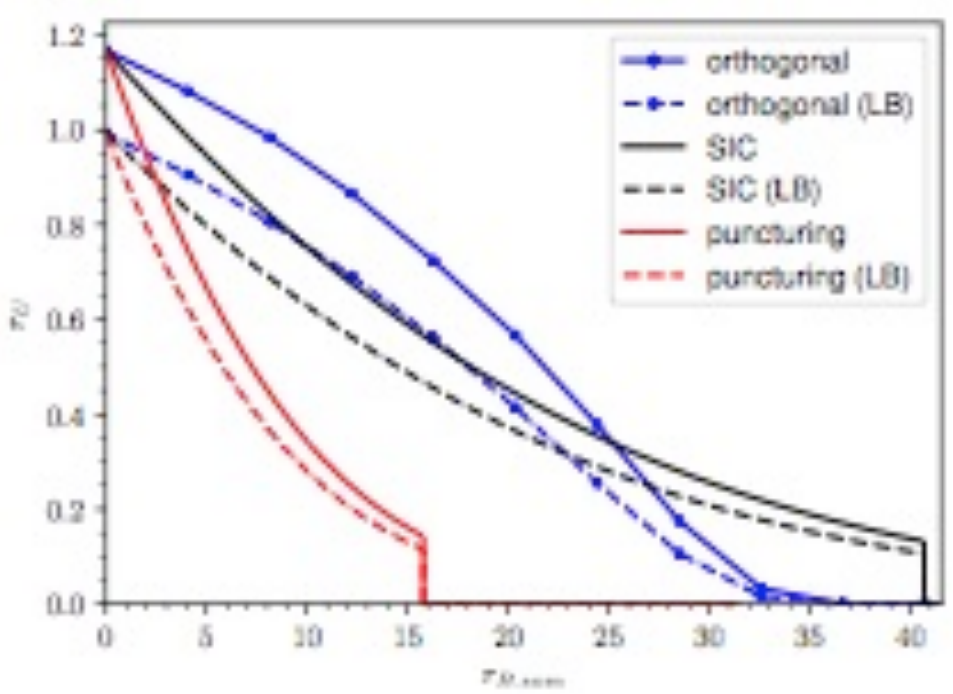
Rate regions for eMBB + URLLC

Three schemes are compared:

- Orthogonal
- Successive Interference Cancellation (SIC) at BS
- Puncturing (erasure)

Key results for URLLC:

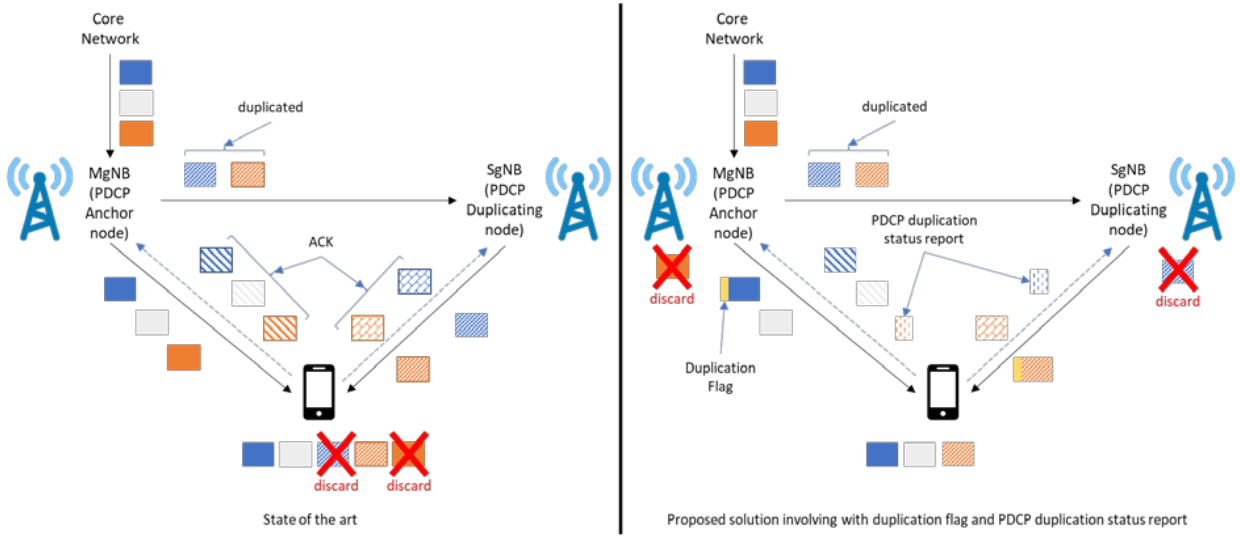
- Puncturing (erasure) is outperformed by others
- If high URLLC rate is desired, orthogonal slicing is best
- If high eMBB rate is desired, SIC is best
  - SIC may be infeasible due to complexity

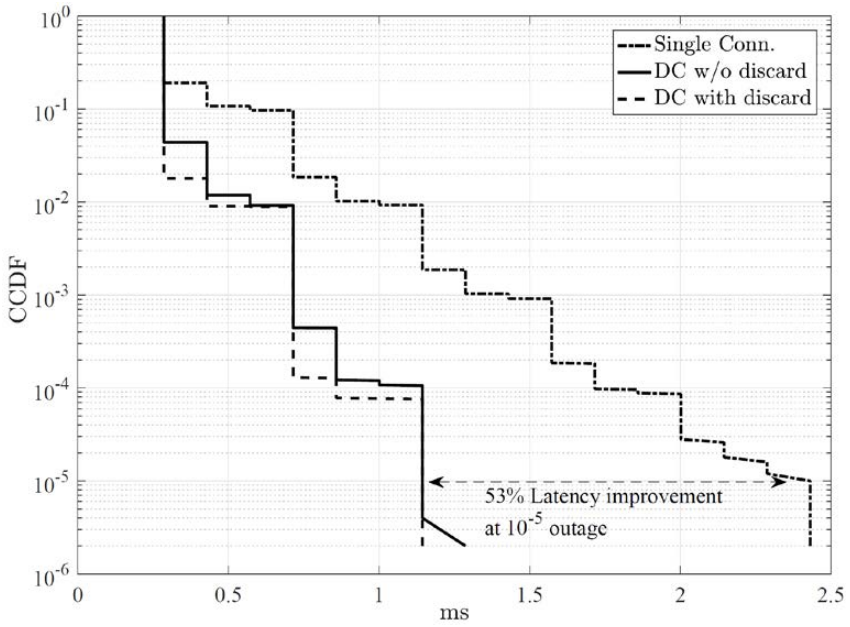


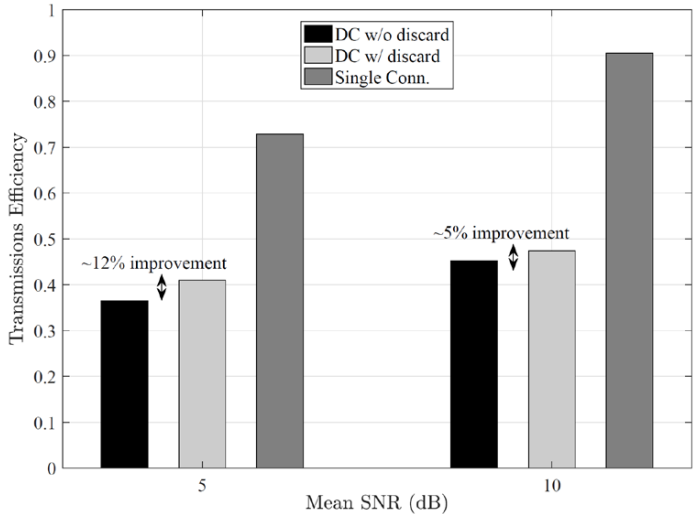
Novel Duplication Status Report for Multi-Connectivity Applications

Novel Duplication Status Report:

- Upon successful decoding of a PDCP packet, the UE sends status report to all nodes.
- Unsent copies of the same PDCP packet in other nodes are thus not transmitted.







Mahmood, N.H., Laselva, D., Palacios, D., Emara, M., Filippou, M.C., Kim, D.M. and de-la-Bandera, I., 2018. Multi-channel access solutions for 5G new radio. *IEEE Wireless Comm*.

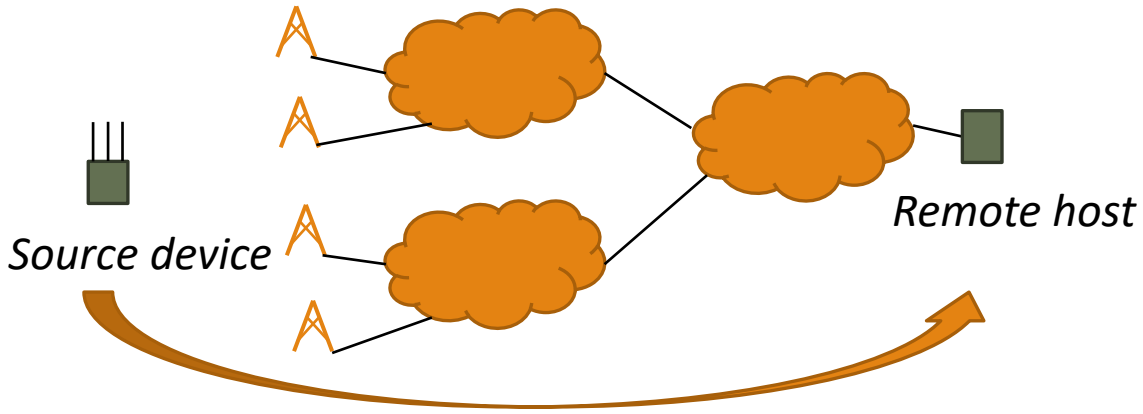
Optimized Interface Diversity for Ultra-Reliable Low Latency Communication (URLLC)

In this work, we have focused on the **integration of multiple communication technologies** to not rely on a single radio technology.

Ensure end-to-end UR(LL)C for packet transmissions.

Exploit multiple available communication interfaces on last hop link to M2M device.

- Or in other words, **Interface Diversity**.
- In principle we can use any communication technology.



Nielsen, J.J., Liu, R. and Popovski, P., 2017. Ultra-reliable low latency communication (URLLC) using interface diversity. *IEEE Transactions on Communications*.

Transmission strategies

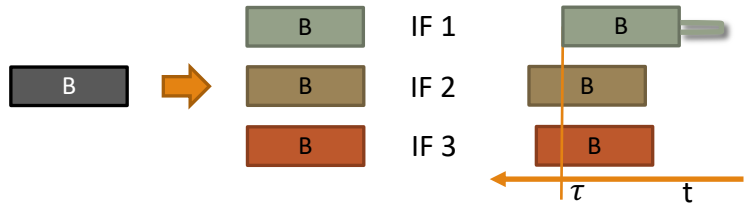
Cloning / packet duplication
- Maximum reliability
- Latency slightly reduced since first packet received defines latency.

Payload splitting through coding of individual packets
- Packet is decodable when slightly more than *B* bytes of coded payload is received, i.e. $\sum \gamma_i > 1$.
- Smaller fragments are (sometimes) faster to send.

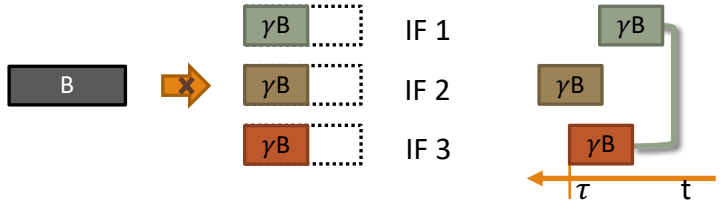
K-out-of-N (needed to decode)
- Equal sized coded fragments sent on each interface.
- Allows to trade-off transmission latency and reliability.

Weighted splitting
- Sizes of coded fragments can be optimized for a specific latency-reliability trade-off.

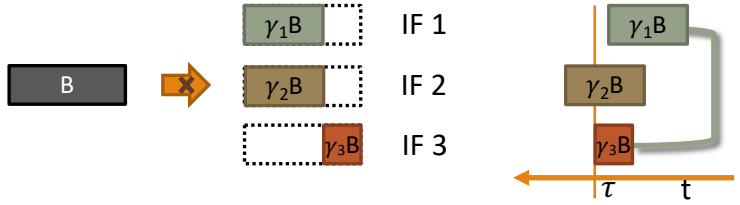
Cloning



2-out-of-3



Weighted splitting



Evaluation framework

Combining of latency CDFs through reliability engineering methods:

Cloning / packet duplication:

- System of components in parallel

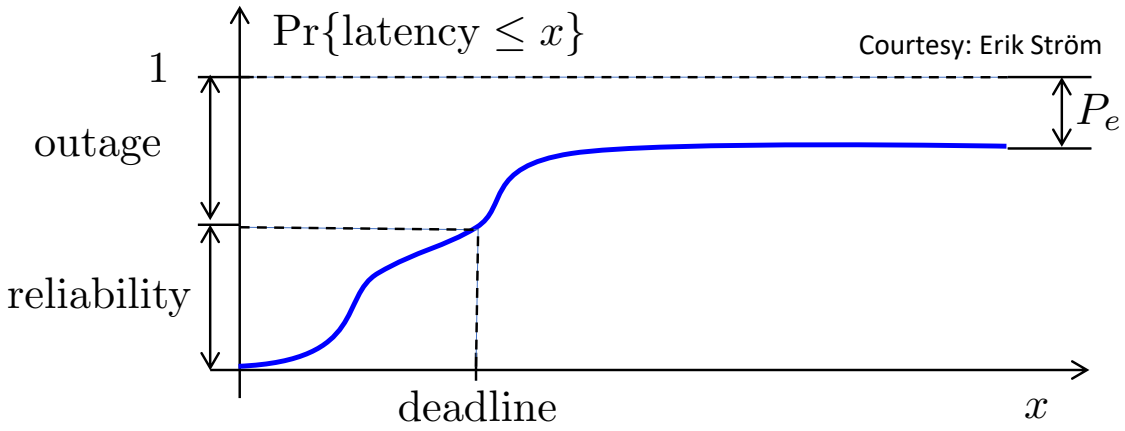
$$F_{N\text{-clon}}(x, \boldsymbol{\gamma}, B) = 1 - \prod_{i=1}^{N}(1 - F_i(x, \gamma_i B))$$

K-out-of-N:

- In case of identical interfaces

$$F_{k\text{-of-}N}(x, \boldsymbol{\gamma} B) = \sum_{r=k}^{N} \binom{N}{r} F(x, \gamma B)^r (1 - F(x, \gamma B))^{n-r}$$

- otherwise, use the following.



Weighted splitting:

- Considers the feasibility of all possible outcomes (all combinations of packet losses on interfaces):

$$F_{\text{weighted}}(x, \boldsymbol{\gamma}, B) = \sum_{h=1}^{2^N} d_h \prod_{i=1}^{N} G_i(x, \gamma_i B)$$

$$d_h = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} c_{h,i} \cdot \gamma_i \ge 1 \\ 0, & \text{otherwise} \end{cases}$$

$$G_i(x, \gamma_i B) = \begin{cases} F_i(x, \gamma_i B), & \text{if } c_{h,i} = 1 \\ 1 - F_i(x, \gamma_i B), & \text{if } c_{h,i} = 0 \end{cases}$$

$$\mathbf{C} = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

Optimization of weights

The vector of weights $\boldsymbol{\gamma}$ can be optimized, e.g. to reach a certain latency-reliability target.

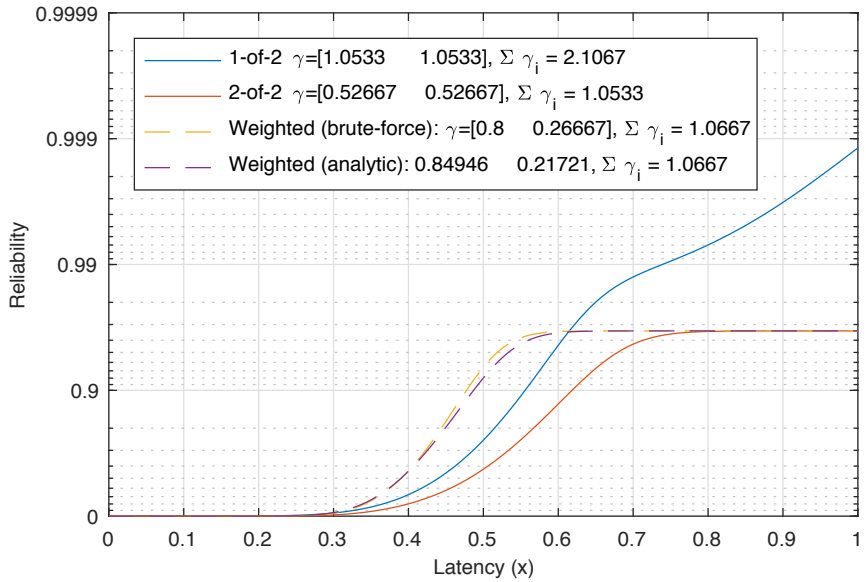
- Combinatorial problem, solution space grows as: $\left(1/\delta_\gamma\right)^N$

$$\begin{aligned} \max_{\boldsymbol{\gamma}} \quad & \sum_{r=1}^{R} F_{\text{weighted}}(l_r, \boldsymbol{\gamma}) \cdot w_r \\ \text{s.t.} \quad & \gamma_i \leq \gamma_{\text{d}} \\ & \sum_{i=1}^{N} \gamma_i \geq \gamma_{\text{d}}. \end{aligned}$$

In general, we use a brute-force algorithm to solve.

However in the paper we present an analytic solution for the simple case of two interfaces:

- Assuming latency distribution is Gaussian, with same variance.
- Based on approximation[1] of $\mathbb{E}[\max(X_A, X_B)]$
- Solution fits well with brute-force result.



[1]C.E. Clark, "The greatest of a finite set of random variables," *Operations Research*, vol. 9, no. 2, pp. 145–162, 1961.

Evaluation scenario

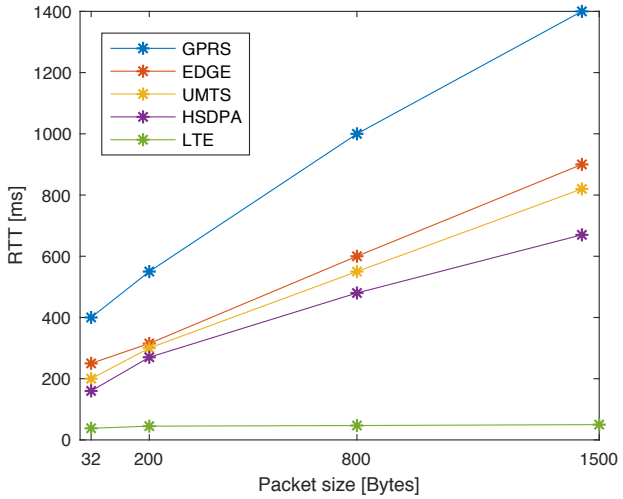
Assumptions:

- Latency distribution is Gaussian with:

$$\mu = \frac{\alpha \cdot \gamma B + \beta}{2}[ms] \qquad \sigma = \frac{\mu}{10}\ [ms]$$

- Based on linear regression model of ping measurements in mobile* network:

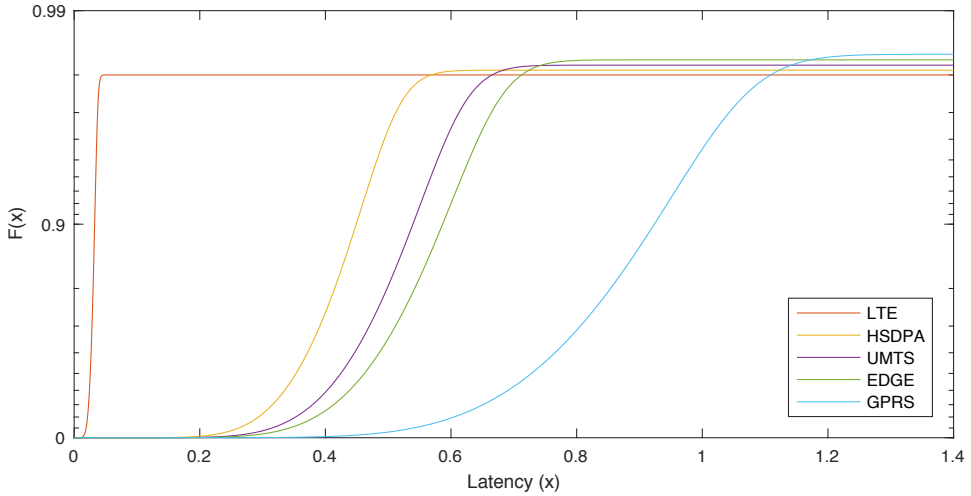
	GPRS	EDGE	UMTS	HSDPA	LTE
α	0.70	0.46	0.43	0.35	0.0067
β	400	230	200	178	41
P_e	0.984	0.983	0.982	0.981	0.980



Scenarios:

	IF1	IF2	IF3	IF4	IF5	B	$\boldsymbol{l}$	$\boldsymbol{w}$
$\mathcal{A}$	UMTS	GPRS	-	-	-	1500 bytes	$[0 \ldots 1]$ s	$[0 \ldots 1]$
$\mathcal{B}$	LTE	HSDPA	UMTS	EDGE	GPRS	1500 bytes	$[0.1, 0.4, 0.9^*]$ s	$[1, 10, 100^*]$
$\mathcal{C}$	HSDPA	HSDPA	GPRS	GPRS	GPRS	1500 bytes	$[0.5]$ s	$[1]$

Utility function: $\sum_{r=1}^{R} F_{\text{weighted}}(l_r, \boldsymbol{\gamma}) \cdot w_r$



Measurements were provided by Telekom Slovenije for the SUNSEED project.

Numerical results

Optimizing for several latency targets:

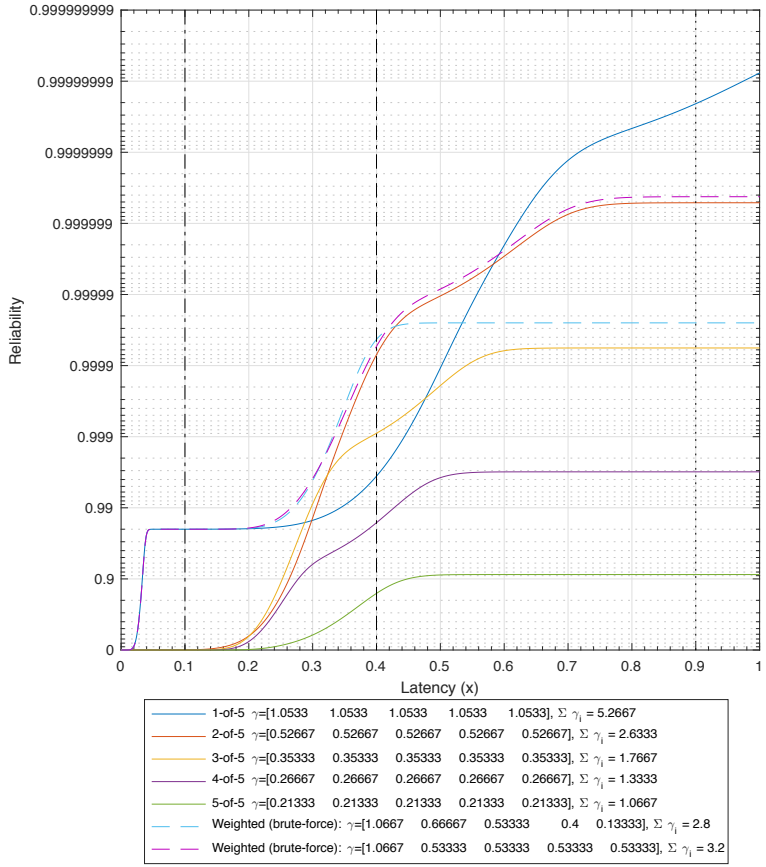


Fig. 4. Reliability results for scenario $\mathcal{B}$. Note: the target latency $l_2 = 0.9$ s only applies to the last strategy.

Optimizing for single latency target:

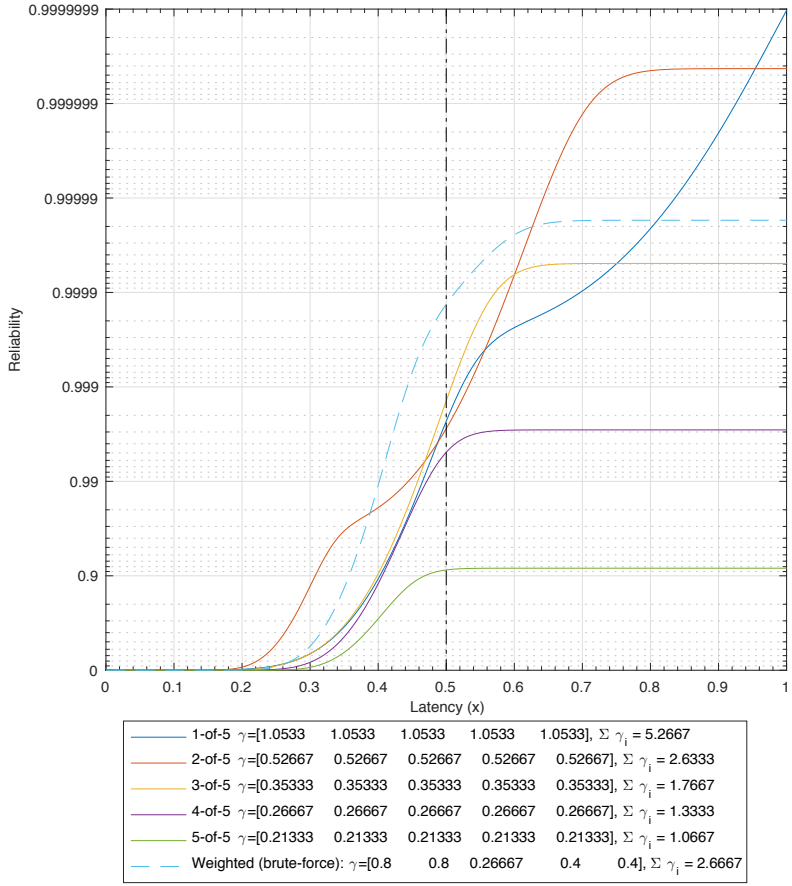


Fig. 5. Reliability results for scenario $\mathcal{C}$.

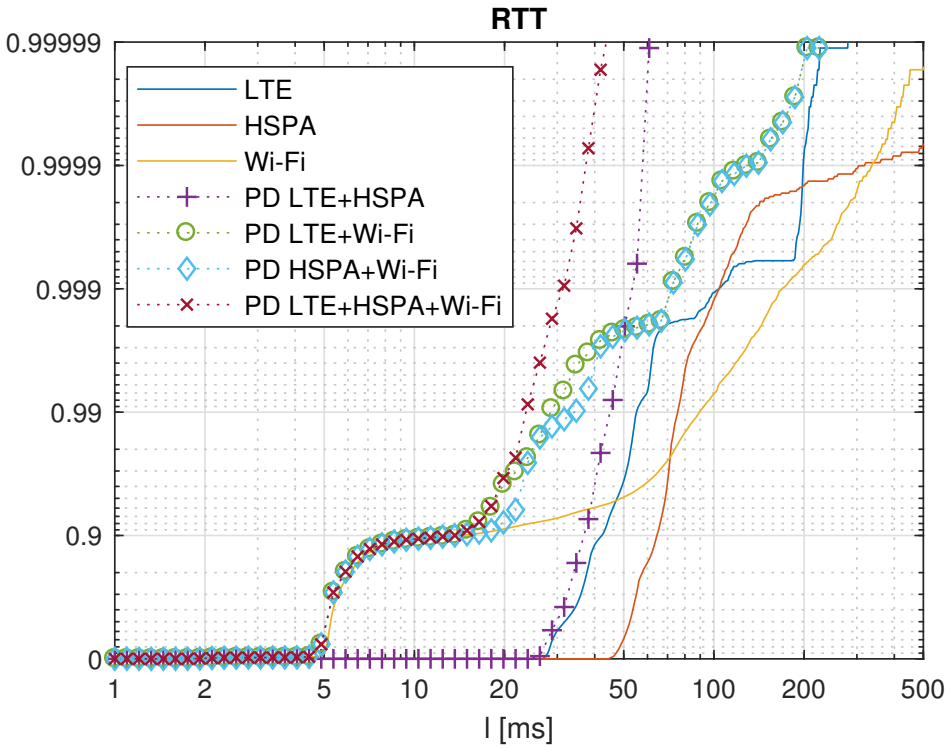
Packet duplication with legacy systems

One-way, end-to-end latency measurements:

- Obtained during full week-day at Aalborg University campus.
- One 128 bytes packet every 100 ms, A→B
- GPS time-synchronization, sub-ms accuracy.
- Three technologies considered:
  - LTE, HSPA, and Wi-Fi

Key observations:

- Single interface:
  - 0.99 within 50-100 ms
- Packet Duplication:
  - LTE+HSPA can reach 0.99999 within ~65 ms deadline
  - LTE+HSPA+Wi-Fi reaches 0.99999 already at ~45 ms.
    - Even though Wi-Fi is terrible alone, it can help to reduce latency in Multi-Connectivity setting.



The end