Connecting ideas, anticipating the future:
collaborative innovation for 5G and 6G networks.

II INTERNATIONAL WORKSHOP xGMobile

Organized by:

xGMobile · Inatel · FAPEMIG · EMBRAPII · GOVERNO DE MINAS — AQUI O TREM PROSPERA. · MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO · GOVERNO DO BRASIL — DO LADO DO POVO BRASILEIRO

uRLLC for 5G and Beyond

Prof. Yonghui Li

The University of Sydney

Organized by:


Lombok
Sumba
Kupang
TIMOR-LESTE
Timor
Timor Sea
Arafura Sea
Torres Strait
Port Moresby
Awara
Ashmore and Cartier Island (AUSTRALIA)
Darwin
Gulf of Carpentaria
Indian Ocean
Coral Sea Islands (AUSTRALIA)
Cairns
Great Barrier Reef
Townsville
KING LEOPOLD RANGE
MACDONNELL RANGE
Mount Isa
GREAT DIVIDING RANGE
Port Hedland
GREAT SANDY DESERT
HAMMERSLEY RANGE
Mackay
Alice Springs
Rockhampton
Gladstone
GIBSON DESERT
AUSTRALIA
SIMPSON DESERT
Toowoomba
Brisbane
Gold Coast
Geraldton
GREAT VICTORIA DESERT
Lake Eyre (lowest point in Australia, -15 m)
DARLING RANGE
Kalgoorlie
FLINDERS RANGE
Broken Hill
Perth
Rockingham
Bunbury
Whyalla
Newcastle
Sydney
Wollongong
Esperance
Adelaide
Canberra
Mount Kosciuszko (highest point in Australia, 2229 m)
Great Australian Bight
Melbourne
Geelong
Bass Strait
Launceston
Hobart
Tasmania
Indian


ABC传媒
© APEX

University of Sydney

- Australia's oldest university established in 1850
- Rank 19th in World University Rankings by US News 2024
- Rank 22nd in QS ranking 2024

THE UNIVERSITY OF SYDNEY

- 73,000 students from more than 140 countries
- 8100 academic and professional staff
- #4 in the world for graduate employability
- #2 in the world for research impact.
- 7 prime ministers
- 5 Nobel laureates
- 145 Olympians

Breakthrough Innovations

- **Ultrasound scanner： by George Kossoff**
- **Artificial Heart Pacemaker**
- **Ventilator**
- **Coclear**
- **Aircraft black box**
- **WI-FI**
- **….**

Centre of Excellence in Telecommunications (CoET)

- Mission: Provide innovative solutions to telecommunication industry
- A world renown and largest Research Centre in Australia in Telecommunications
- Telecommunication Engineering (Shanghai ranking): 13th globally
- R&D team in communications
-                 - 9 Academic staff (5 IEEE Fellows)
-                 - 6 Research Fellow
-                 - More than 50 PG students
- Areas: 5G, Wireless communications, Networks, Industrial Internet of Things, Signal processing, Wireless Networked Control, AI in Wireless Communications
- Activities: Research, product development, system design and industry consultancy
- Publications: >200 IEEE journal papers in the past five years

Track Record

- Over 40 industry projects, more than $20million
- Design of last mile and home area networks for $500 million Australian Smart Grid Smart City project
- Participate in 7-years' $100million SmartSat CRC project
- Contributions to the IEEE Standard (802.X)
- Soft output detection and decoding used in 2G, 3G and 4G base stations and terminals
- Adaptive CDMA receivers used in 3G cellular systems
- Soft frequency reuse has been adopted by 3GPP standard
- Develop Ultra-high speed FTN (faster than Nyquist Sampling) based microwave systems (achieving 11.2bit/s/Hz super-high spectrum efficiency), high speed train communication systems for Huawei
- Patents: 15 (wireless system architecture, solutions, protocols, core technology)

Collaborators

acma

DATA 61

CSIRO

ABB

Radisys

starnet

SCHAEFFLER

NORTEL

NEC Empowered by Innovation

vodafone

'yes' OPTUS

NOKIA

QUALCOMM

ZTE中兴

Canon

HUAWEI

ERICSSON TAKING YOU FORWARD

EnergyAustralia®

Current Research Projects of Centre

- Open source 5G open platform and SDN development: Developed the open source 5G open platform
- AI and Machine Learning in Wireless Communications: Developed an AI based signal propagation prediction algorithm, which is much accurate than the commercial propagation tools. It can be widely used for cell planning, data driver propagation modelling and communication infrastructure deployment
- Ultra-reliable and low latency communication networks : redesign communication networks to significantly reduce the latency and improve the reliability. Key applications are in factory automation, industry control, autonomous vehicles
- High performance Industrial communications networks : develop an ultra-reliable and low latency converged wireless and wired networks for industrial control and factory
- Millimetre wave communications and Massive MIMO Research and SDRprototype development : develop high speed communication networks based on mmWave and massive MIMO
- Cooperative robots with wireless networked control : To separate manipulator (controller) from Robot to enable low-precision robotic arms to perform high-precision tasks via a closed-loop wireless networked control system designed for its manipulator

uRLLC for 5G and Beyond

Prof. Yonghui Li

Centre for IoT and Telecommunications

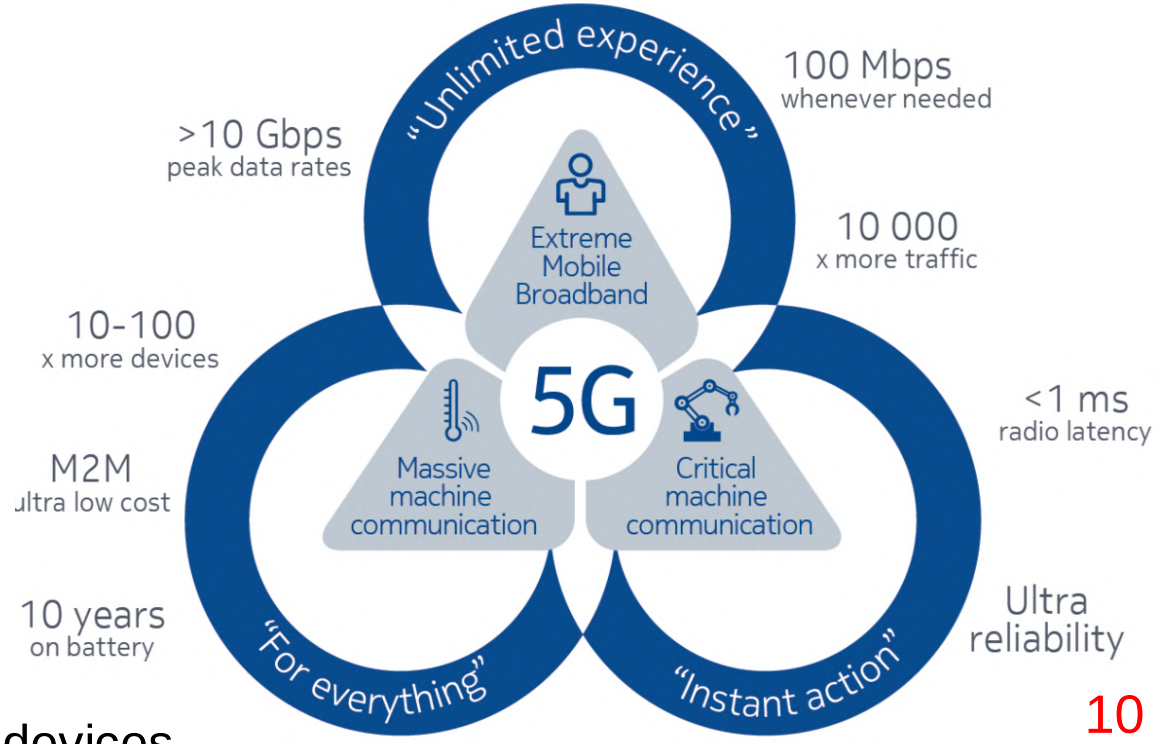
The University of Sydney

THE UNIVERSITY OF SYDNEY

5G Technical Requirements

1000 X data capacity per cell

100 X typical user data rate



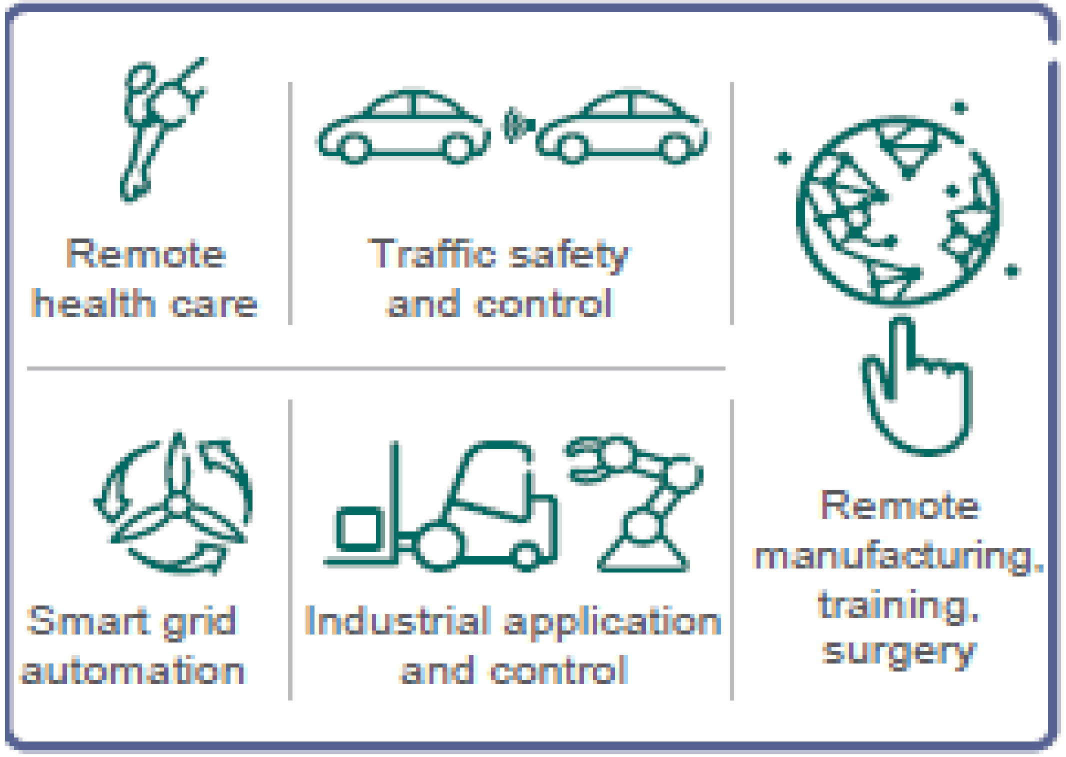
100 X number of connected devices

10 X longer battery life

10 X reduced latency

1000X improved reliability

URLLC for Mission Critical IoT



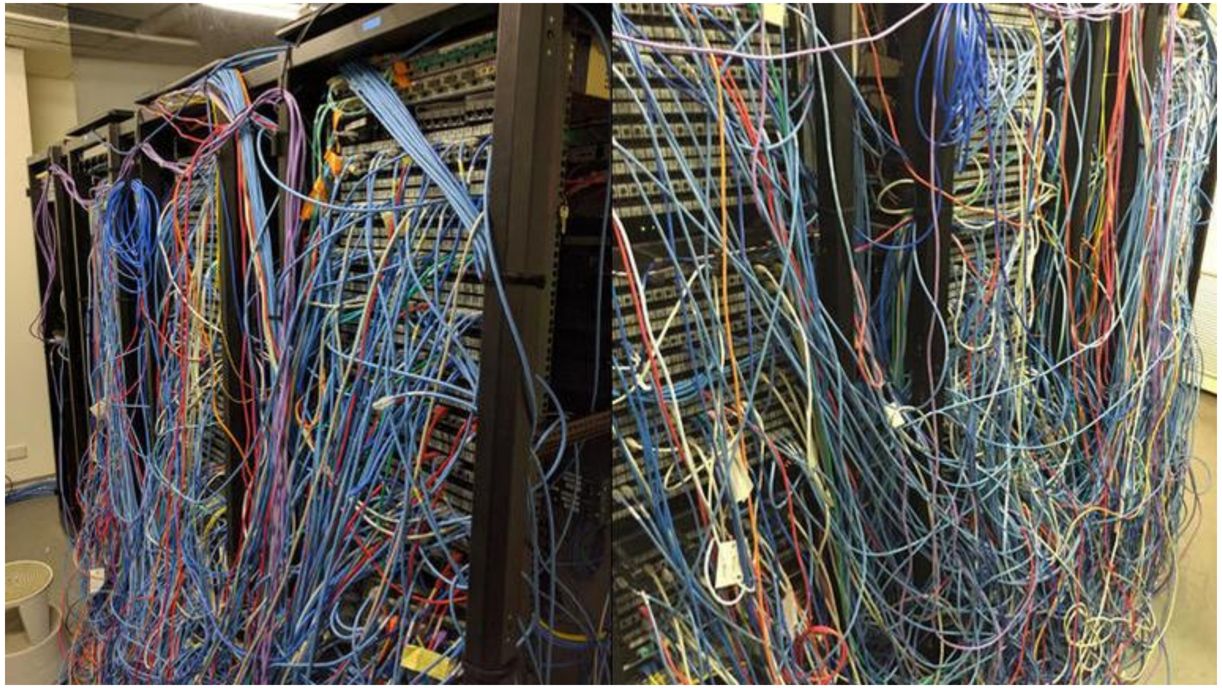
- 5G URLLC targets: Latency <1ms air link; <10ms E2E; Reliability: BLER < 10^{-5}
- 6G URLLC targets: Latency < 100µs; Reliability: BLER < 10^{-9}

Industrial Network Requirements

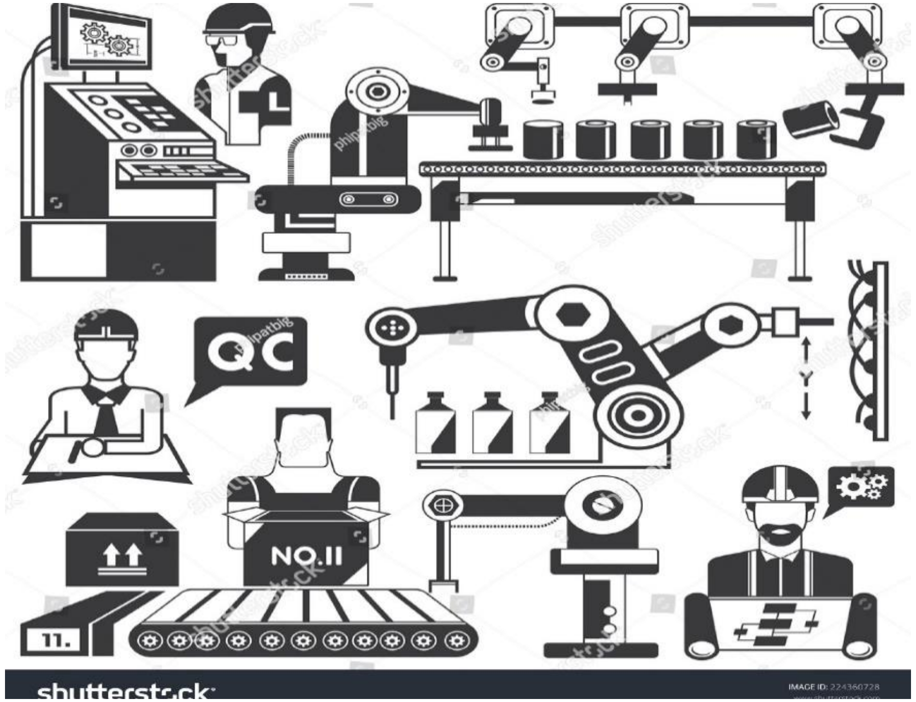
- Ultralow latency (<1ms on the link level, vs 100ms in 4G)
- Ultrahigh reliability (packet error rate $<10^{-9}$, vs 10^{-2} in 4G)
- Determinism
- Criticality
- Scalability

Wireless vs Wired Networks for IoT

- Currently wired networks are dominant
- Trends towards wireless due to
  - lower installation cost
  - lower maintenance
  - easier redundancy
  - higher flexibility
  - enable mobile applications
  - even higher long term reliability



IoT in Industrial Automaion



- Current industrial robots form a closed system of manipulators and controllers
  - ✓Connected via wired networks
  - ✓Fixed degree of freedom (DOF)
  - ✓Expensive set-up
  - ✓Challenging for Cobot and HMC

- Fog Robot Architecture:
  - ✓Enables the separation of manipulators and controllers
  - ✓Powerful edge controller wirelessly controls robot arms
  - ✓Flexible DOF
  - ✓Key enabler for cobot and HMC

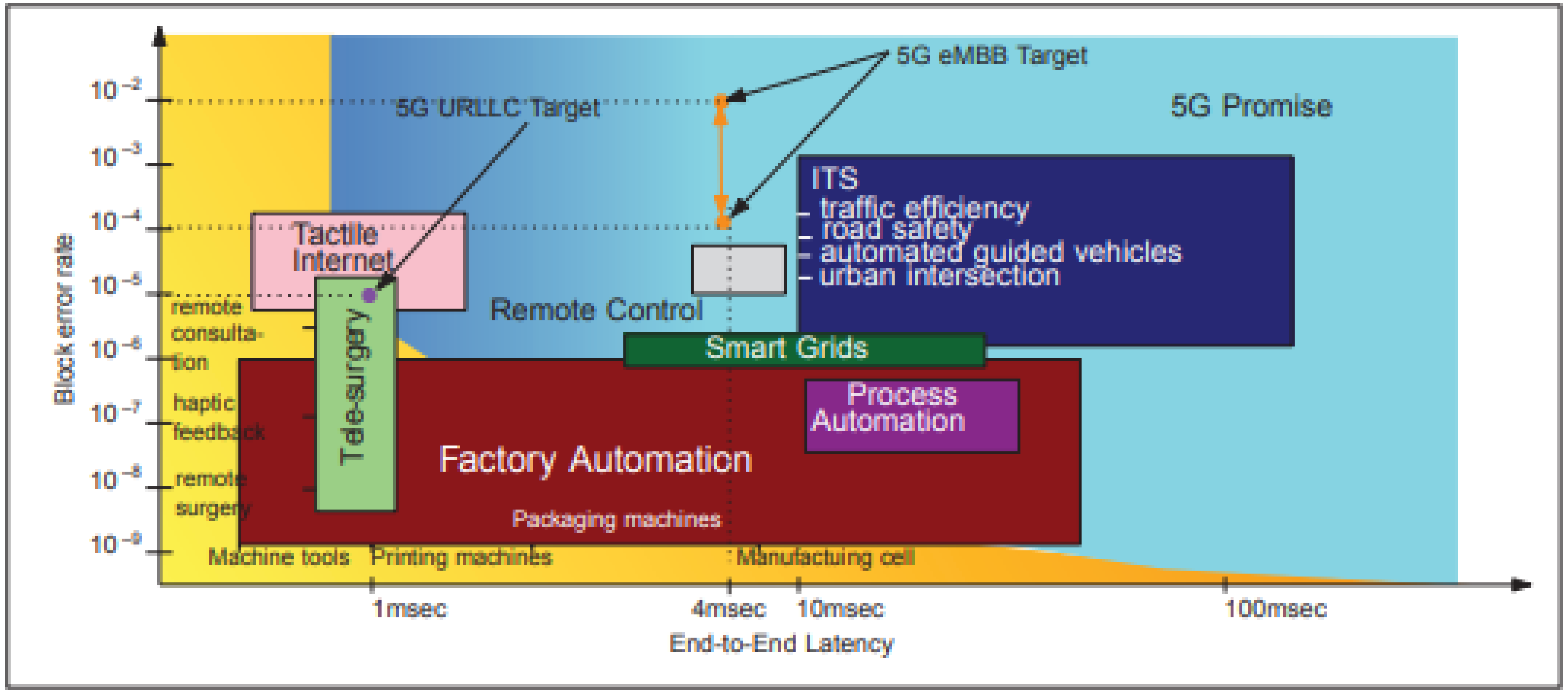
URLLC Applications, Requirements & Markets

Industry Vertical	Market segment size	Remarks
Healthcare	GBP 43b 2018 global GBP 2.9b 2018 UK 66% digital health 18% Tele-healthcare 35% mHealth (mobile apps and wearables) Medical robot systems global market of GBP 15b	10 ms Round trip delay to support haptic feedback Tendencies to evolve towards remote applications Mutual gain in merging tele-healthcare with connected home 50% Potential Market share to network operators
Automotive	GBP 104.2 billion by 2019 connected car global: (a) driver assistance: GBP 42b (b) safety technologies: GBP 28b (c) safety and autonomous driving: 66%	10 ms One way delay in Cooperative driving 36m cars with SIM cards to be sold by 2018 Partnerships between Telecoms and Manufacturers Global revenue of GBP 3.4b for telecom companies
Entertainment	Media and Entertainment UK Market: GBP 85.32b Global AR and VR market is GBP 118.5b Driver for VR is gaming with 76% of content in gaming Gaming in the UK: GBP 3.9b consumer spend in 2015 Mobile Gaming in UK: GBP 548m	7 ms Round trip delay to support VR/AR Largest industry in wearable segment growth VR and AR are driven by Gaming industry Revenues are dependent on level of added value provided
Manufacturing	ICT Sector in Manufacturing market (EU): GBP 755.7b Robotics and autonomous system technologies: GBP 13b by 2025 in the UK Ethernet and Wireless growing in industrial automation	Sub-1 ms One way delay in control applications Ethernet solutions are growing fast Cyber-physical systems to play primary role Opportunities in low latency and data analytics

Lema, Maria A., et al. "Business case and technology analysis for 5G low latency applications." *IEEE Access* 5 (2017): 5917-5935.

Why do we need URLLC?

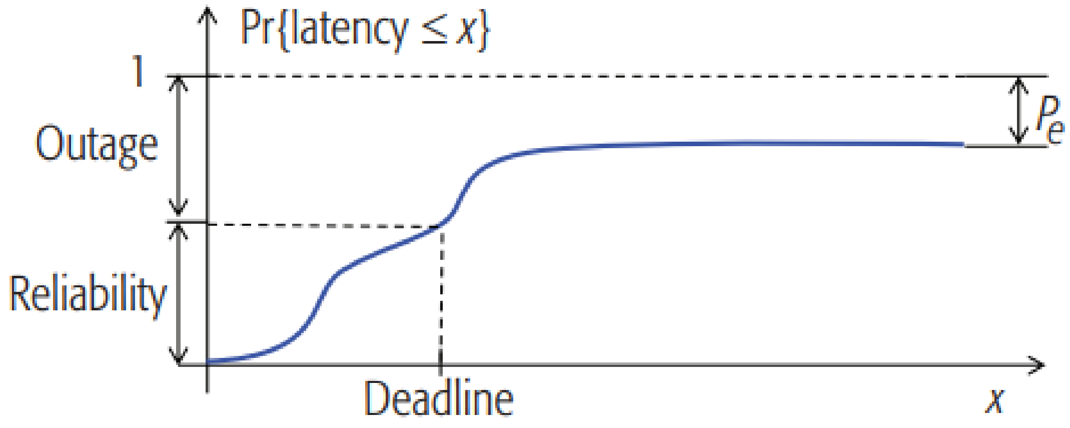
What levels of latency and reliability do we need?



[S-2] Chen, He, et al. "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches." *IEEE Communications Magazine* 56.12 (2018): 119-125. https://doi.org/10.1109/MCOM.2018.1701178

Reliability in URLLC

What kind of reliability are we after?

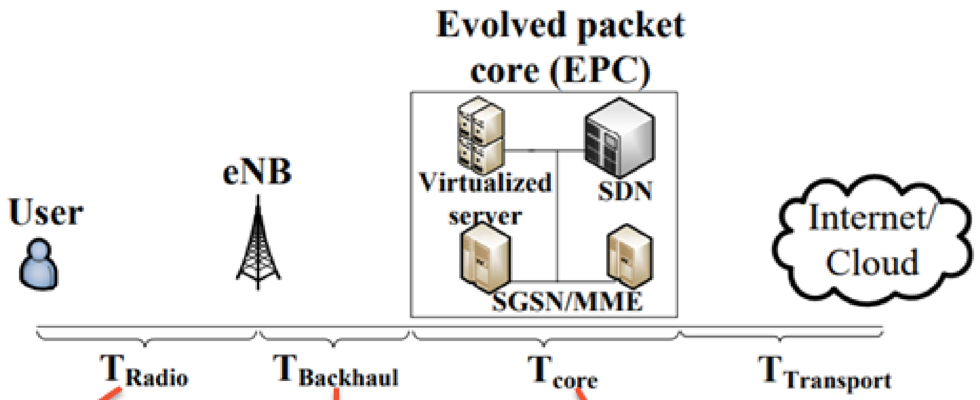


- Tightly coupled Reliability and latency metrics
- Reliability is defined with respect to a predetermined deadline
- The exact deadline and the reliability level are application-dependent.

Popovski, P., Nielsen, J.J., Stefanovic, C., De Carvalho, E., Strom, E., Trillingsgaard, K.F., Bana, A.S., Kim, D.M., Kotaba, R., Park, J. and Sorensen, R.B., 2018. Wireless access for ultra-reliable low-latency communication: Principles and building blocks. *Ieee Network*, *32*(2), pp.16-23.

Latency in URLLC

Latency Components



T_{Radio}	$T_{Backhaul}$	T_{core}	$T_{Transport}$
Packet transmission time	Time for building connections between BS and core network	Processing time at core network	Date communication time between core network and Internet

$$T_{\mathrm{E2E}} = 2\left(T_{\mathrm{Radio}} + T_{\mathrm{Backhual}} + T_{\mathrm{Core}} + T_{\mathrm{Transport}}\right)$$

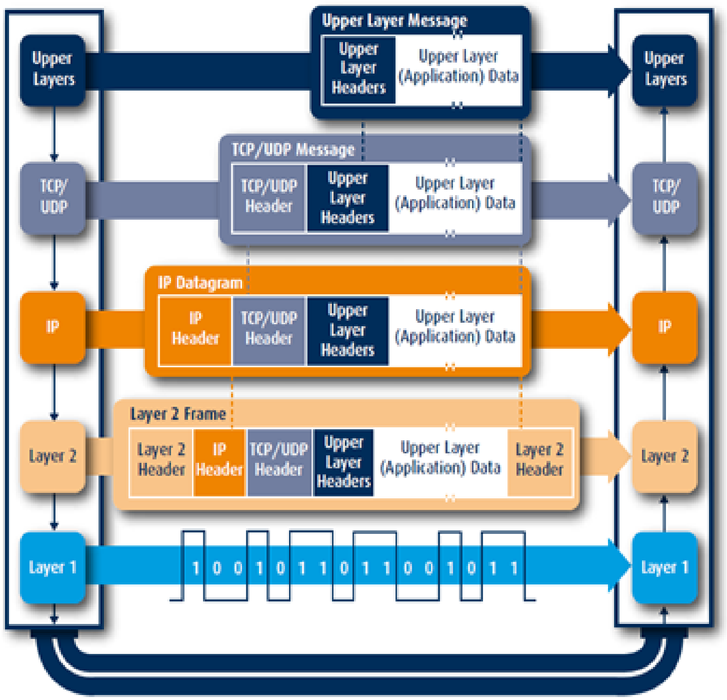
Latency in Wireless Network

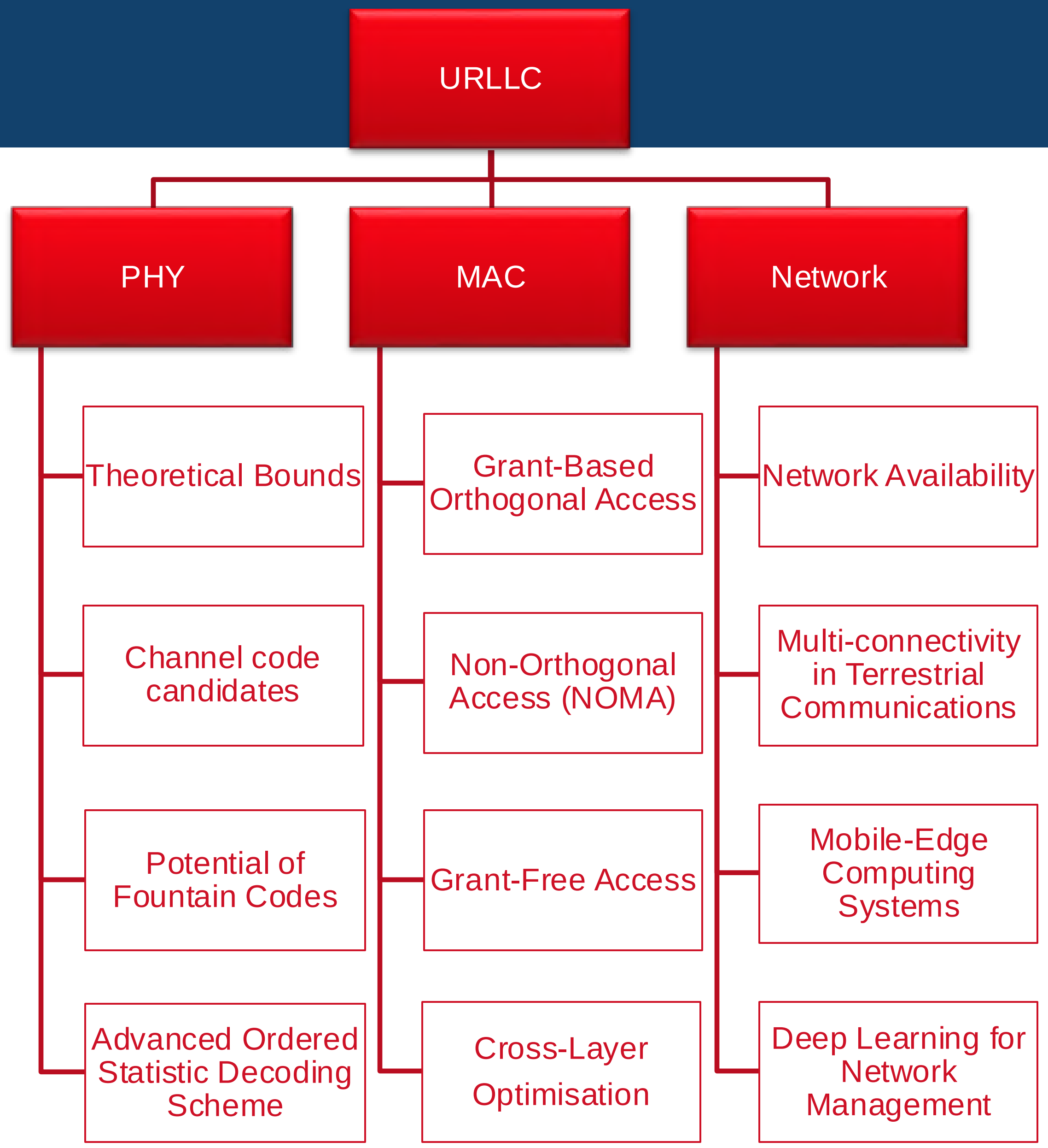
$$T_{radio} = T_{TTI} + T_{ARQ} + T_{link} + T_{MAC} + T_{BPT} + T_{MPT} + T_{prop}$$

- Current LTE end-to-end latency <100ms
- It is not guaranteed

- Sources of latency

- T_{TTI} : Transmit Time Interval – 1ms
- T_{ARQ}: HARQ retransmission > 8ms
- T_{link} : link establishment – 4ms
- T_{MAC}: multiple access – 10ms
- T_{BPT} + T_{MPT} : receiver processing – 3ms
- T_{prop}: 1ms per 300km
- Core network/Internet – Vary widely




URLLC
PHY
MAC
Network
Theoretical Bounds
Channel code candidates
Potential of Fountain Codes
Advanced Ordered Statistic Decoding Scheme
Grant-Based Orthogonal Access
Non-Orthogonal Access (NOMA)
Grant-Free Access
Cross-Layer Optimisation
Network Availability
Multi-connectivity in Terrestrial Communications
Mobile-Edge Computing Systems
Deep Learning for Network Management

How do we address URLLC at the Physical Layer (PHY)?

PHY

Theoretical Bounds

- Revisit the Shannon Bound
- Normal Approximation

Channel code candidates

- Turbo Codes
- Convolutional Codes
- BCH Codes
- LDPC Codes
- Polar Codes

Potential of Fountain Codes

- Motivation for Rateless Codes
- Rateless Code Design for Short Block lengths
- Rateless Code Performance

Advanced Ordered Statistic Decoding Scheme

- Approach near MLD Performance
- Reduced Complexity
- Reduced Decoding Latency

Constraints for achieving uRLLC

- One metric is optimized for improvement at the degradation of another metric
- High capacity needs large control overhead (e.g., cyclic prefix, transmission mode, and pilot symbols); this makes the portion of overhead unacceptably high in shorter TTI.
- In LTE, packet retransmission takes around 8ms, and removal of retransmission will affect packet error significantly.

Fundamental Limit of Finite Block Length

- Fundamental trade-offs between capacity, latency, and reliability
- Existing systems are designed based on Shannon formula

$$R = \frac{v}{m} = \log_2(1+\gamma)$$

- Achieving Shannon capacity requiring transmitting extremely long data packets

Fundamental Limit of Finite Block Length

› URLLC mandates short packet communications and extremely low BER

› Errors cannot be avoided even the rate is below the Capacity.

› Shannon bound is no longer accurate!

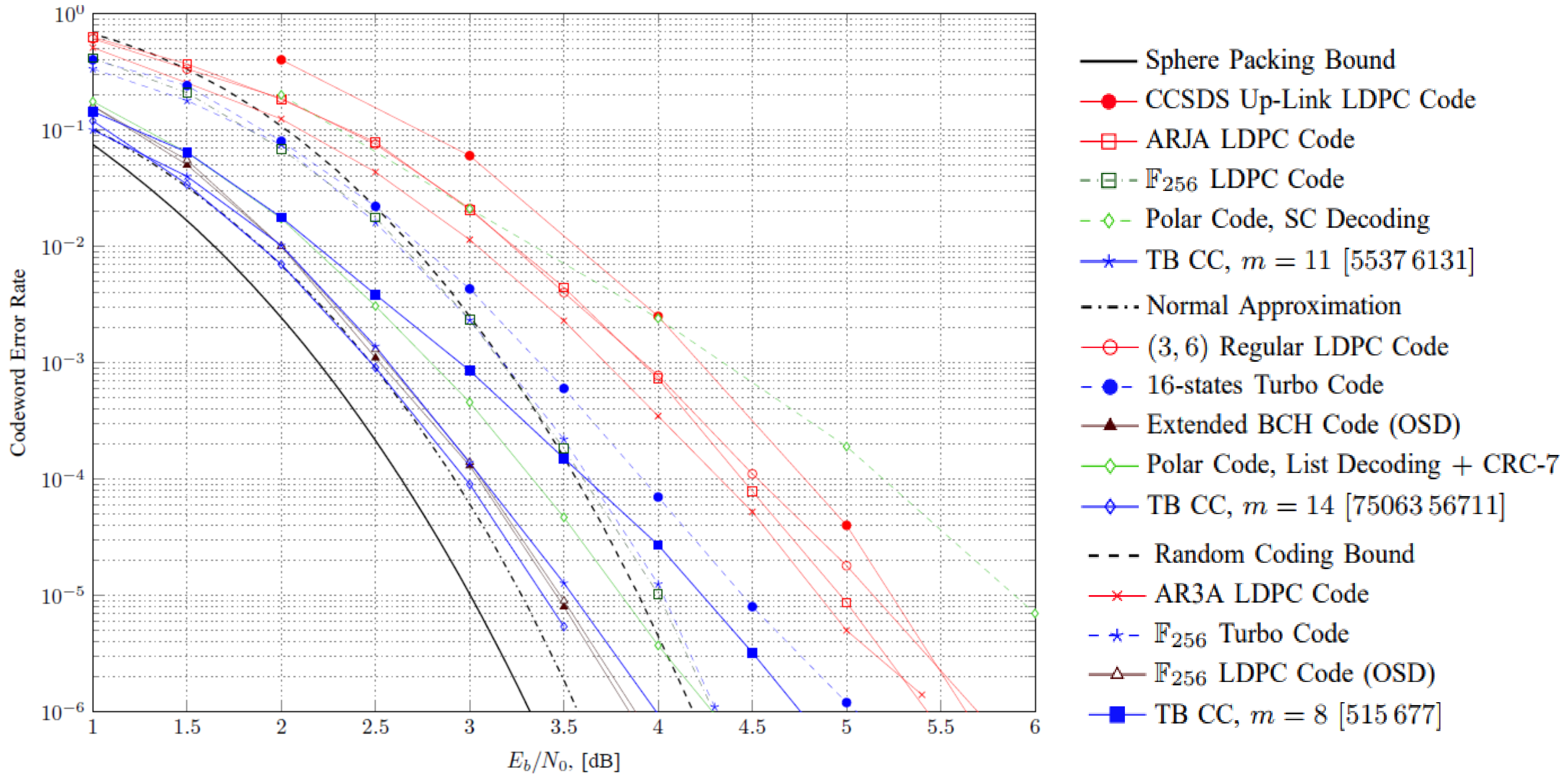
› Polyanskiy-Verdu-Poor Bound:

$$R = \frac{v}{m} = \log_2(1+\gamma) - Q^{-1}(\varepsilon)\sqrt{\left(1 - \frac{1}{(1+\gamma)^2}\right)(\log_2 e)^2/m}$$

› New bound requires new designs of communication algorithms and protocols

Channel coding for uRLLC: a revisit

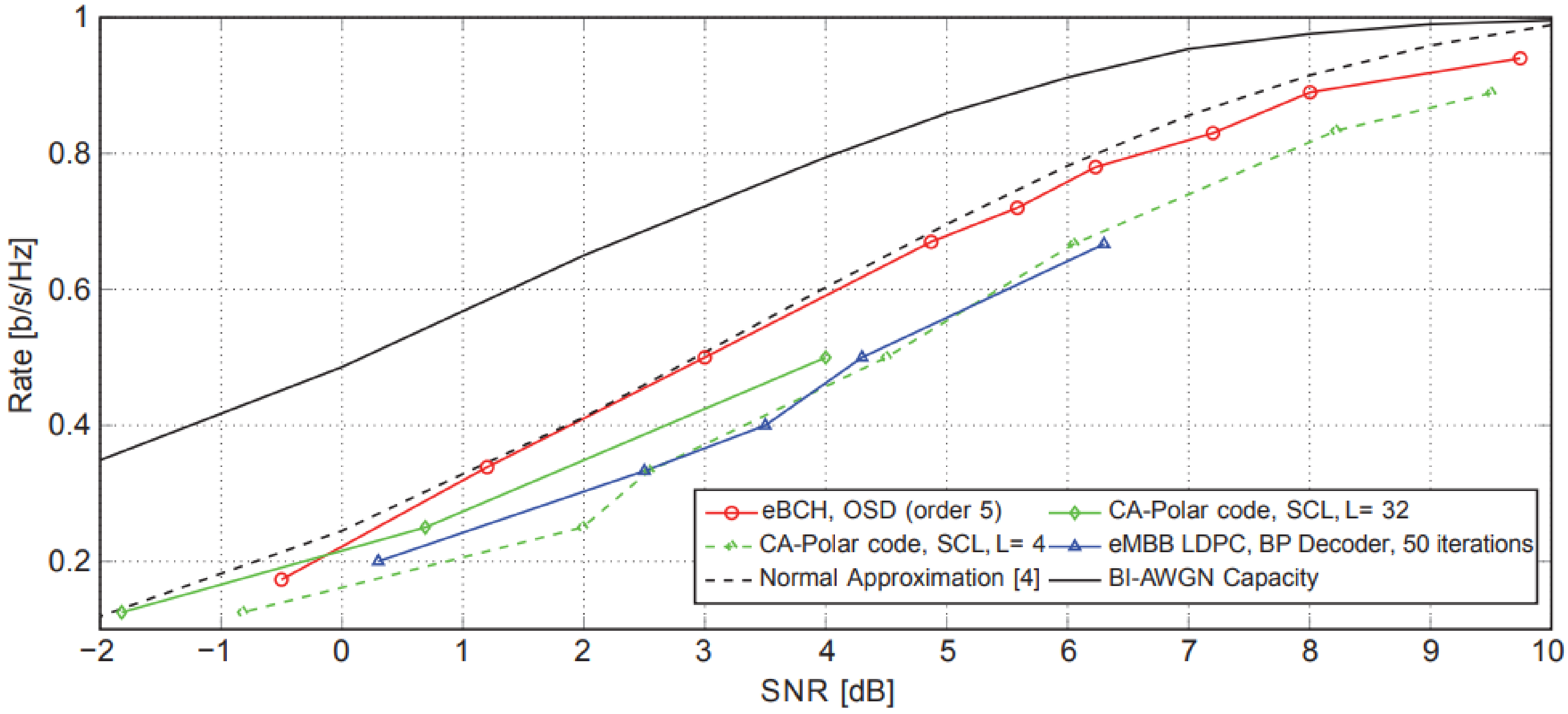
- •LDPC + BP: Poor performance for short codes, simple decoding
- •BCH + MLD: Best error-correction but exponential decoding complexity



Liva, Gianluigi, et al. "Code design for short blocks: A survey." *arXiv preprint arXiv:1610.00873* (2016).

Comparison

Rate performance of different candidate codes at a BLER of 10–4 when the codeword length is N = 128.



M. Shirvanimoghaddam *et al.*, "Short Block-Length Codes for Ultra-Reliable Low Latency Communications," in *IEEE Communications Magazine*, vol. 57, no. 2, pp. 130-137, February 2019.

Receiver processing delay

Receive Modules	B = 1.4MHz	B = 5MHz	B = 10MHz
CFO Compensation	0.0010	0.0023	0.0037
FFT	2.9004e-04	6.2917e-04	8.3004e-04
Disassemble Reference Signal	1.2523e-04	2.2708e-04	3.1685e-04
Channel Estimation (MMSE)	0.0015	0.0141	0.0878
Disassemble Symbols	0.0013	0.0045	0.0087
MIMO Detection (MMSE-SIC)	0.0028	0.0242	0.0760
SINR Calculation	2.4947e-04	6.6754e-04	0.0012
Layer Demapping	4.3253e-05	1.0988e-04	3.8987e-04
Turbo Decoding	0.0129	0.0498	0.1048
Obtained Throughput	2.2739Mbps	10.073Mbps	20.41Mbps

Decoder Candidates for URLLC in 6G

Performance metrics for decoder design for URLLC

- **Optimality:** defined as the ability of decoders to achieve the code ML performance.
- **Complexity:** needs to be as low as possible to conserve the overall budget of latency.
- **Universality:** refers to the ability to decode any linear block codes.

Main Candidate Decoders for URLLC

- OSD: Ordered Statistics Decoding
- GRAND: Guessing Random Additive Noise Decoding
- SCL/SCS: Successive Cancellation List/Stack Decoding

All these decoders have the (near) optimality; however, only OSD and GRAND have the universality.

Ordered Statistic Decoder

The decoder receives a message encoded by $C(N, K, d_{min})$ from channel.

Hard-decision results: $\mathbf{y} = [y_1 \quad y_2 \quad y_3 \quad \dots \quad y_N]$

- **ML Decoding:** Flip every bit to find the maximum-likelihood result （complexity 2^N）
- **OSD:** Only flip the first K bits, and generate the rest bits by encoding

$$\mathbf{c} = [\mathbf{y}_{[1,K]} \oplus \mathbf{e}]\mathbf{G}$$

Test error pattern (TEP)

- OSD's Techniques:
  - puts high reliable bits at the first K positions by permutations (Linear property of linear Code)
  - Permute elements of $\mathbf{y}$, columns of $\mathbf{G}$ according to Channel LLR (reliability) simultaneously
  - Output codeword estimation by applying inverse permutation.

The total number of TEPs $\binom{K}{0} + \binom{K}{1} + \binom{K}{2} + \dots + \binom{K}{d_{min}/4} \ll 2^N$

C. Yue, M. Shirvanimoghaddam, B. Vucetic, and Y. Li, "A revisit to ordered statistics decoding: Distance distribution and decoding rules," IEEE Trans. Inf. Theory, vol. 67, no. 7, 2021.

GRAND Decoder

- GRAND is a universal decoder capable of rigorously achieving the ML performance.
- In BSC, Guesses a possible noise sequence w′ to retrieve

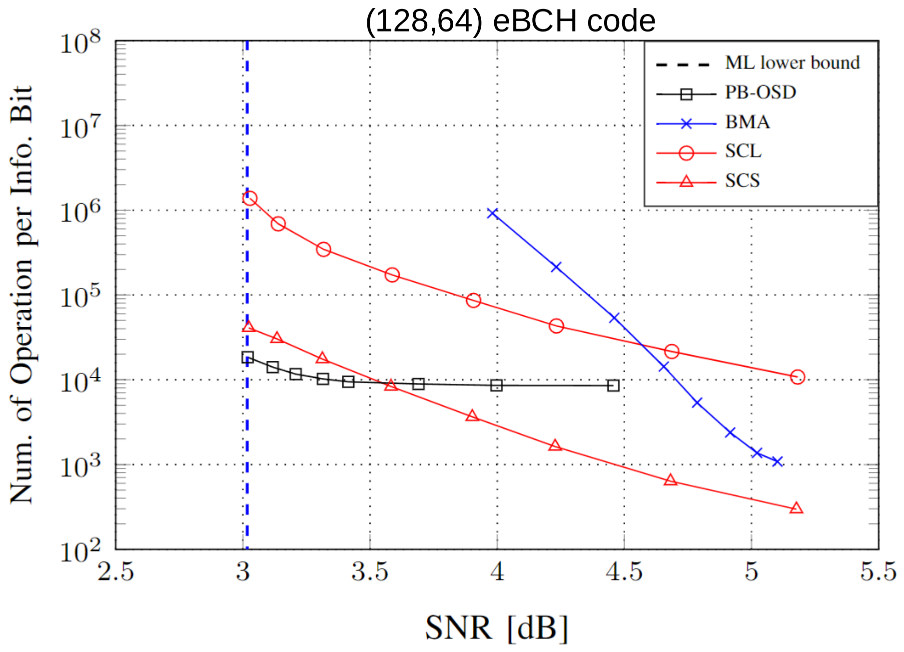
$$c' = y \oplus w',$$

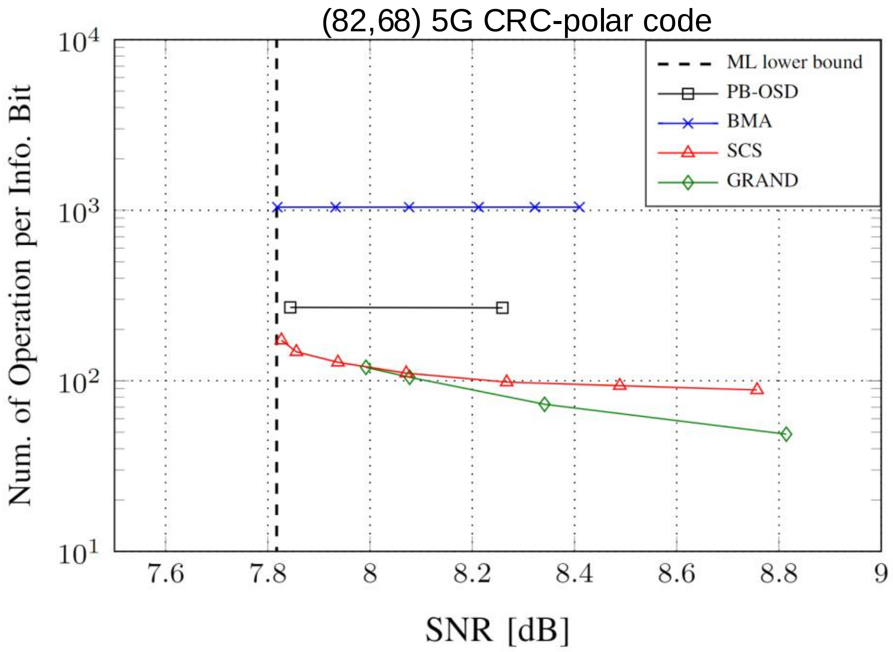
where c′ is a valid estimate and w′ is the noise sequence guess.
- Check a group of noise sequences, each check is referred to as a noise query.
- Query noise sequences in ascending order of the Hamming weight
- First discovered codeword will have the lowest Hamming distance to y, which is the ML solution.
- Potential to be less complex than the ML decoding, because it stops at the first discovered codeword.
- Efficient when SNR or the coding rate k/n is high.
- High SNRs ensure the low Hamming weight of the true noise sequence
- High coding rates increase the probability of passing codebook membership checks.

K. R. Duffy, J. Li, and M. M´edard, "Capacity-achieving guessing random additive noise decoding," IEEE Trans. Inf. Theory, vol. 65, no. 7, pp. 4023–4040, 2019.

Decoder Candidates for URLLC in 6G

Decoding complexity vs. Gap to ML decoding bound [1]





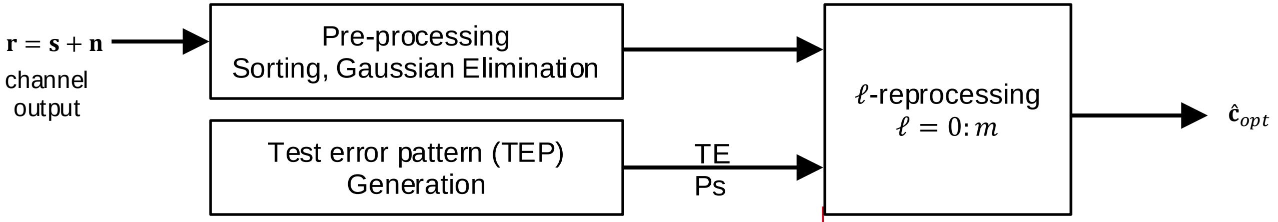
- OSD is competitive in decoding length-128 eBCH codes and 5G CRC-polar codes
- GRAND is efficient only for high-rate codes

[1] C. Yue, V. Miloslavskaya, M. Shirvanimoghaddam, B. Vucetic, and Y. Li, “Efficient decoders for short block length codes in 6g URLLC,” arXiv preprint arXiv:2206.09572, 2022.

Low complexity Universal Decoder

Probability-based Universal Ordered Statistics Decoder (OSD)

PB-OSD has significantly lowered complexity than existing OSD-based approaches



- Soft Individual Stopping Rule (SISR) [2]
- Soft Discarding Rule (SDR) [2]
- Tree-based TEP search

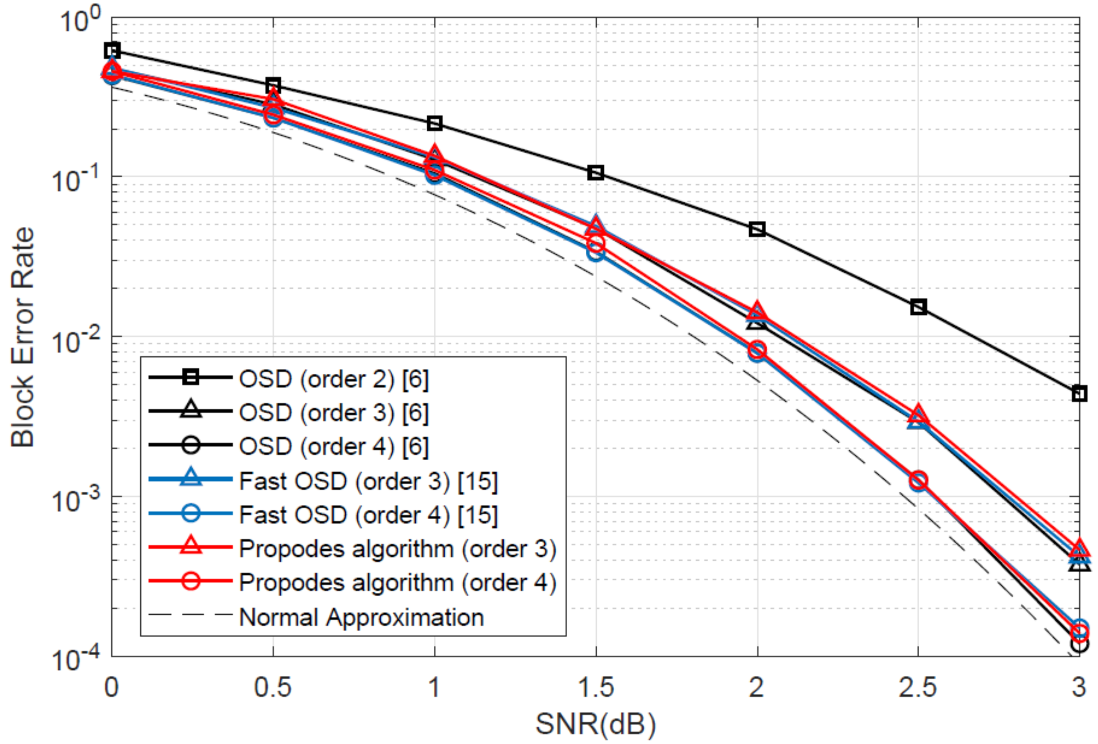
- SISR identified the correct codeword estimate and terminate the decoding early
- SDR avoids generating unpromising codeword estimates
- Tree-based TEP search decreases the overhead of computing SISR and SDR

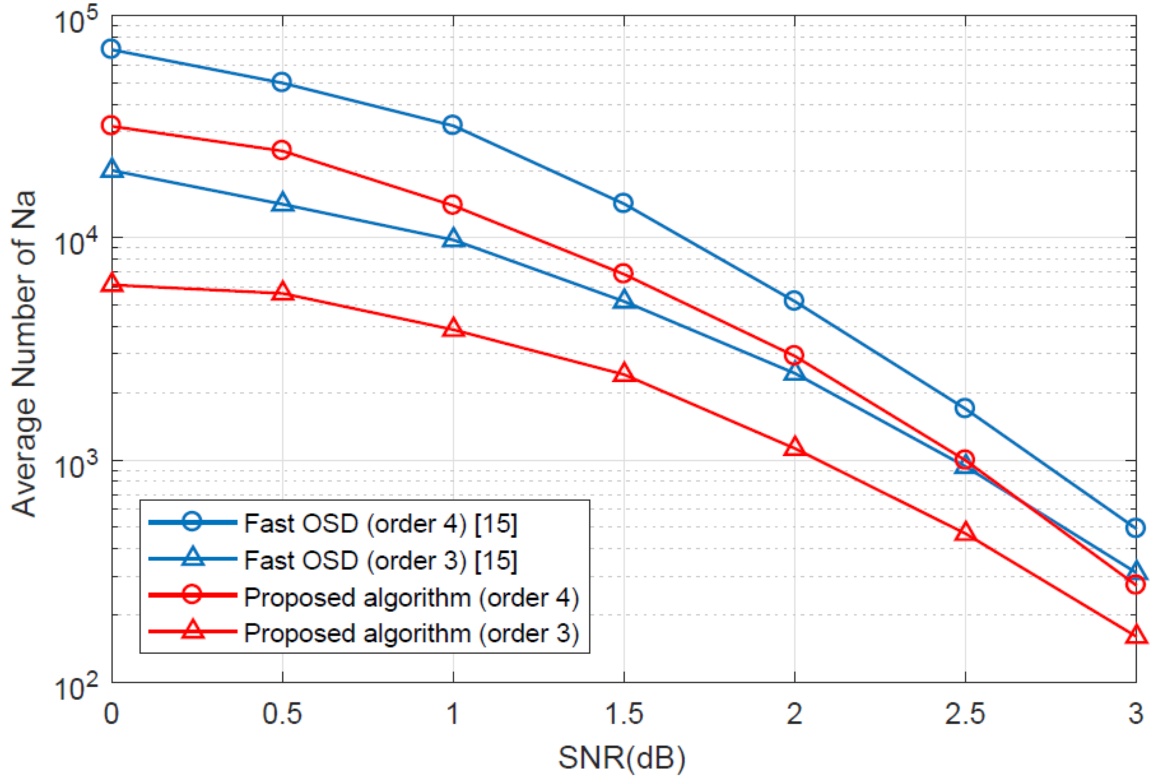
[1] C. Yue, M. Shirvanimoghaddam, B. Vucetic, and Y. Li, “A revisit to ordered statistics decoding: Distance distribution and decoding rules,” IEEE Trans. Inf. Theory, vol. 67, no. 7, 2021

[2] C. Yue, M. Shirvanimoghaddam, G. Park, O.-S. Park, B. Vucetic, and Y. Li, “Probability-based ordered-statistics decoding for short block codes,” IEEE Commun. Lett., 2021

Performance and Complexity Comparison

(128,64,22) eBCH code

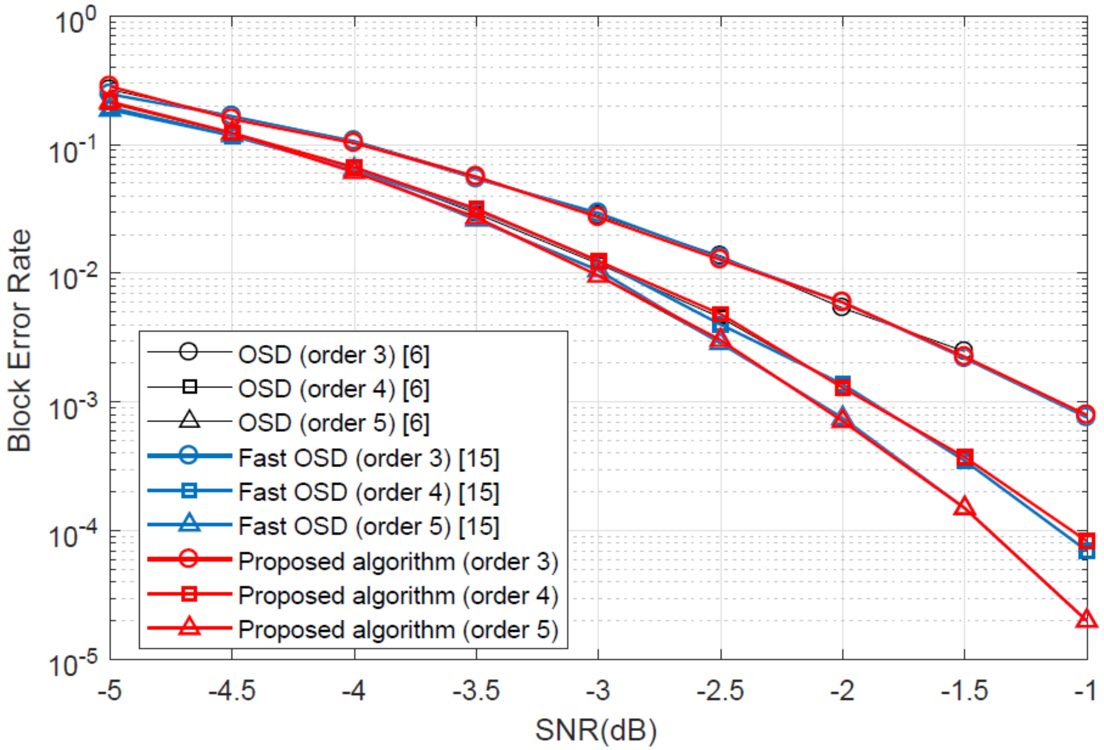


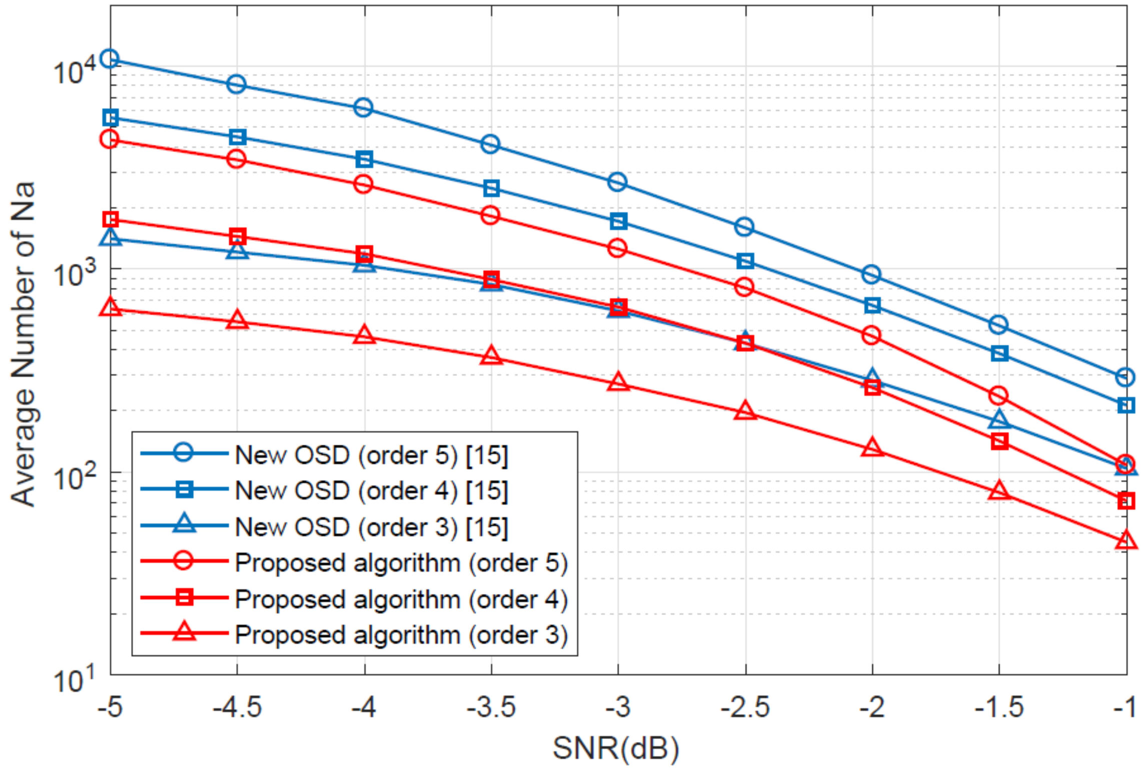


- C. Yue, M. Shirvanimoghaddam, Y. Li and B. Vucetic, "A Revisit to Ordered Statistics Decoding: Distance Distribution and Decoding Rules" *IEEE Transactions on Information Theory*, 2021
- C. Yue, M. Shirvanimoghaddam, Y. Li and B. Vucetic, Probability-Based Ordered-Statistics Decoding for Short Block Codes. IEEE Commun. Lett. 2021.

Performance and Complexity Comparison

(128,22,48) eBCH code

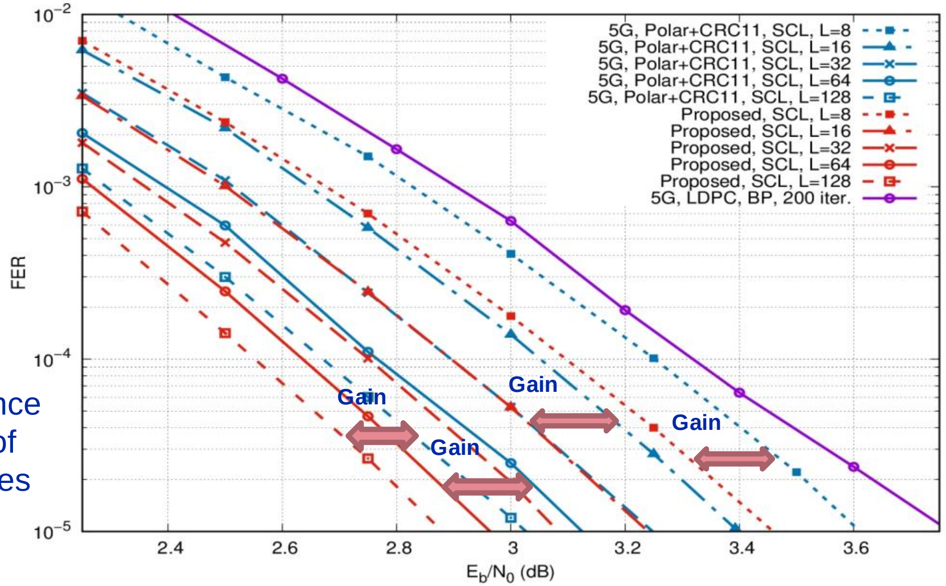




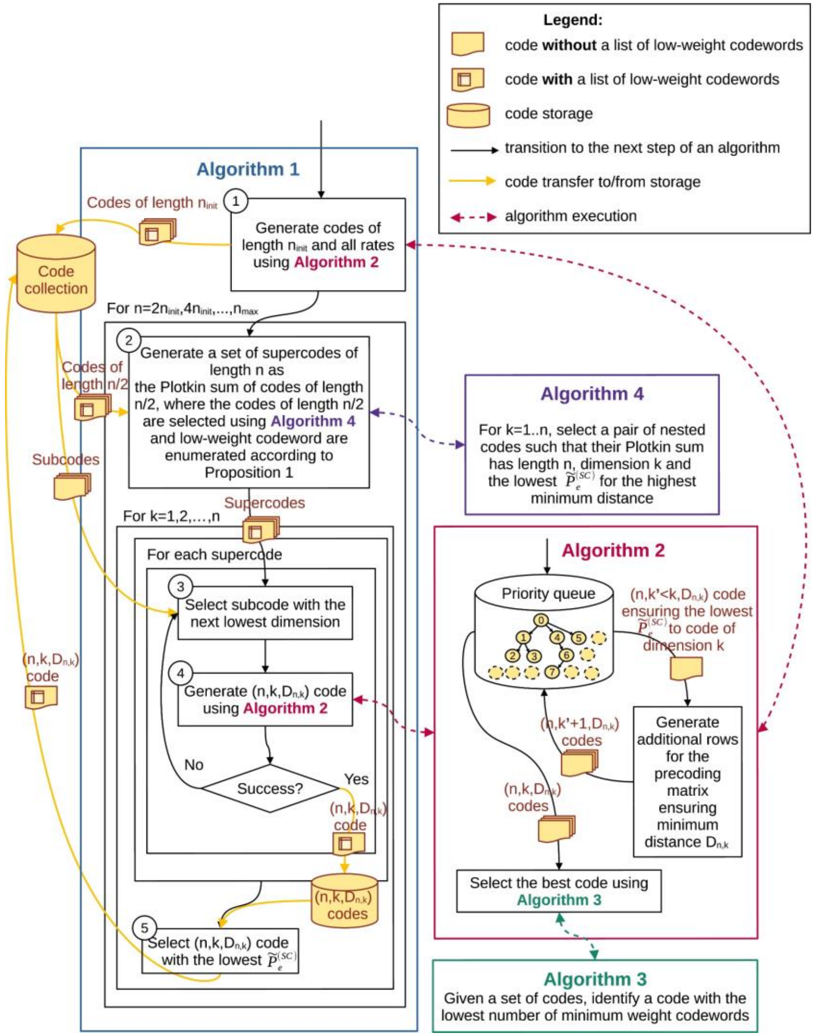
- C. Yue, M. Shirvanimoghaddam, Y. Li and B. Vucetic, "A Revisit to Ordered Statistics Decoding: Distance Distribution and Decoding Rules" *IEEE Transactions on Information Theory*, 2021
- C. Yue, M. Shirvanimoghaddam, Y. Li and B. Vucetic, Probability-Based Ordered-Statistics Decoding for Short Block Codes. IEEE Commun. Lett. 2021

Proposed Precoded Polar Codes of Lengths 128 and 256

- Recursive design of precoded polar codes optimized for successive cancellation list (SCL) decoding
- Inspired by the recursive structure of Reed-Muller codes
- High error-correction capability is achieved by explicitly enumerating low-weight codewords and eliminating them
- The optimization complexity is reduced by introducing a number of constrains, e.g., a subcode and a supercode
- Outperform polar codes with various CRC polynomials
- 0.2 dB performance gain compared to 5G polar codes with CRC11



FER performance comparison of (256,128) codes



V. Miloslavskaya, B. Vucetic, Y. Li, G. Park, and O.-S. Park, “Recursive Design of Precoded Polar Codes for SCL Decoding”, IEEE Transactions on Communications, vol. 69, pp. 7945–7959, 2021.

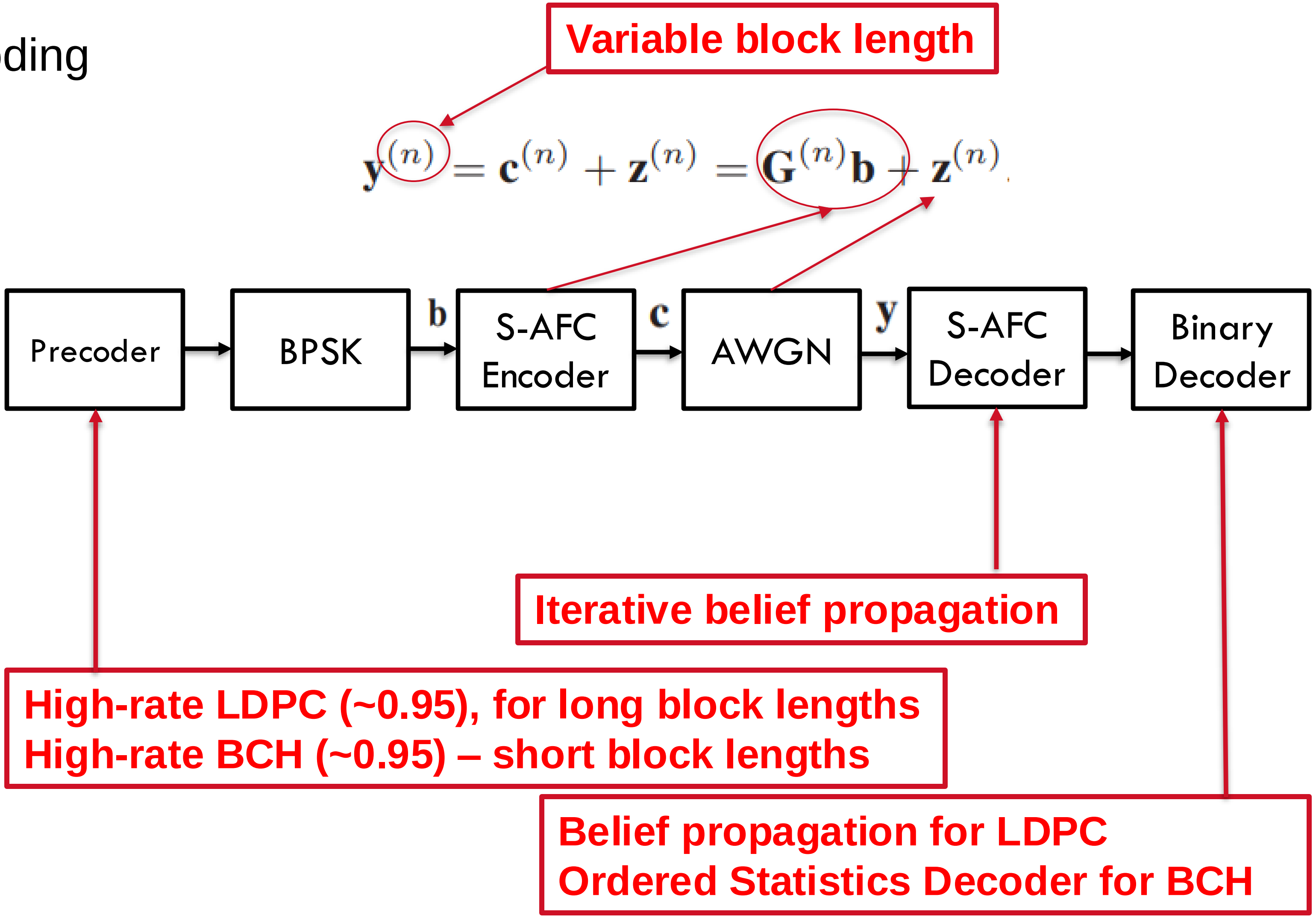
Avoid Retransmission Delay via Rateless Code

- Each retransmissions introduces 8ms delay
- AMC channel feedback - further 5–8 ms delay
- ~30–50% signaling overhead for payloads of length 200 symbols with 7–10 users
- Analog fountain codes (AFC) to avoid the retransmission and channel feedback
- AFC naturally adapt code rates to channel conditions and
- AFC approaches the capacity over a range of SNRs without CSIT

- M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Near-capacity adaptive analog fountain codes for wireless channels," IEEE Communications Letters, vol. 17, no. 12, Dec. 2013, pp. 2241-2244.
- R. Abbas, M. Shirvanimoghaddam, Y. Li and B. Vucetic, Novel Design for Short Analog Fountain Codes, IEEE Communications Letters, 31 March 2019.

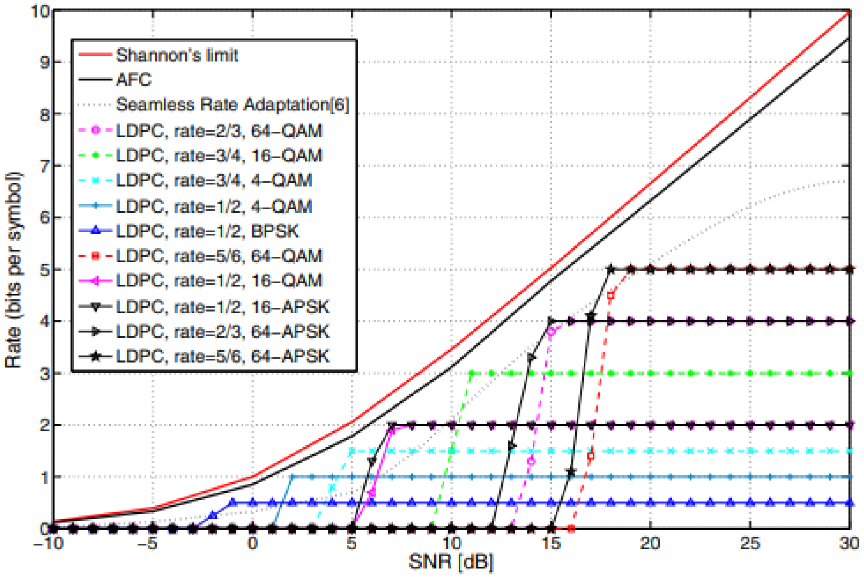
Encoding

$$\mathbf{y}^{(n)} = \mathbf{c}^{(n)} + \mathbf{z}^{(n)} = \mathbf{G}^{(n)}\mathbf{b} + \mathbf{z}^{(n)}$$

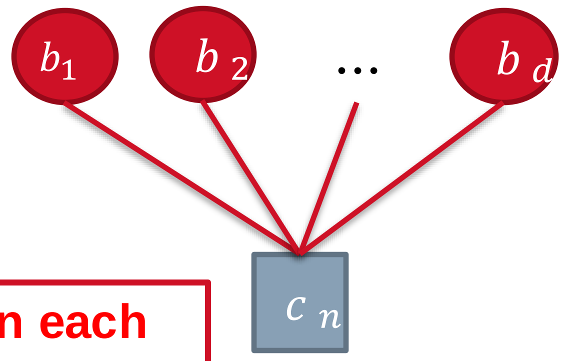


Performance of AFC in the Asymptotic Block-Length Regime

Near-Capacity Performance when k = 10,000 bits



- AFC can achieve near capacity performance for asymptotically long blocks
- The design of AFC is based on a degree d and a weight set W(set of real numbers)

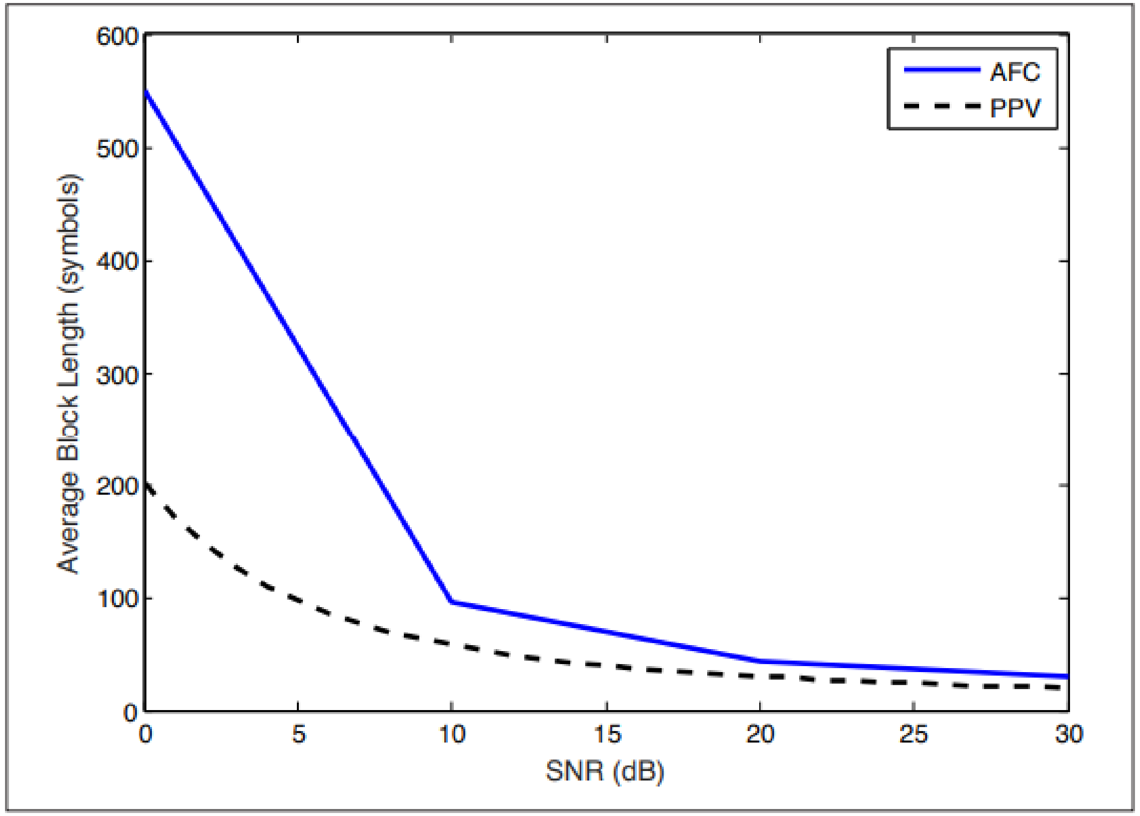


Set of chosen information symbols in each encoding stage is random!

[P-6] M. Shirvanimoghaddam, Y. Li and B. Vucetic, "Near-Capacity Adaptive Analog Fountain Codes for Wireless Channels," in *IEEE Communications Letters*, vol. 17, no. 12, pp. 2241-2244, December 2013. https://doi.org/10.1109/LCOMM.2013.101813.131972

Performance of AFC in the Short Block-Length Regime

Poor performance at low SNR in comparison to the Polyanskiy-Poor and Verdu Bound (Normal Approximation), k = 192

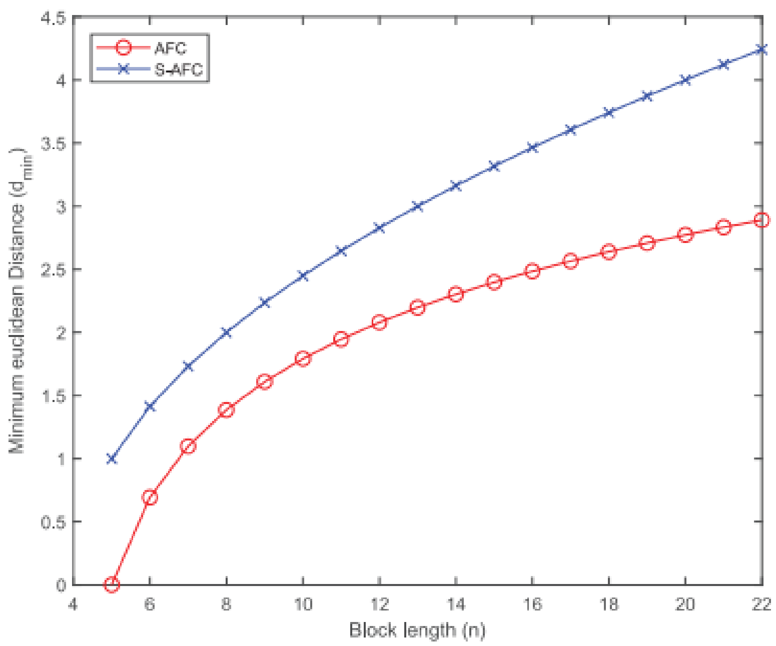


[P-1] R. Abbas, M. Shirvanimoghaddam, T. Huang, Y. Li and B. Vucetic, "Novel Design for Short Analog Fountain Codes," in *IEEE Communications Letters*, vol. 23, no. 8, pp. 1306-1309, Aug. 2019. https://doi.org/10.1109/LCOMM.2019.2910517

Design of Short AFC

d	$\mathcal{W}$
2	{0.8949, 0.4462}
3	{0.8736,0.4354,0.2172}
4	{0.8686,0.4329,0.2159,0.1075}
5	{0.8674,0.4322,0.2156,0.1073,0.05312}
6	{0.8671,0.4320,0.2155,0.1073,0.0531,0.0261}
7	{0.8670,0.4320,0.2155,0.1073,0.0531,0.0261,0.0125}
8	{0.8670, 0.4320, 0.2155, 0.1073, 0.0531, 0.0261, 0.0125, 0.0058}

TABLE I
DESIGNED WEIGHT SETS FOR S-AFC ($\delta = 0.001$) [8]



We propose a new weight set design based on two rules:

- Power constraint
- Distance constraint

$$d_{\min}^{(n)} = \min_{1\leq j\leq k} \left\{ 2\sqrt{\sum_{i=1}^{n} (g_{i,j})^2} \right\}$$

We want to maximize the minimum Euclidean distance of the constellation while ensuring the power constraint is met!

R. Abbas, M. Shirvanimoghaddam, T. Huang, Y. Li and B. Vucetic, "Novel Design for Short Analog Fountain Codes," in *IEEE Communications Letters*, no. 8, pp. 1306-1309, Aug. 2019.

Performance of Short AFC

Much better performance in comparison to original AFC

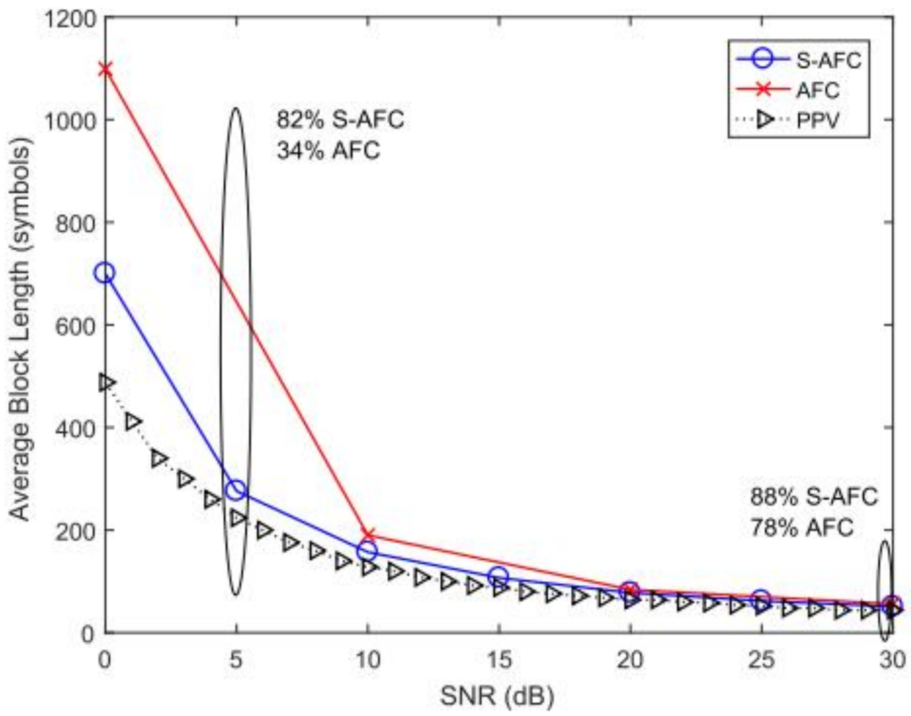


Fig. 2. Average block length of S-AFC in comparison to AFC (both with a (256,192) LDPC precoder) and the PPV bound ($\epsilon = 10^{-4}$). The percentages shown represent the percentage of block length calculated from the PPV bound at a given SNR to that achieved by S-AFC and AFC.

R. Abbas, M. Shirvanimoghaddam, T. Huang, Y. Li and B. Vucetic, "Novel Design for Short Analog Fountain Codes," in *IEEE Communications Letters*, no. 8, pp. 1306-1309, Aug. 2019.

Performance of Short AFC

Orders of magnitude improvement for very short blocks

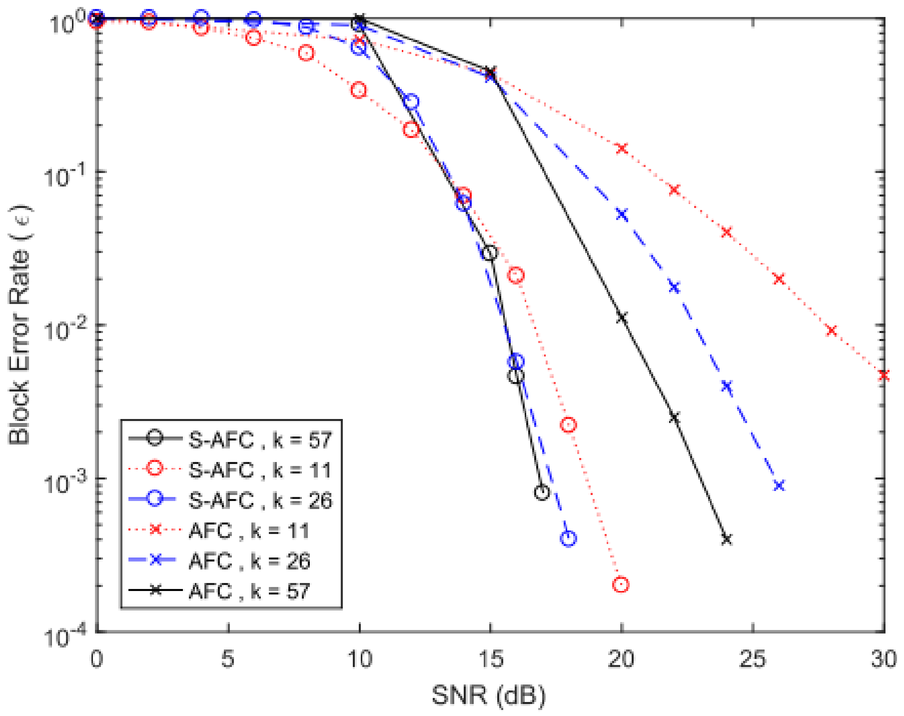


Fig. 5. The block error rate of S-AFC and AFC with single-error correcting BCH precoders for a code rate of 2. Both setups have the same decoder [7].

How do we address URLLC at MAC?

MAC

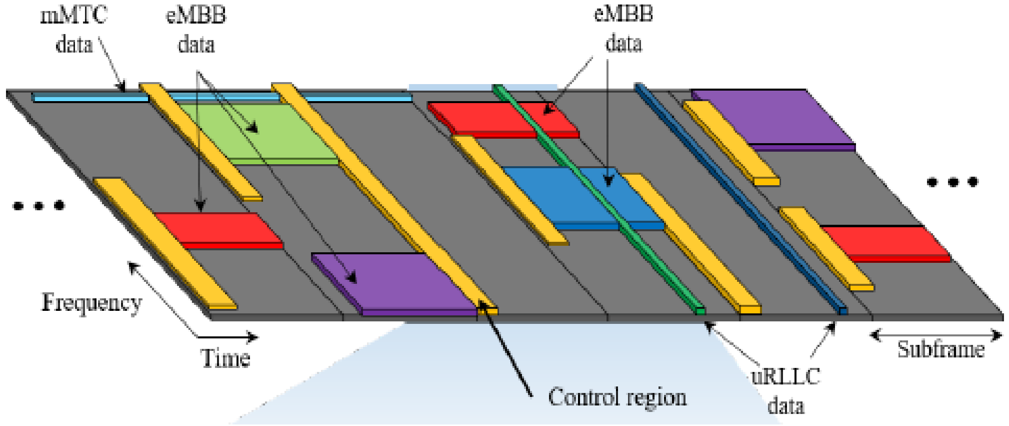
- **Grant-Based Orthogonal Access Techniques**
  - Why can't 4G support URLLC ?
  - Why can't 5G support massive URLLC?
- **Non-Orthogonal Access (NOMA)**
  - Information theoretic bounds
  - Existing NOMA schemes
- **Grant-Free Access**
  - Motivation for Grant-Free Access
  - Performance analysis
  - Different transmission schemes for Grant-Free NOMA
  - Deep Learning Approach for Grant-Free NOMA
- **Cross-Layer Design**
  - Queueing Delay
  - Effective Bandwidth
  - Packet Dropping

Resource reservation for uRLLC

Problem and challenge

In the request contention period, uRLLC needs to compete with other services, leading to uncertain access delay;

In the joint scheduling among uRLLC and other services, uRLCC need to be served immediately, which would interrupt ongoing transmission of other services.



Solution

reserve resources for uRLLC to ensure its immediate transmission

- trade-off between latency and reserved resources
- Prediction and Communication Co-design

Z. Hou, C. She, Y. Li, Z. Li, and B. Vucetic, "Prediction and Communication Co-design for Ultra-Reliable and Low-Latency Communications", IEEE TWC, 2019.

Grant-free access for reducing access delay

- Grant acquisition and random access procedures in current standards are two major sources of delay
- BS needs to first identify the users through contention-based random access.
- Key problems: severe collisions and high latencies when the number of users increases

- In grant-free multiple access, users encode and transmit their IDs and data together without grant acquisition.
- Eliminate the contention and random access phase, significantly reducing the latency, at the expense of larger interference.
- Signals from multiple users are decoded using successive interference cancellation (SIC)

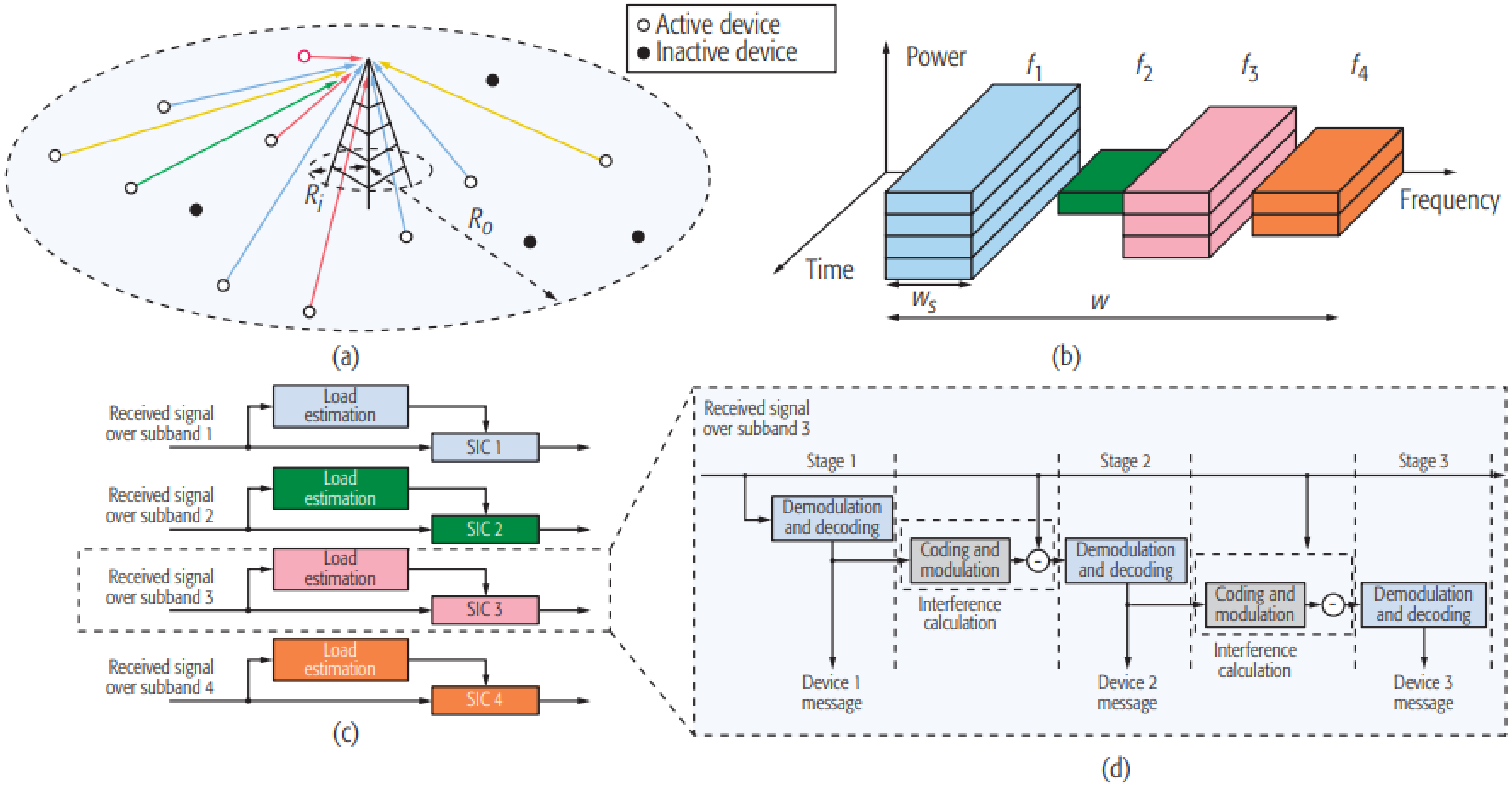
1. R. Abbas, M. Shirvanimoghaddam, Y. Li and B. Vucetic, “On the Performance of Massive Grant-Free NOMA,” Proc. of PIMRC 2017, Oct. 2017, Montreal, Canada.
2. R. Abbas; M. Shirvanimoghaddam, Y. Li, B. Vucetic, “A Novel Analytical Framework for Massive Grant-Free NOMA,” IEEE Transactions on Communications, March 2019.
3. R. Abbas, M. Shirvanimoghaddam, Y. Li, B. Vucetic, “Random Access for M2M Communications with QoS Guarantees,” IEEE Transactions on Communications, July 2017.

Coded Grant Free Multiple Access

– Develop Multi-Layer Grant free Multiple Access schemes with rateless codes

– Develop low complexity decoders for joint user decoding

– Develop deep learning techniques that aim at joint detection and decoding of users in Grant-Free Multiple Access



How do we address URLLC at the Network Layer ?

Network

Network Availability

- Definitions

Multi-connectivity

- Improving Network Availability
- Terrestrial Communications
- Air-to-Ground and Ground-to-Air Communications

Mobile-Edge Computing Systems

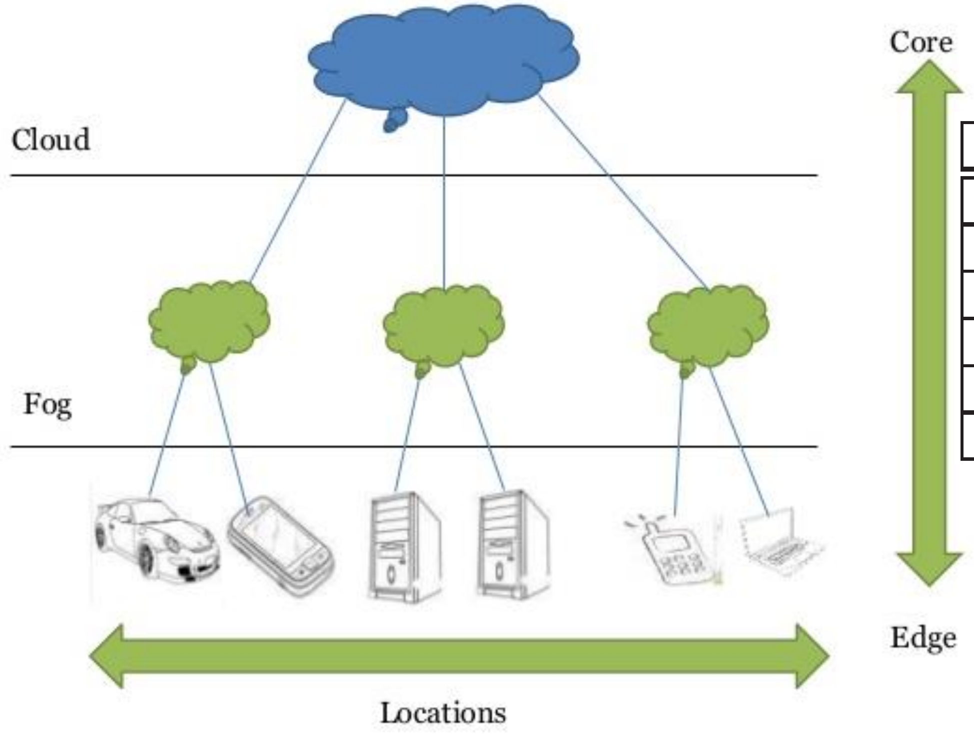
- Analysis of the computing delay using short packets
- Optimisation of user association, computing offloading, and radio resource allocation

Deep Learning for Network Management

- Summary of our recent results
- Motivation of using deep learning in URLLC
- Examples

Edge computing for reducing backhaul latency

- To reduce backhaul latency, computation and content resources should be moved from cloud to the edge – VR, AR, Vehicular networks (driving, urban sensing, content distribution, mobile advertising and intelligent transportation);
- Proximal users are allowed to communicate directly;

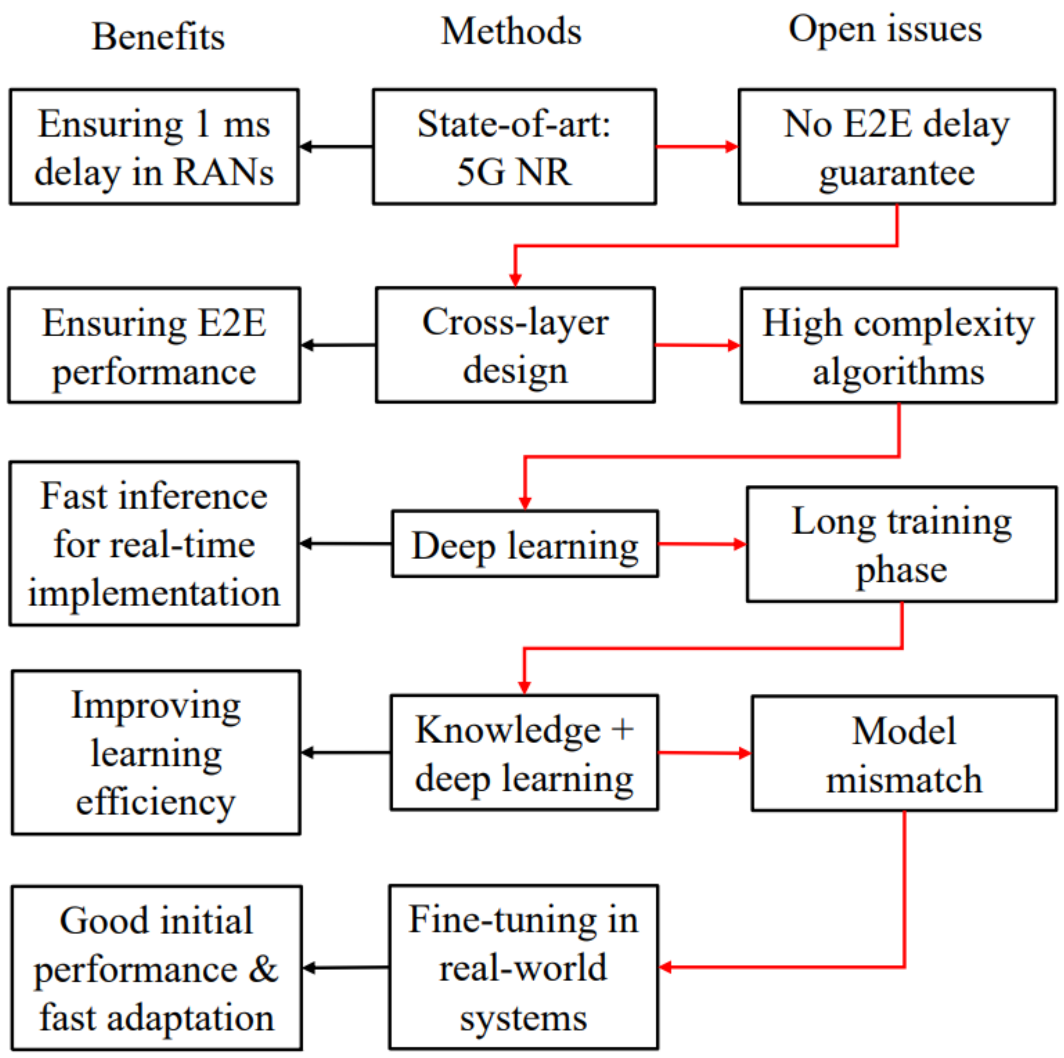


Technical aspect	cloud	Edge computing
Deployment	Centralized	Distributed
Distance to the UE	High	Low
Latency	High	Low
Jitter	High	Low
Computational power	Ample	Limited
Storage capacity	Ample	Limited

1. C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li and B. Vucetic, Optimizing Resource Allocation in Short Blocklength Regime for Ultra-reliable and Low-latency Communications, IEEE Trans. on Wireless Commun., vol. 18, no. 1, pp. 402-415, Jun. 2019.

AI for uRLLC

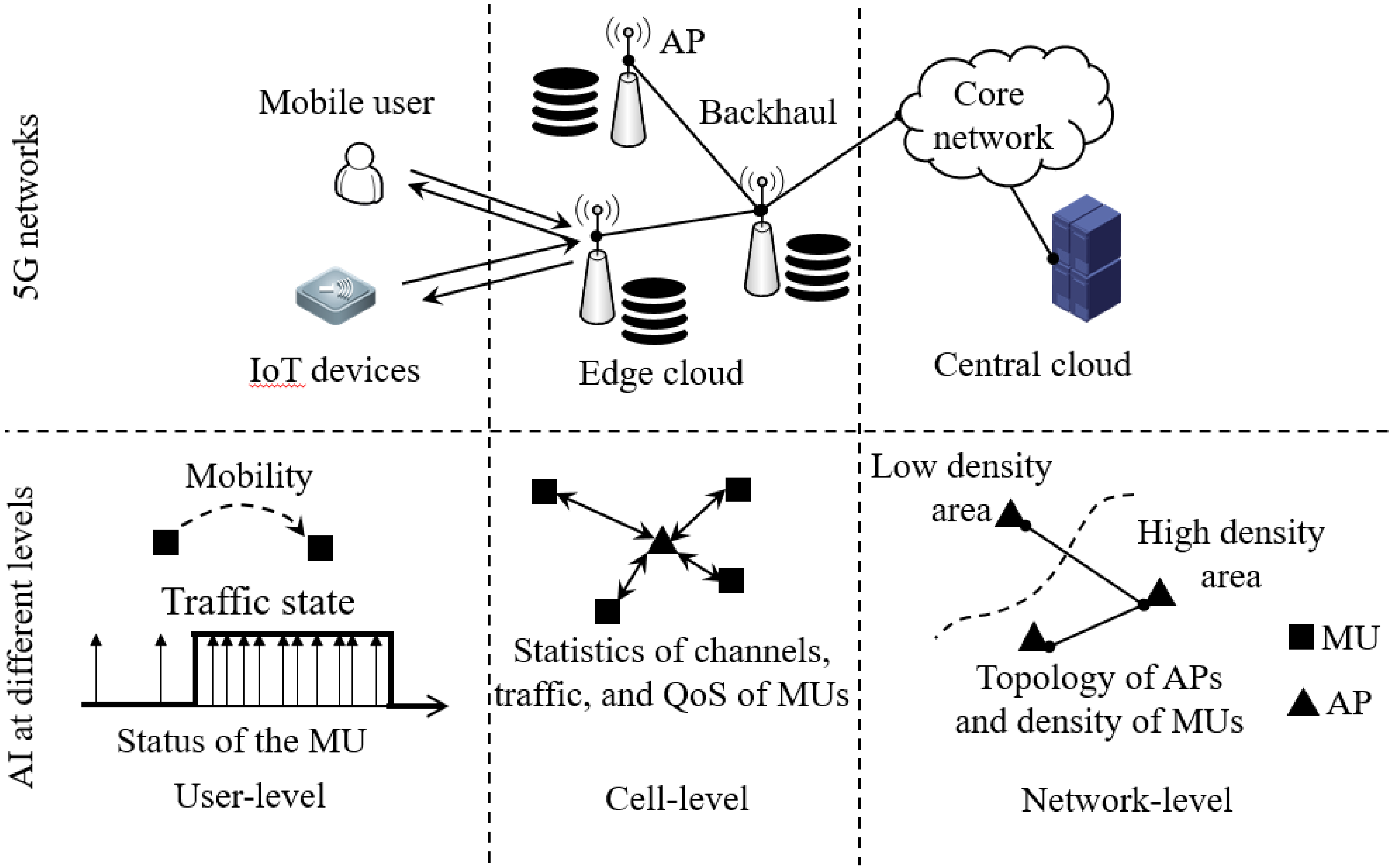
- AI/ML will create more intelligent networks for real-time communications in 6G.
- AI will increase the efficiency and reduce the processing delay of the communication steps.
- Time-consuming tasks, such as handover and network selection, can be performed promptly by using AI.



[1] C. She, et. al., "A Tutorial on Ultrareliable and Low-Latency Communications in 6G: Integrating Domain Knowledge into Deep Learning," **Proceedings of the IEEE**, Mar. 2021.

Deep learning for URLLC

Learning at three levels (User-, Cell-, Network-levels)



Our Results on AI for URLLC

User-level

- Burstiness aware bandwidth reservation (AI for traffic state classification in tactile internet) [N-5]
- Prediction & communication co-design (AI for mobility prediction in remote control applications) [N-8]

Cell-level

- Deep learning for resource management [N-9]
- Deep reinforcement learning for downlink scheduler design

Network-level

- Deep learning for user association [N-11]
- Deep reinforcement learning in software-defined networks

AI/ML based Communication Systems

- Processing blocks interdependent and separate processing is not optimum
- Mathematical modelling with ideal assumptions is suboptimal in practical systems
- Exponential complexity of analytical methods with the number of nodes
- Semantic instead of bit-level transmission needed for high data rates in 6G
- Future networks demand new frameworks

Shannon Model based Design	AI/ML based Design
Separate module optimisation	Joint, end-to-end optimisation
Mathematical model driven with ideal assumptions	Data-driven
Idealised assumptions (e.g., Gaussian, perfect CSI, linear, etc…)	No assumptions; learns directly from data
Static models	Dynamic, adaptive networks
Bit-level transmission	Semantic focus – transmit meaning

Autoencoder based Communication Design

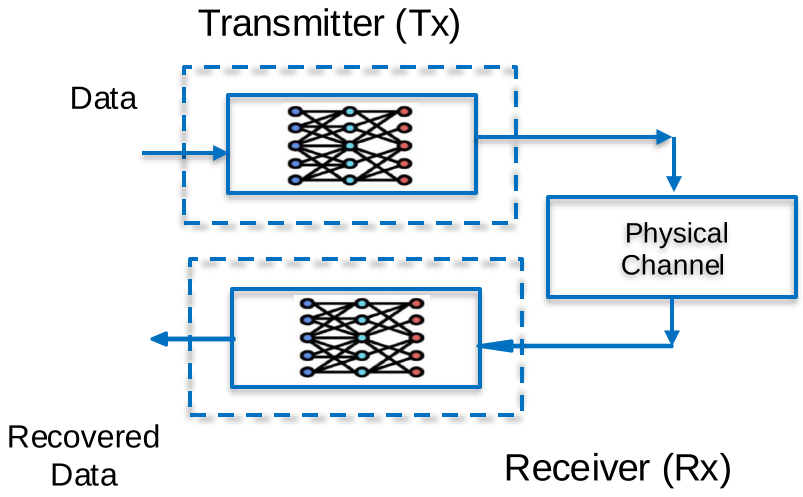
- Sequence of blocks in the Shannon model replaced by NN-based Tx and Rx
- The NNs learn channel characteristics from data
- Joint training optimises the entire system
- System transmits meaning rather than bits

Advantages

- Reduced delay, higher reliability and efficiency

Challenges

- High training complexity, no scalability, no optimality guarantee, deployment incompatibility



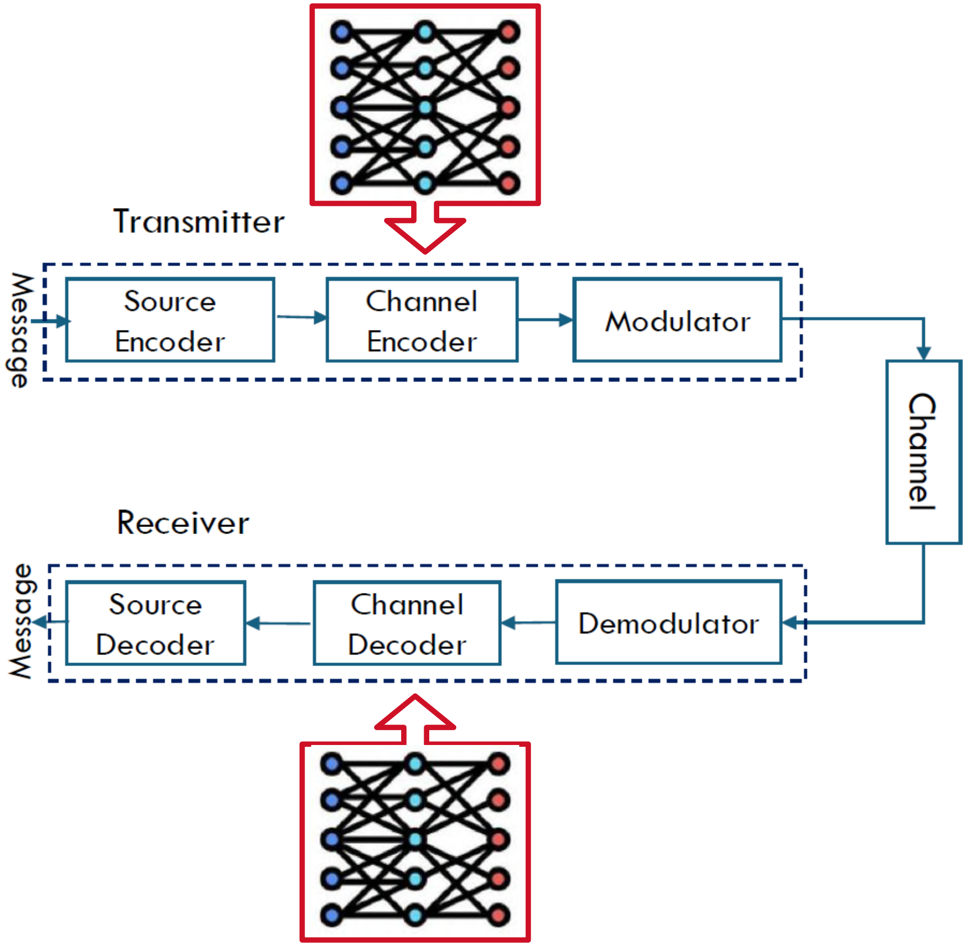
1. D. Gunduz et al., “Beyond transmitting bits: Context, semantics,” IEEE Journal on Selected Areas in Commun., vol. 41, no. 1, pp. 5–41, 2023.

2. Sagduyu, Erpek, Yener et al., Joint Sensing and Semantic Communications with Multi-Task Deep Learning, arXiv, 2024.

Channel Coding for 6G

Existing Autoencoder Channel Coding

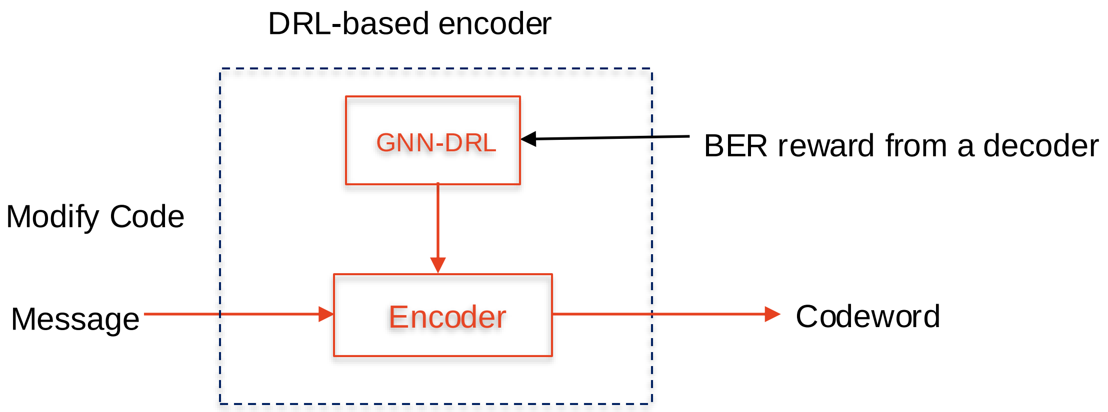
- Use model agnostic FNNs, CNNs and RNNs
- Require a large training set
- Applicable to very short codes



[1] Timothy O'shea and Jakob Hoydis. "An introduction to deep learning for the physical layer", *IEEE Trans. on Cogni. Comms. and Networking* (2017), pp. 563–575.
[2] Karl Chahine et al. "Turbo autoencoder with a trainable interleaver". IEEE ICC pp. 3886–3891.
[3] Guillaume Larue et al. "Neural Belief Propagation Auto-Encoder for Linear Block Code Design". *IEEE Trans. Communi., Vol.* 70, No 11 (2022), pp. 7250–7264.

GNN-DRL Encoder

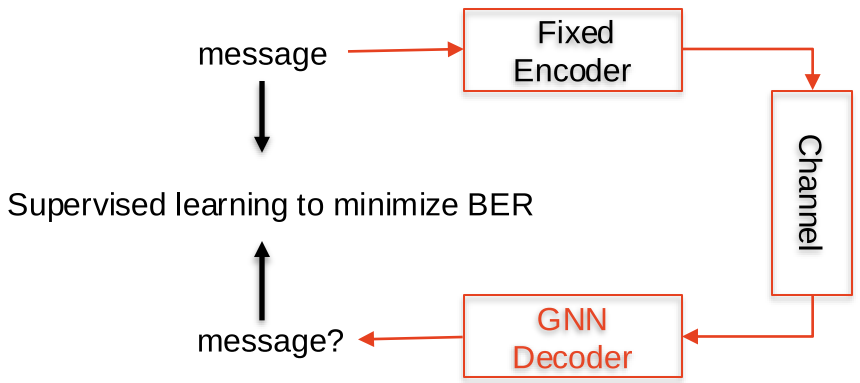
- Our approach: Graph neural network + deep reinforcement learning for the code parity check matrix generation



- Training the encoder to find a better code with a fixed decoder
- Initialised with Standard 5G LDPC codes or BCH codes

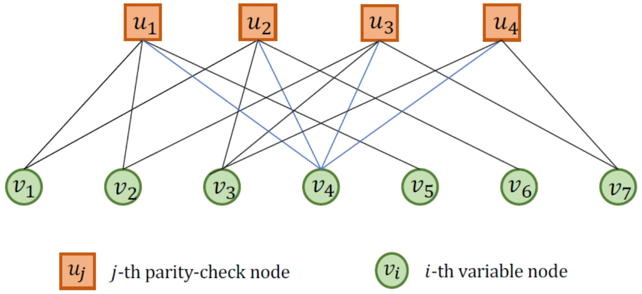
GNN-based Decoder and Its Training

- The decoder is based on GNN and Tanner graph, while the encoder is fixed
- Training with supervised learning, cross-entropy loss function, and gradient descent

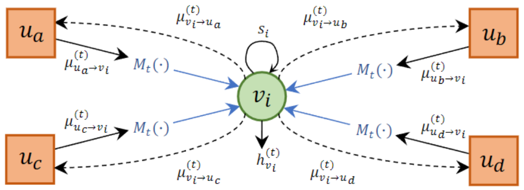


Tian, K., She, C., Yue, C., Li, Y., and Vucetic, B., A Scalable Graph Neural Network Decoder for Short Block Codes, IEEE 2023 IEEE ICC, May 2023, Rome, Italy

GNN in the decoder (supervised learning):



Tanner graph of
code parity check matrix

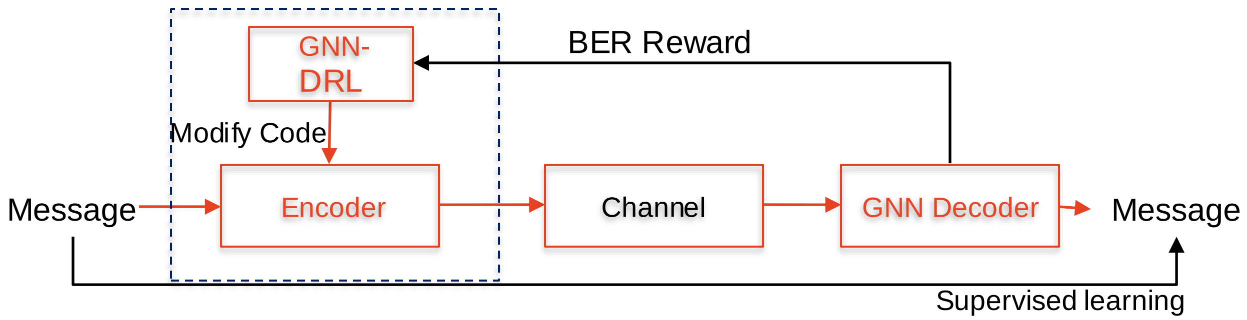


Message-passing decoding by GNN
(Example of a variable node with its four neighbors)

- GNN decoder learns the graph structure, message passing, and operations at the nodes.
- It significantly outperforms BP.

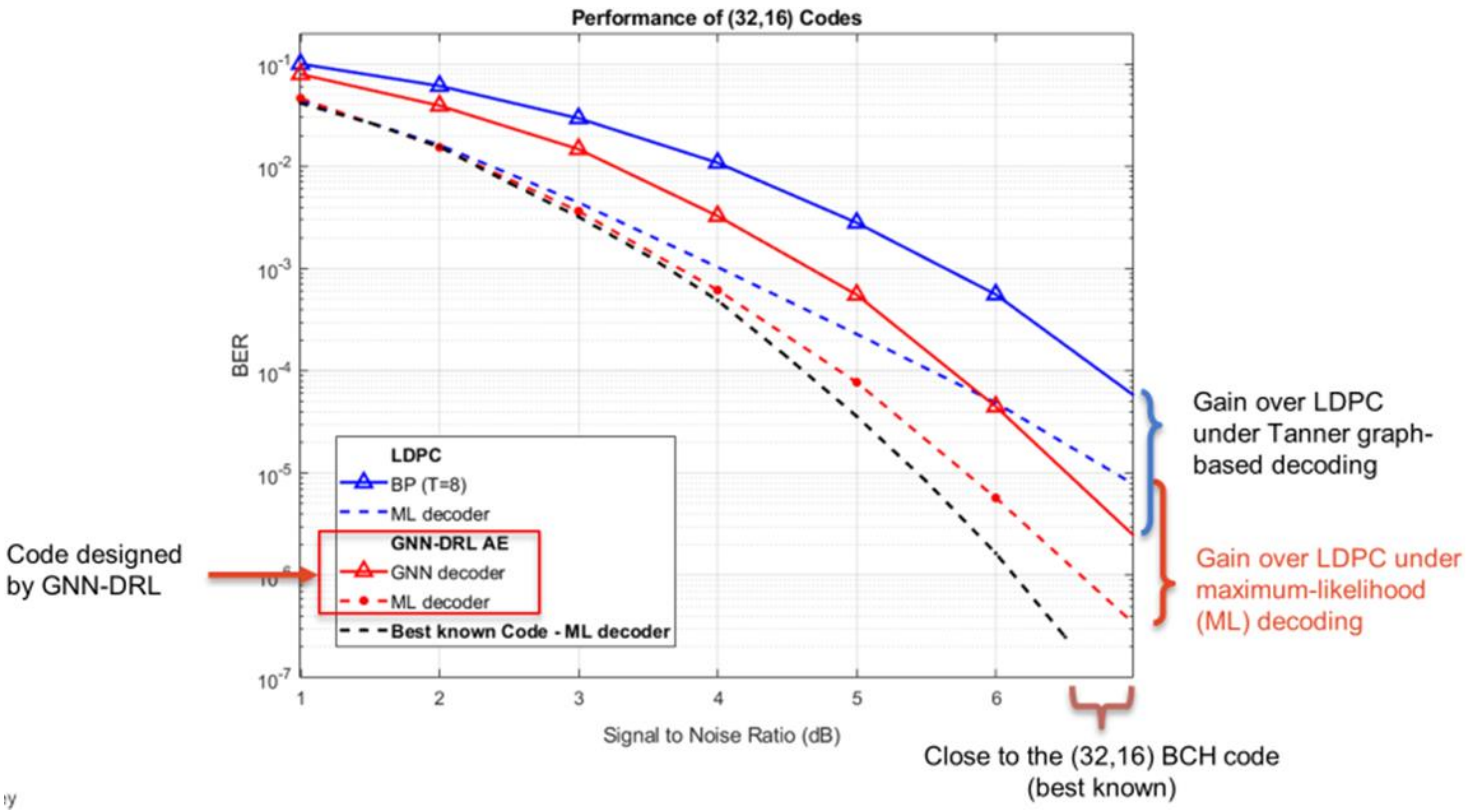
GNN-DRL Auto-Encoder for Channel Coding

- Our approach: Graph neural network + deep reinforcement learning and joint training and optimization



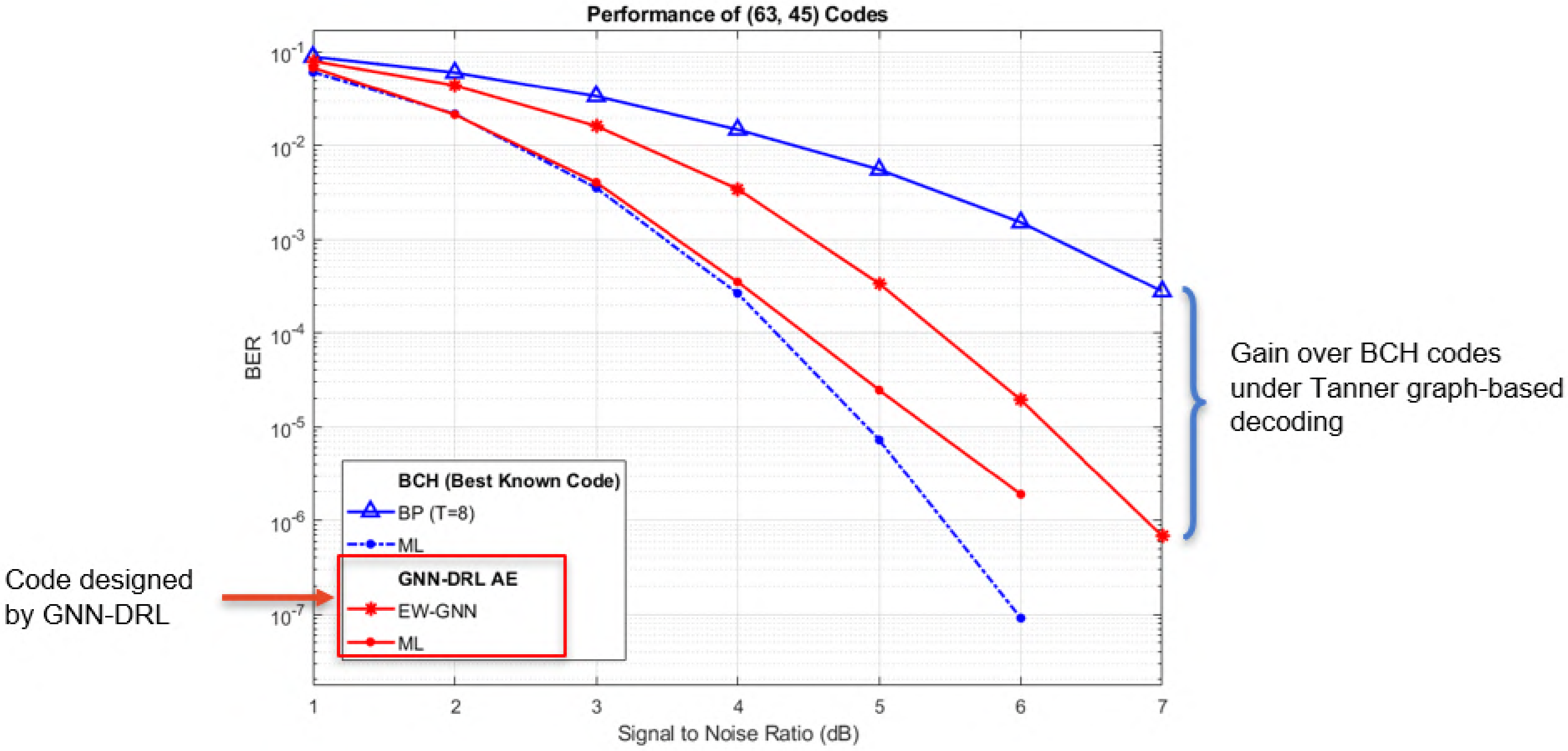
- Joint training to find the best encoder and decoder
- As both the DRL agent and decoder are realized by GNN, the model is adaptive to varying code lengths after a single training.

Performance of the GNN-DRL Auto-Encoder for Channel Coding



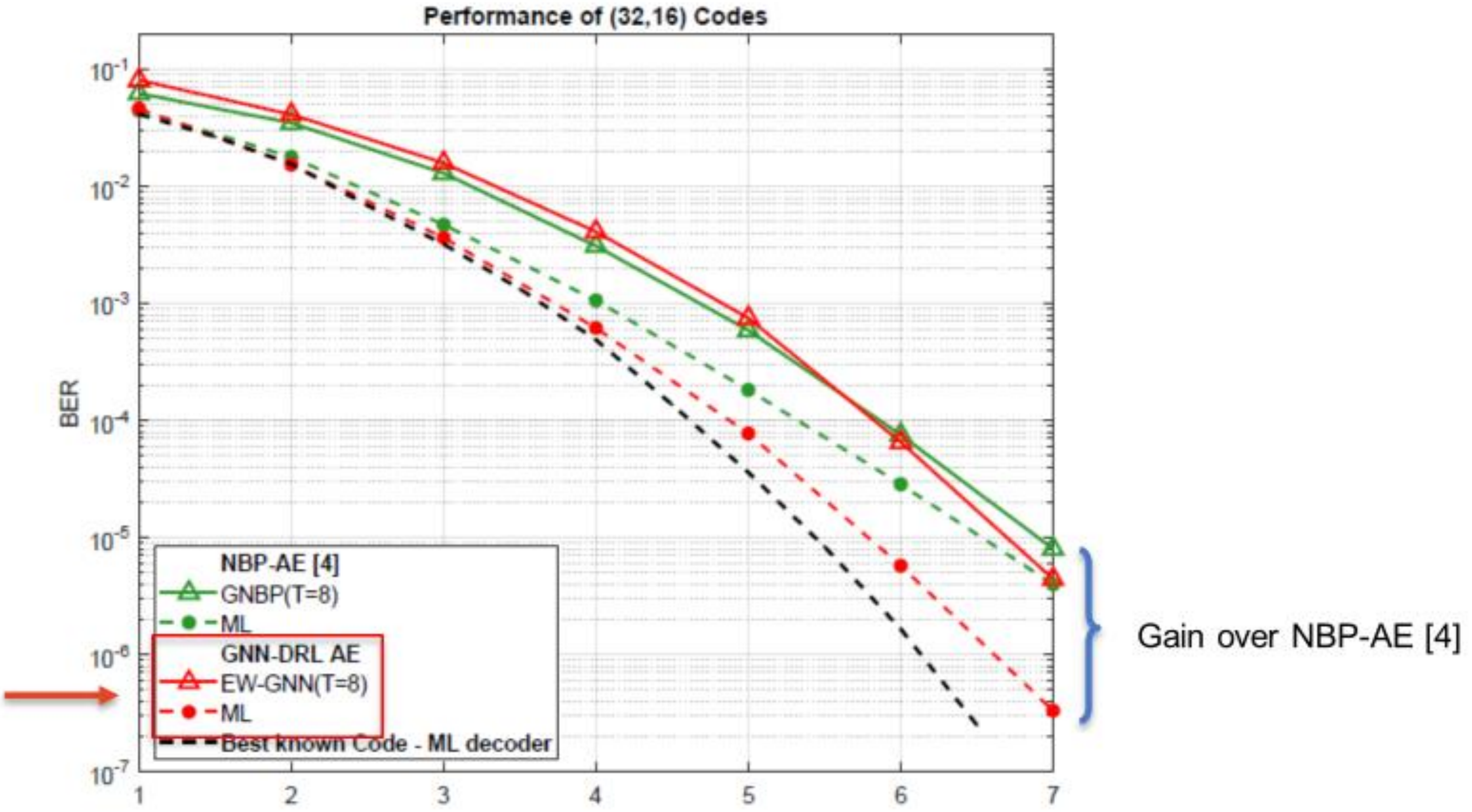
Performance of the GNN-DRL Auto-Encoder for Channel Coding

- Results of our GNN-DRL Auto-encoder【(63,45) BCH initialization】



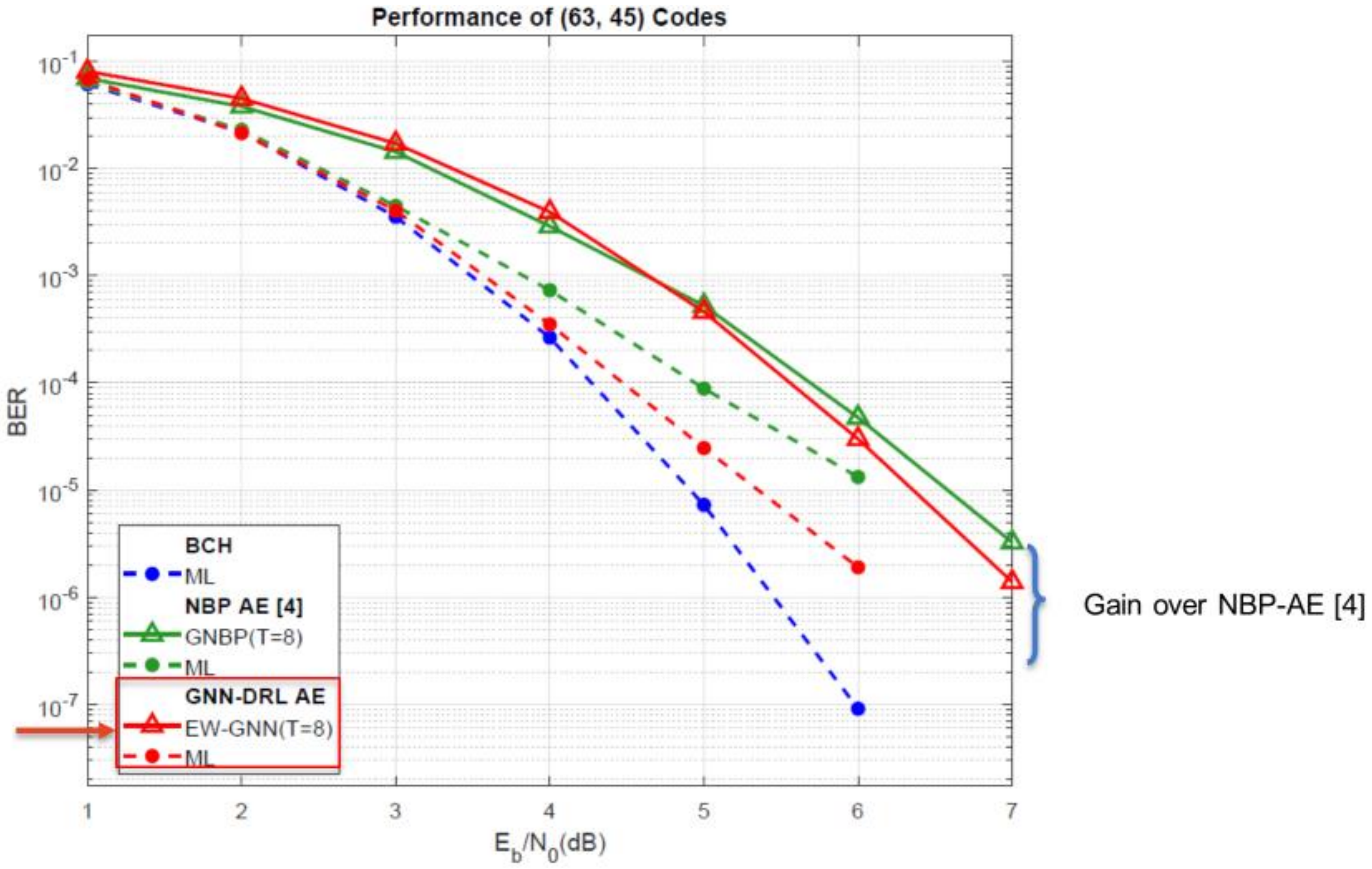
Performance of the GNN-DRL Auto-Encoder for Channel Coding

- Results of our GNN-DRL Auto-encoder【(32,16) BCH initialization】

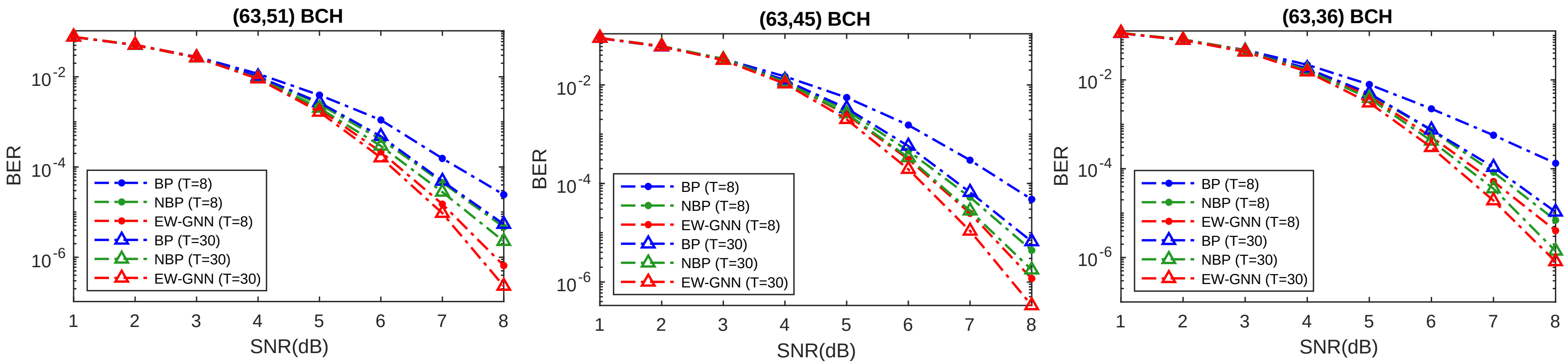


Performance of the GNN-DRL Auto-Encoder for Channel Coding

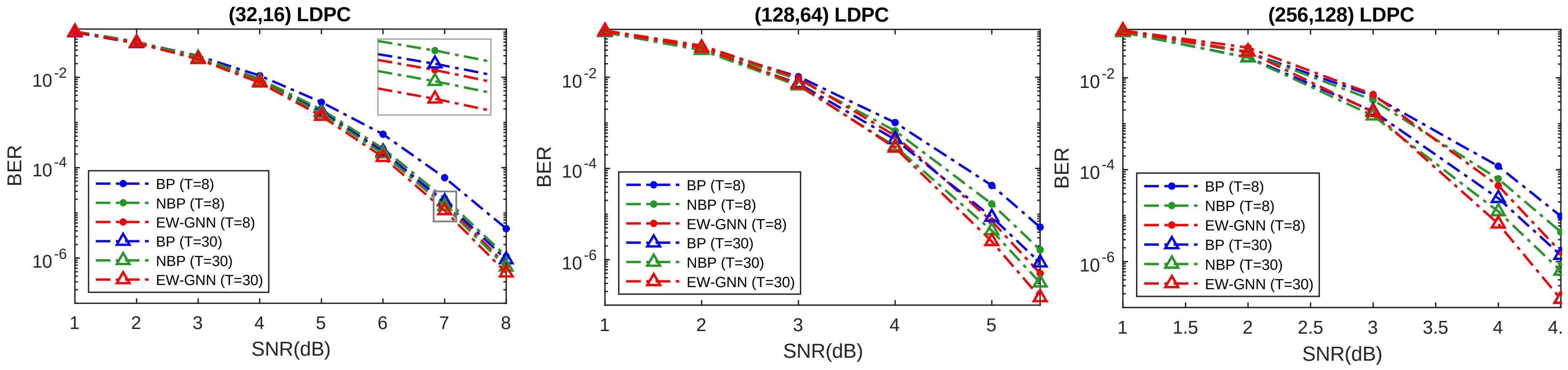
- Results of our GNN-DRL Auto-encoder【(63,45) BCH initialization】



Results



BER performance for BCH codes of length n = 63. EW-GNN is only trained with the (63, 51) BCH code and T = 8.



BER performance for LDPC codes at rate 1/2. EW-GNN is only trained with the (32, 16) LDPC code and T = 8.

Machine learning communication systems

- Google recently announced new record-breaking 72-qubit quantum processor, achieving quantum supremacy, the point at which quantum computers can perform calculations that are beyond the capabilities of even the most advanced supercomputers.
- Time for decryption can be reduced from years to minutes – new challenge for security protocols;
- Google is also developing Quantum machine learning chip, which can significantly speed up the machine learning process – complexity may not be a big issue with quantum processing
- How to develop parallel communication architecture tailored for Quantum processor? Machine learning based communication systems may be one possible approach.

Dense LEO Satellite Networks – Internet from Space

- Existing communication network design has mainly focused on terrestrial communication networks
- It is strategically important but a challenging issue to provide a full coverage of mainland and offshore economic zone.
- LEO Satellite network has a key role to play here
- Elon Musk's SpaceX is going to launch 42000 LEOs in next few years to provide Global internet communications
- Altitude of 1150 – 1275 km, latency to 7ms
- Employ optical feeder links/ inter-satellite links and phased array beam forming and digital processing technologies in the Ku- and Ka-band
- The system will be able to provide high speed (up to 1 Gbps per user, which is 200 times faster than current average internet speed), low latency broadband services for consumers and businesses
- Reliability due to rain attenuation, inter-beam-interference, real-time data services, low latency, channel estimation due to fast moving satellite
- joint design of Satellite and Terrestrial communication networks

Connecting ideas, anticipating the future:
collaborative innovation for 5G and 6G networks.

II INTERNATIONAL WORKSHOP
xGMobile

Organized by:

xGMobile Inatel FAPEMIG EMBRAPII GOVERNO DE MINAS MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO GOVERNO DO BRASIL