

Edited by
Trung Q. Duong • Saeed R. Khosravirad
Changyang She • Petar Popovski
Mehdi Bennis • Tony Q.S. Quek

Ultra-Reliable and Low-Latency Communications (URLLC) Theory and Practice

Advances in 5G and Beyond



WILEY

Ultra-Reliable and Low-Latency Communications (URLLC) Theory and Practice

Ultra-Reliable and Low-Latency Communications (URLLC) Theory and Practice

Advances in 5G and Beyond

*Edited by Trung Q. Duong, Saeed R. Khosravirad,
Changyang She, Petar Popovski, Mehdi Bennis and
Tony Q.S. Quek*

WILEY

This edition first published 2023

© 2023 John Wiley & Sons Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Trung Q. Duong, Saeed R. Khosravirad, Changyang She, Petar Popovski, Mehdi Bennis and Tony Q.S. Quek to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

A catalogue record for this book is available from the Library of Congress

Hardback ISBN: 9781119818304; ePub ISBN: 9781119818335; ePDF ISBN: 9781119818311;

oBook ISBN: 9781119818366

Cover image: © Krunja/Shutterstock

Cover design by Wiley

Set in 9.5/12.5pt STIXTwoText by Integra Software Services Pvt. Ltd, Pondicherry, India

Contents

Preface *vii*

List of Contributors *ix*

- 1 **URLLC: Faster, Higher, Stronger, and Together** *1*
*Changyang She, Trung Q. Duong, Saeed R. Khosravirad,
 Petar Popovski, Mehdi Bennis, and Tony Q.S. Quek*
- 2 **Statistical Characterization of URLLC: Frequentist and Bayesian Approaches** *15*
Tobias Kallehauge, Pablo Ramirez-Espinosa, Anders E. Kjør, and Petar Popovski
- 3 **Characterizing and Taming the Tail in URLLC** *61*
Chen-Feng Liu, Yung-Lin Hsu, Mehdi Bennis, and Hung-Yu Wei
- 4 **Unsupervised Deep Learning for Optimizing Wireless Systems with Instantaneous and Statistic Constraints** *85*
Chengjian Sun, Changyang She, and Chenyang Yang
- 5 **Channel Coding and Decoding Schemes for URLLC** *119*
Chentao Yue, Mahyar Shirvanimoghaddam, Branka Vucetic, and Yonghui Li
- 6 **Sparse Vector Coding for Ultra-reliable and Low-latency Communications** *169*
Byonghyo Shim

- 7 Network Slicing for URLLC 215**
*Peng Yang, Xing Xi, Tony Q. S. Quek, Jingxuan Chen,
Xianbin Cao, and Dapeng Wu*
- 8 Beamforming Design for Multi-user Downlink OFDMA-URLLC
Systems 241**
Walid R. Ghanem, Vahid Jamali, Yan Sun, and Robert Schober
- 9 A Full-Duplex Relay System for URLLC with Adaptive Self-Interference
Processing 259**
Hanjun Duan, Yufei Jiang, Xu Zhu, and Fu-Chun Zheng
- 10 Mobility Prediction for Reducing End-to-End Delay in URLLC 291**
Zhanwei Hou, Changyang She, Yonghui Li, and Branka Vucetic
- 11 Relay Robot-Aided URLLC in 5G Factory Automation with
Industrial IoTs 321**
*Dang Van Huynh, Saeed R. Khosravirad, Yuexing Peng,
Antonino Masaracchia, and Trung Q. Duong*
- Index 343**

Preface

Pursuing ever higher data rates has been the central design goal in all the previous generations of mobile communications. This has been changed in the 5th generation (5G) mobile communications that aims to support various new emerging services with diverse and stringent quality-of-service requirements. The most formidable challenge in 5G is to achieve Ultra-Reliable Low-Latency Communications (URLLC) for many mission-critical services including autonomous vehicles, industry automation, and tele-robotic surgery, i.e. the roundtrip delay of 1 millisecond and less than 1 out of a million in packet loss. In the 4th generation (4G) systems, the average latency is usually a few hundred milliseconds, and the packet loss probability is around 1%. 5G systems need to significantly improve the latency and reliability by several orders of magnitude compared to 4G systems. This presents unprecedented challenges.

This book covers a range of topics from fundamental theories to practical solutions in URLLC.

In Chapters 2 and 3, the authors analyze the statistical features and tail distributions of wireless channels and provide useful insights on the performance of URLLC. *Chapter 2* presents the statistical aspects of URLLC in both frequentist and Bayesian approaches. The authors have analyzed the statistical features and guarantees for outage probability in a narrowband wireless channel. *Chapter 3* considers various metrics of URLLC including tail distribution, higher-order statistics, extreme events with very low occurrence probabilities, worst-case metrics, and reliability/latency. The authors have introduced readers to the entropic risk measure in financial mathematics, generalized extreme value distribution, and generalized Pareto distribution to investigate these metrics.

In Chapters 4–7, the authors have introduced several techniques to guarantee the reliability and latency of URLLC, including machine learning, candidate channel codes, sparse vector coding, and network slicing. Two problems of resource

allocation in URLLC are addressed in *Chapter 4* with an unsupervised learning approach. The results have shown that bandwidth utilization efficiency of URLLC can be improved more significantly by exploiting frequency diversity than by multi-user diversity. *Chapter 5* discusses the channel coding and decoding schemes for URLLC. This chapter reviews state-of-the-art channel codes for URLLC and analyzes them in terms of performance and complexity. Furthermore, the Ordered Statistics Decoding (OSD) is promoted as one of the potential universal decoding algorithms for URLLC. In *Chapter 6*, a new approach to support short packet transmissions, referred as Sparse Vector Coding (SVC) is introduced. The numerical evaluations and performance analysis which validate the proposed SVC technique is highly effective in URLLC transmissions. *Chapter 7* studies a CoMP-enabled RAN slicing system simultaneously supporting URLLC and eMBB services. The authors address a joint bandwidth and CoMP beamforming optimization problem to maximize the long-term total slice utility.

In Chapters 8 and 9, downlink Orthogonal Frequency Division Multiple Access systems (OFDMA) and full-duplex relay system are optimized for URLLC, respectively. *Chapter 8* investigates the beamforming design for downlink OFDMA to enable the stringent delay requirement. In particular, the authors address a non-convex optimization problem to maximize the weighted system sum throughput subject to Quality-of-Service (QoS) of URLLC users. *Chapter 9* presents an up-to-date overview of the end-to-end latency for a Full-Duplex (FD) relay system in the context of URLLC. The authors not only provide an insightful investigation of reliability and latency together for FD relay assisted URLLC but also discuss possible relaying latency reduction solutions in the chapter.

In Chapters 10 and 11, the authors investigate URLLC in vertical industries: Tactile Internet and Industrial Internet-of-Things. More specifically, *Chapter 10* addresses an optimization problem that maximizes the number of URLLC services by jointly optimizing time and frequency resources and the prediction horizon. The numerical results clearly demonstrate the effectiveness of the proposed solution. In addition, a proof-of-concept experiment with the remote control in a virtual factory is also provided to illustrate a typical application of Tactile Internet. Finally, *Chapter 11* considers relay robots-aided URLLC in 5G factory automation, which consists of multiple relay robot deployment and decoding error probability minimization problems. There are two different approaches introduced for relay robot deployment, including Deep Neural Networks (DNN) and the K-means clustering algorithm. A low-complexity iterative algorithm is also provided to deal with the joint blocklength and power allocation problem to minimize the decoding error probability.

List of Contributors

Mehdi Bennis

Centre for Wireless Communications
University of Oulu
Oulu, Finland

Xianbin Cao

School of Electronics and
Information Engineering
Beihang University
Beijing, China

Jingxuan Chen

School of Electronics and
Information Engineering
Beihang University
Beijing, China

Hanjun Duan

The School of Electronic and
Information Engineering
Harbin Institute of Technology
Shenzhen, China

Trung Q. Duong

School of Electronics Electrical
Engineering and Computer Science

Queen's University
Belfast, UK

Walid R. Ghanem

Friedrich-Alexander-University
Erlangen-Nuremberg (FAU)
Erlangen, Germany

Zhanwei Hou

The School of Electrical and
Information Engineering
The University of Sydney
Sydney, Australia

Yung-Lin Hsu

Graduate Institute of
Communication Engineering
National Taiwan University
Taipei, Taiwan

Dang Van Huynh

School of Electronics Electrical
Engineering and Computer Science
Queen's University
Belfast, UK

Vahid Jamali

Technical University Darmstadt
Darmstadt, Germany

Yufei Jiang

The School of Electronic and
Information Engineering
Harbin Institute of Technology
Shenzhen, China

Tobias Kallehauge

Connectivity Section at the
Department of Electronic Systems
Aalborg University
Aalborg, Denmark

Anders E. Kalør

Connectivity Section at the
Department of Electronic Systems
Aalborg University
Aalborg, Denmark

Saeed R. Khosravirad

Nokia Bell Labs
Murray Hill
New Jersey, US

Yonghui Li

School of Electrical and Information
Engineering
University of Sydney
Sydney, Australia

Chen-Feng Liu

Technology Innovation Institute
Masdar City
Abu Dhabi, UAE

Antonino Masaracchia

School of Electronics Electrical
Engineering and Computer Science

Queen's University
Belfast, UK

Yuexing Peng

Beijing University of Posts and
Telecommunications
Beijing, China

Petar Popovski

Connectivity Section at the
Department of Electronic Systems
Aalborg University
Aalborg, Denmark

Tony Q.S. Quek

Information Systems Technology
and Design
Singapore University of Technology
and Design
Singapore

Pablo Ramirez-Espinosa

Connectivity Section at the
Department of Electronic Systems
Aalborg University
Aalborg, Denmark

Robert Schober

Friedrich-Alexander-University
Erlangen-Nuremberg (FAU)
Erlangen, Germany

Changyang She

The School of Electrical and
Information Engineering
The University of Sydney
Sydney, Australia

Byonghyo Shim

Institute of New Media and
Communications and Department of

Electrical and Computer Engineering
Seoul National University
Seoul, Korea

Mahyar Shirvanimoghaddam

The School of Electrical and
Information Engineering
The University of Sydney
Sydney, Australia

Chengjian Sun

School of Electronics and
Information Engineering
Beihang University
Beijing, China

Branka Vucetic

The School of Electrical and
Information Engineering
The University of Sydney
Sydney, Australia

Hung-Yu Wei

Department of Electrical
Engineering
National Taiwan University
Taipei, Taiwan

Dapeng Wu

Department of Electrical and
Computer Engineering
University of Florida
Gainesville, USA

Xing Xi

School of Electronics and
Information Engineering
Beihang University
Beijing, China

Chenyang Yang

School of Electronics and
Information Engineering
Beihang University
Beijing, China

Peng Yang

Information Systems Technology
and Design
Singapore University of Technology
and Design
Singapore

Chentao Yue

The School of Electrical and
Information Engineering
The University of Sydney
Sydney, Australia

Fu-Chun Zheng

The School of Electronic and
Information Engineering
Harbin Institute of Technology
Shenzhen, China

Xu Zhu

The School of Electronic and
Information Engineering
Harbin Institute of Technology
Shenzhen, China

1

URLLC: Faster, Higher, Stronger, and Together

Changyang She^{1,*}, Trung Q. Duong², Saeed R. Khosravirad³, Petar Popovski⁴, Mehdi Bennis⁵, and Tony Q.S. Quek⁶

¹ School of Electrical and Information Engineering, University of Sydney, 2006, NSW, Australia

² School of Electronics Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN, Belfast, UK

³ Nokia Bell Labs, NJ 07974-0636, Murray Hill, USA

⁴ Connectivity Section at the Department of Electronic Systems, Aalborg University, 9220, Aalborg, Fredrik Bajers Vej 7A, Denmark

⁵ Centre for Wireless Communications, University of Oulu, Oulu, Finland

⁶ Information Systems Technology and Design, Singapore University of Technology and Design, 487372, Singapore

* Corresponding Author

As one of the new communication scenarios in the 5th Generation (5G) mobile communications, Ultra-Reliable and Low-Latency Communications (URLLC) are crucial for enabling a wide range of emerging applications, including industry automation, intelligent transportation, telemedicine, Tactile Internet, and Virtual/Augmented Reality (VR/AR). According to the requirements in 5G standards, to support emerging mission-critical applications, the End-to-End (E2E) delay cannot exceed 1 ms and the packet loss probability should be 10^{-5} – 10^{-7} . Compared with the existing cellular networks, the delay and reliability require significant improvements by at least two orders of magnitude for 5G networks. This capability gap cannot be fully resolved by the 5G New Radio (NR), i.e. the physical-layer technology for 5G, even though the transmission delay in Radio Access Networks (RANs) achieves the 1 ms target. Transmission delay contributes only a small fraction of the E2E delay, as the stochastic delays in upper networking layers, such as queuing delay, processing delay, and access delay, are key bottlenecks for achieving URLLC. Beyond 5G systems or so-called 6th Generation (6G) systems should guarantee the E2E delay bound with high reliability.

In addition to the latency and reliability requirements, some other Key Performance Indicators (KPIs) should also be taken into account, including Spectrum

Efficiency (SE), throughput, Energy Efficiency (EE), Age of Information (AoI), jitter of latency, round-trip delay, network availability, and security (shown in Table 1.1). These requirements will pose unprecedented challenges in terms of design methodologies and enabling technologies in the 6th Generation (6G) mobile communications. To fill the gap between 5G URLLC and the diverse KPI requirements, we shall investigate novel methodologies and innovative technologies for the next generation URLLC (xURLLC), also known as eXtreme URLLC, [23]. This book will cover various methods and technologies to achieve URLLC from the physical layer, link layer, and network layer, to diverse applications in vertical industries of 5G/6G communications.

Table 1.1 KPIs and research challenges. ©IEEE 2021. Reprinted with permission from [28].

Indoor large-scale scenarios		
Applications	KPIs	Research Challenges
Factory automation	SE, EE, and AoI	Scalability and network congestions
VR/AR applications	SE and throughput	Processing/transmission 3D videos
Indoor wide-area scenarios		
Applications	KPIs	Research Challenges
Tele-surgery	Round-trip delay, throughput, and jitter	Propagation delay and high data rate
eHealth monitoring	EE and network availability	Propagation delay and localization
Outdoor large-scale scenarios		
Applications	KPIs	Research Challenges
Vehicle safety	AoI, SE, security, and network availability	High mobility and scalability
Outdoor wide-area scenarios		
Applications	KPIs	Research Challenges
Smart grid	SE	Propagation delay and scalability
Tele-robotic control	SE, security, network availability, and jitter	Propagation delay and high data rate
UAV control	EE, security, network availability, and AoI	Propagation delay and high mobility

1.1 Requirements of URLLC: Faster, Higher, Stronger, and Together

The next generation URLLC is expected to be “faster, higher, stronger - together”. The specific requirements and research challenges are discussed in the sequel.

1.1.1 Faster Responses and Movement

In factory automation and autonomous vehicles, mobile devices need to make decisions according to their local observation in a real-time manner. Given the fact that the energy budget and the computing capability of each device are limited, it may need the help of Mobile Edge Computing system (MEC). Unlike centralized mobile cloud computing with routing delay and propagation delay in backhubs and core networks, the E2E delay in MEC systems consists of Uplink (UL) and Downlink (DL) transmission delays, queuing delays in the buffers of users and Base Stations (BSs), and the processing delay in the MEC [26]. Although MEC helps reduce latency in communication systems, there are two bottlenecks for providing fast responses to mobile devices. First, optimization problems in MEC systems are generally non-convex. To find the optimal solution, such as user association and task offloading. The computing complexity for executing searching algorithms is too high to be implemented in real-time. Second, exchanging information among different edge servers will lead to high overheads and latency. To avoid this issue, edge servers need to make decisions in a distributed manner. As a result, a local decision may not be optimal for all the devices.

Supporting high mobility URLLC is critical for some outdoor applications, e.g. Unmanned Aerial Vehicle (UAV) control. Since the mobile devices are moving fast, the Doppler frequency shift is large. Thus, the inter-symbol interference is strong and the receiver needs to adjust the carrier frequency according to the Doppler shift. Besides, frequent handovers in high mobility URLLC will result in service interruption. The BS with good channel quality may not have sufficient radio resources due to the dynamic traffic load in high mobility scenarios. Therefore, how to serve high mobility URLLC remains an open issue in 6G.

1.1.2 Higher Throughput and Density

As one of the killer applications in 5G networks, VR/AR applications require ultra-reliable and low-latency tactile feedback and high data rate 360° videos [8]. Meanwhile, as the sizes of devices shrinks, battery lifetime will become a bottleneck for enabling high data rate URLLC [24]. To implement VR/AR applications in future wireless networks, we need to investigate the fundamental trade-offs

among throughput, energy efficiency, reliability, and latency in communications, caching, and computing systems [34], as well as enabling technologies such as touch user interface and haptic codecs [4, 31].

Due to the explosive growth of the numbers of autonomous vehicles and mission-critical IoT devices [9], future wireless networks are expected to support massive URLLC. To support massive URLLC, novel communication and learning techniques are needed. With orthogonal multiple access technologies, the required bandwidth increases linearly with the number of devices. To achieve better trade-offs among delay, reliability, and scalability, other multiple access technologies should be used, such as non-orthogonal multiple access and contention-based multiple access technologies [29, 30]. Meanwhile, we may need to exploit the above 6 GHz spectrum including mmWave [16] and the Terahertz band [36].

1.1.3 Stronger Connectivity and Security

Multi-connectivity is a promising approach to provide seamless services to URLLC users [17]. As illustrated in [22], one way to improve network availability without sacrificing spectrum efficiency is to serve each user with multiple BSs over the same subchannel (or subcarrier). The disadvantage of this intra-frequency multi-connectivity is that the failures of different links are highly correlated. For example, if there is a strong interference on a subchannel, then the signal to interference plus noise ratios of all the links are low. To alleviate cross-correlation among different links, different nodes can connect to one user with different subchannels or even with different communication interfaces [21]. Further considering that terrestrial networks may not cover rural areas and marine areas, we need to use non-terrestrial networks to provide global connectivity for long-distance URLLC services, e.g. Tactile Internet.

Future URLLC systems will suffer from different kinds of attacks that result in inefficient communications [19]. The widely used cryptography algorithms require high-complexity signal processing, and may not be suitable for URLLC, especially for IoT devices with low computing capacities. To defend against eavesdropping attacks in URLLC, physical layer security is a viable solution [3]. The maximal secret communication rate in the short blocklength regime over a wiretap channel was derived in [37]. The results show that there are trade-offs among delay, reliability, and security. Based on this fundamental result, we can further investigate the technologies for improving physical-layer security.

1.1.4 Human Intelligence Together with Artificial Intelligence in URLLC

As illustrated in Figure 1.1, a new trend of developing communication networks is to integrating human intelligence (expert knowledge in wireless

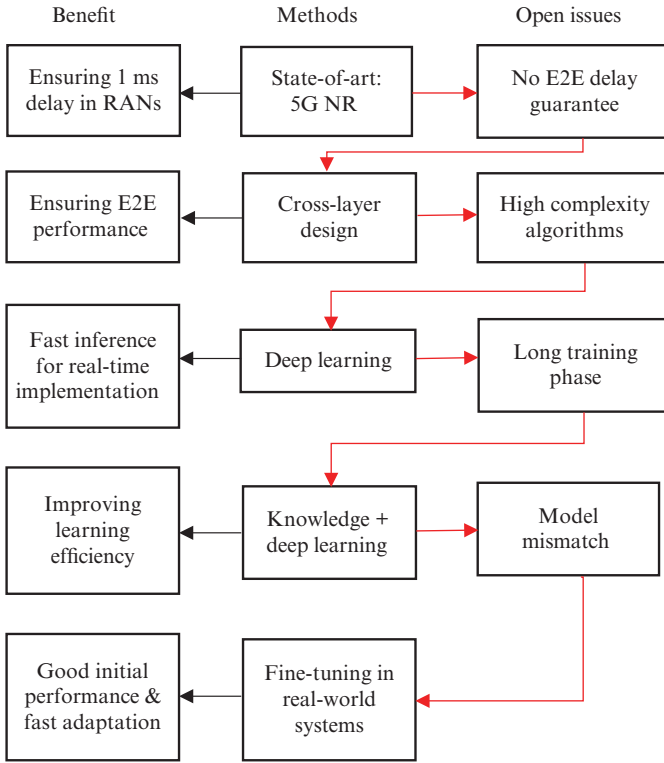


Figure 1.1 Wireless AI for developing URLLC systems. ©IEEE 2021. Reprinted with permission from [28].

communications) into artificial intelligence (deep learning) for optimizing communication systems.

Cross-layer Design Existing design methods divide communication networks into multiple layers according to the Open Systems Interconnection model [18]. Communication technologies in each layer are often developed without considering the impacts on other layers, despite the fact that the interactions across different layers are known to significantly impact on the E2E delay and reliability. Most existing approaches do not reflect such interactions; this leads to suboptimal solutions and thus we are yet to be able to meet the stringent requirements of URLLC. To guarantee the E2E delay and the reliability of the communication system, we need accurate and analytically tractable cross-layer models to reflect the interactions across different layers.

Deep Learning With 5G NR, the radio resources are allocated in each Transmission Time Interval (TTI) with a duration of $0.125 \sim 1$ ms [1]. To implement

optimization algorithms in 5G systems, the processing delay should be less than the duration of one TTI. Since the crosslayer models are complex, related optimization problems are non-convex in general. Most of the existing optimization algorithms incur high computing overheads, and hence can hardly be implemented in real-world systems. Deep learning has significant potential to address the above issue in beyond 5G/6G networks. The basic idea is to approximate the optimal policy with a Deep Neural Network (DNN). After the training phase, a near-optimal solution of an optimization problem can be obtained from the output of the DNN in each TTI. Essentially, by using deep learning, we are trading off the online processing time with the computing resource for off-line training.

Integrating Knowledge into Learning Algorithms Although deep learning algorithms have shown significant potential, the application of deep learning in URLLC is not straightforward. As shown in [27], deep learning algorithms converge slowly in the training phase and need a large number of training samples to evaluate or improve the E2E delay and reliability. If some knowledge of the environment is available, such as the estimated packet loss probability of a certain decision, the system can exploit this knowledge to improve the learning efficiency [13]. Domain knowledge of communications and networking including models, analytical tools, and optimization frameworks have been extensively studied in the existing literature [12, 14]. How to exploit them to improve deep learning algorithms for URLLC has drawn significant attention as well, including in [15, 32, 38].

Fine-tuning in Real-world Systems Communication environments in wireless networks are non-stationary in general. Theoretical models used in off-line training may not match this non-stationary nature of practical networks. As a result, a DNN trained off-line cannot guarantee the Quality-of-Service (QoS) constraints of URLLC. Such an issue is referred to as the model mismatch problem in [5]. To handle the model mismatch, wireless networks should be intelligent to adjust themselves in dynamic environments, explore unknown optimal policies, and transfer knowledge to practical networks.

1.2 Scope of This Book

We invited leading researchers in both academia and industry from diverse backgrounds to share their recent studies in URLLC.

In Chapter 2, T. Kallehauge et al. focus on the physical layer and present the statistical aspects of URLLC, detailing both frequentist and Bayesian approaches. Specifically, the authors analyze the statistical features and guarantees for outage probability in a narrowband wireless channel. As a motivating example, they treat the practical case in which a Base Station (BS) collects channel statistics for users

at different locations and attempts to predict the performance of a user at a new location. Their results show that the BS can obtain high-quality predictions of the reliability performance even for locations that are not in proximity.

In Chapter 3, instead of analyzing the average latency and the delay outage probability, C.-F. Liu et al. investigate the statistical information and/or metrics, rooted in the tail behavior of probability distributions to gain insights in URLLC systems. Specifically, the authors analyzed the tail distribution of the delay, channel fading, or packet inter-arrival time; variance and higher-order statistics; threshold deviation with a very low occurrence probability; worst-case metrics; age of information. Useful methodologies and extensive numerical results are provided in this chapter.

In Chapter 4, C. Sun et al. establish a unified framework of using unsupervised deep learning to solve both kinds of problem with both instantaneous and statistic constraints. For a constrained variable optimization, the authors first convert it into an equivalent functional optimization problem with instantaneous constraints. Then, to ensure the instantaneous constraints in the functional optimization problems, the authors use DNN to approximate the Lagrange multiplier functions, which is trained together with a DNN to approximate the policy. By taking resource allocation problems in URLLC as examples, the authors show that unsupervised learning outperforms supervised learning in terms of quality-of-service violation probability and approximation accuracy of the optimal policy.

In Chapter 5, C. Yue et al. overview candidate channel codes for URLLC, and compare them in terms of performance and complexity. Their respective strengths and weaknesses are investigated in terms of the performance gap to theoretical limits and the computational complexity of practical decoding algorithms. Furthermore, Ordered Statistics Decoding (OSD) is introduced as one of the potential universal decoding algorithms for URLLC, which can achieve near-optimal performance for any block code. The error performance and computational complexity of OSD are investigated in this chapter. Finally, recent improvements on OSD, including decoding rules and complete decoder design, are studied.

In Chapter 6, B. Shim et al. introduce a new type of short packet transmission framework named the Sparse Vector Coding (SVC) technique. The key idea behind SVC is to transform an information vector into the sparse vector in the transmitter and to exploit the sparse recovery algorithm in the receiver. Metaphorically, SVC can be thought as marking dots on the empty table. As long as the number of dots is small enough and the measurements contains enough information to figure out the marked cell positions, accurate decoding of the SVC packet can be guaranteed. The numerical results demonstrate that SVC is very effective in the short packet transmission for URLLC scenarios.

In Chapter 7, P. Yang et al. consider a CoMP-enabled RAN slicing system simultaneously supporting URLLC and eMBB traffic transmission. In the presence of eMBB traffic, the authors orchestrate the shared network resources of the system to guarantee a more reliable bursty URLLC service provision from the perspectives of lowering both URLLC packet blocking probability and codeword decoding error probability. The authors formulate the problem of RAN slicing for bursty URLLC and eMBB service multiplexing as a resource optimization problem and develop a joint bandwidth and CoMP beamforming optimization algorithm to maximize the long-term total slice utility. Several bandwidth allocation and beamforming algorithms are evaluated in the RAN slicing system.

In Chapter 8, W. R. Ghanem et al. investigate the beamforming design for downlink Orthogonal Frequency Division Multiple Access (OFDMA) URLLC systems. To enable the stringent URLLC delay requirements, finite blocklength transmission is adopted for the beamforming algorithm design. The authors formulate the beamforming algorithm design as a non-convex optimization problem for maximization of the weighted system sum throughput subject to constraints on the Quality of Service (QoS) of the URLLC users. A sub-optimal algorithm is proposed based on Sequential Convex Approximation (SCA). Numerical results reveal that the proposed design can achieve a considerable gain compared to several baseline schemes.

In Chapter 9, H. Duan et al. first present an up-to-date overview of the end-to-end latency for a Full-Duplex (FD) relay system. The authors investigate the possible solutions in the literature to achieve the goal of URLLC. The efficient solution is to allow a simple Amplify-and-Forward (AF) FD relay mode with low-complexity SI radio frequency and analog cancellations, and process the residual SI alongside the desired signal at base station in an adaptive manner, rather than being canceled at relay in digital domain. Their results show that the FD relay assisted system with adaptive SI utilization or cancellation enables extended network coverage, enhanced reliability, and reduced latency, compared to the existing overview work.

In Chapter 10, Z. Hou et al. aim to reduce the user experienced delay through prediction and communication co-design, where each mobile device predicts its future states and sends them to a data center in advance. Since predictions are not error-free, the authors consider prediction errors and packet losses in communications when evaluating the reliability of the system. Then, the authors formulate an optimization problem that maximizes the number of URLLC services supported by the system by optimizing time and frequency resources and the prediction horizon. Simulation and experiment results verify the effectiveness of the proposed method, and show that the trade-off between user experienced delay and reliability can be improved significantly via prediction and communication co-design.

In Chapter 11, D. V. Huynh et al. investigate the URLLC supported Industrial Internet-of-Things (IIoT) devices in industry automation. To enhance the URLLC performance, the authors propose two approaches to optimize the deployment of multiple relay robots in assisting URLLC system, namely DNN-based deployment, and the K-means clustering algorithm. An optimal resource allocation scheme is proposed to minimize the error probability at the IIoT devices. To solve the highly non-convex optimization problems of URLLC, the authors propose an effective iterative algorithm for solving the reliability maximization. Representative numerical results demonstrate the proposed scheme can significantly improve the reliability over various conventional approaches.

1.3 Future Directions

Considering the diverse application scenarios and KPIs of URLLC in 6G, the design of future communication systems requires considerable additional research efforts beyond what the community has done so far. We discuss some promising research directions in this section.

1.3.1 Constrained Deep Learning for URLLC

When applying deep learning for URLLC applications, the reward is usually defined as a weighted sum of different KPIs. With different weighting coefficients, the final achieved KPIs are different. Thus, we need to select these weighting coefficients manually to achieve satisfactory KPIs. A potential way to overcome this difficulty is to formulate the problem as a constrained optimization problem and use constrained unsupervised deep learning to find the optimal policy [7]. If the problem turns out to be a sequential-decision making problem with constraints, constrained deep reinforcement learning algorithms can be applied [20]. Nevertheless, the reliability of URLLC is extremely high and the closed-form results may not be available. Therefore, the required number of training samples is extremely large. How to achieve the target KPIs with a reasonable amount of training samples remains an open issue.

1.3.2 Distributed Learning for URLLC

In IIoT, collecting the status of all the devices in the central server will bring considerable overheads to wireless networks, and hence the devices should be controlled in a distributed manner. To achieve this goal, distributed learning with partial observation is a promising framework. With this framework, edge computing servers, VR glasses, and IoT devices can take actions based

on local observations. In this way, the overheads for exchanging control information and updating global status can be reduced remarkably [25]. The major issues of distributed learning include long convergence time, poor performance with partial observation, and limited computing and storage resources of mobile devices.

1.3.3 Graph Neural Networks for Network Management of URLLC

Since the dimensions of the input and the output of a Fully-connected Neural Network (FNN) grows with the number of devices, we need to adjust the hyper-parameters of the FNN and retrain the parameters whenever the number of devices varies. Thus, FNNs are not flexible in managing dynamic networks. To address this issue, a promising approach is to use GNNs to represent the topology of wireless networks [11]. As indicated in [35], GNN is a very general structure that can be applied to solve large-scale problems with non-Euclid data structure. Since the number of parameters of a GNN does not increase with the dimension of the input, GNNs are suitable for resource management in dynamic networks [6]. In most of the existing literature, the network status is assumed to be perfectly known at the central control plane. How to find optimal policy with inaccurate/outdated/partial network status remain open problems.

1.3.4 Few-shot Learning for URLLC

The real-world data samples from practical systems are limited and could be non-stationary in wireless networks. A promising learning framework for fast adaptation is known as few-shot learning [33]. The basic idea is to use meta learning to optimize hyper-parameters, including initial parameters, learning rates, and the structures of neural networks [2, 10]. After off-line training in existing tasks, the neural networks can be transferred to new tasks with limited data samples. Most of the existing few-shot learning algorithms for image classification cannot achieve high reliability. Whether it is possible to meet the reliability requirement of URLLC with few-shot learning remains open.

Bibliography

- [1] 3GPP. Study on new radio (NR) access technology; Physical layer aspects (release 14). Stockholm, Sweden, Jun.2017. TR 38.802 V2.0.0.
- [2] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning

- to learn by gradient descent by gradient descent. In *Proc. Neural information processing systems (NIPS)*, pages 3981–3989, 2016.
- [3] Riqing Chen, Chunhui Li, Shihao Yan, Robert Malaney, and Jinhong Yuan. Physical layer security for ultra-reliable and low-latency communications. *IEEE Wireless Commun.*, 26 (5): 6–11, 2019.
 - [4] Victor Adriel de Jesus Oliveira, Luciana Nedel, and Anderson Maciel. Assessment of an articulatory interface for tactile intercommunication in immersive virtual environments. *Computers & Graphics*, 76: 18–28, 2018.
 - [5] Rui Dong, Changyang She, Wibowo Hardjawana, Yonghui Li, and Branka Vucetic. Deep learning for radio resource allocation with diverse quality-of-service requirements in 5G. *IEEE Trans. Wireless Commun.*, 20 (4): 2309–2324, 2020.
 - [6] Mark Eisen and Alejandro Ribeiro. Optimal wireless resource allocation with random edge graph neural networks. *IEEE Trans. Signal Process.*, 68: 2977–2991, 2020.
 - [7] Mark Eisen, Clark Zhang, Luiz FO Chamon, Daniel D Lee, and Alejandro Ribeiro. Learning optimal resource allocations in wireless systems. *IEEE Trans. Signal Process.*, 67 (10): 2775–2790, 2019.
 - [8] Mohammed S Elbamy, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. Toward low-latency and ultra-reliable virtual reality. *IEEE Netw.*, 32 (2): 78–84, 2018.
 - [9] Ericsson. Internet of things forecast. *Ericsson Mobility Report*, 2019. URL <https://www.ericsson.com/en/mobility-report/internet-of-things-forecast>.
 - [10] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. International Conference on Machine Learning (ICML)*, pages 1126–1135, 2017.
 - [11] Fernando Gama, Joan Bruna, and Alejandro Ribeiro. Stability properties of graph neural networks. *IEEE Trans. Signal Process.*, 68: 5680–5695, 2020.
 - [12] Andrea Goldsmith. *Wireless Communications*. Cambridge university press, 2005.
 - [13] Shixiang Gu, Timothy Lillicrap, Ilya Sutskever, and Sergey Levine. Continuous deep Q-learning with model-based acceleration. In *Proc. International conference on machine learning (ICML)*, pages 2829–2838. PMLR, 2016.
 - [14] Mor Harchol-Balter. *Performance modeling and design of computer systems: Queueing Theory in Action*. Cambridge University Press, 2013.
 - [15] Hengtao He, Shi Jin, Chao-Kai Wen, Feifei Gao, Geoffrey Ye Li, and Zongben Xu. Model-driven deep learning for physical layer communications. *IEEE Wireless Commun.*, 26 (5): 77–83, 2019.
 - [16] Ibrahim A Hemadeh, Katla Satyanarayana, Mohammed El-Hajjar, and Lajos Hanzo. Millimeter-wave communications: Physical channel models, design

- considerations, antenna constructions, and link-budget. *IEEE Commun. Surveys & Tuts.*, 20 (2): 870–913, 2017.
- [17] Tom Hobler, Lucas Scheuven, Norman Franchi, Meryem Simsek, and Gerhard P. Fettweis. Applying reliability theory for future wireless communication networks. In *Proc. IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2017.
 - [18] Xiaolin Jiang, Hossein Shokri-Ghadikolaie, Gabor Fodor, Eytan Modiano, Zhibo Pang, Michele Zorzi, and Carlo Fischione. Low-latency networking: Where latency lurks and how to tame it. *Proc. IEEE*, 107 (2): 280–306, 2018.
 - [19] Chong Li, Chih-Ping Li, Kianoush Hosseini, Soo Bum Lee, Jing Jiang, Wanshi Chen, Gavin Horn, Tingfang Ji, John E Smee, and Junyi Li. 5G-based systems design for tactile internet. *Proc. IEEE*, 107 (2): 307–324, 2018.
 - [20] Qingkai Liang, Fanyu Que, and Eytan Modiano. Accelerated primal-dual policy optimization for safe reinforcement learning. *Proc. Neural Information Processing Systems (NIPS)*, 2019.
 - [21] Jimmy J. Nielsen, Rongkuan Liu, and Petar Popovski. Ultra-reliable low latency communication (URLLC) using interface diversity. *IEEE Trans. Commun.*, 66 (3): 1322–1334, Mar. 2018.
 - [22] David Ohmann, Ahmad Awada, Ingo Viering, Meryem Simsek, and Gerhard P. Fettweis. Modeling and analysis of intra-frequency multi-connectivity for high availability in 5G. In *Proc. IEEE Vehicular Technology Conference (VTC) Spring*, 2018.
 - [23] Jihong Park, Sumudu Samarakoon, Hamid Shiri, Mohamed K Abdel-Aziz, Takayuki Nishio, Anis Elgabli, and Mehdi Bennis. Extreme URLLC: Vision, challenges, and key enablers. *arXiv preprint arXiv:2001.09683*, 2020.
 - [24] Walid Saad, Mehdi Bennis, and Mingzhe Chen. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Netw.*, 34 (3): 134–142, 2019.
 - [25] Mohit K Sharma, Alessio Zappone, Mohamad Assaad, Mérouane Debbah, and Spyridon Vassilaras. Distributed power control for large energy harvesting networks: A multi-agent deep reinforcement learning approach. *IEEE Trans. Cog. Commun. and Netw.*, 5 (4): 1140–1154, 2019.
 - [26] Changyang She, Yifan Duan, Guodong Zhao, Tony Q. S. Quek, Yonghui Li, and Branka Vucetic. Cross-layer design for mission-critical IoT in mobile edge computing systems. *IEEE Internet of Things J.*, 6 (6): 9360–9374, 2019.
 - [27] Changyang She, Rui Dong, Zhouyou Gu, Zhanwei Hou, Yonghui Li, Wibowo Hardjawana, Chenyang Yang, Lingyang Song, and Branka Vucetic. Deep learning for ultra-reliable and low-latency communications in 6G networks. *IEEE Netw.*, 34 (5): 219–225, 2020.
 - [28] Changyang She, Chengjian Sun, Zhouyou Gu, Yonghui Li, Chenyang Yang, H Vincent Poor, and Branka Vucetic. A tutorial on ultra-reliable and

- low-latency communications in 6G: Integrating domain knowledge into deep learning. *Proc. IEEE*, 109 (3): 204–246, Mar. 2021.
- [29] Mahyar Shirvanimoghaddam, Mischa Dohler, and Sarah J Johnson. Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations. *IEEE Commun. Mag.*, 55 (9): 55–61, 2017.
 - [30] Bikramjit Singh, Olav Tirkkonen, Zexian Li, and Mikko A Uusitalo. Contention-based access for ultra-reliable low latency uplink transmissions. *IEEE Wireless Commun. Lett.*, 7 (2): 182–185, 2017.
 - [31] Eckehard Steinbach, Matti Strese, Mohamad Eid, Xun Liu, Amit Bhardwaj, Qian Liu, Mohammad Al-Ja’afreh, Toktam Mahmoodi, Rania Hassen, Abdulmotaleb El Saddik, et al. Haptic codecs for the tactile internet. *Proc. IEEE*, 107 (2): 447–470, 2018.
 - [32] Chengjian Sun and Chenyang Yang. Learning to optimize with unsupervised learning: Training deep neural networks for URLLC. In *Proc. IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2019.
 - [33] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 403–412, 2019.
 - [34] Yaping Sun, Zhiyong Chen, Meixia Tao, and Hui Liu. Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff. *IEEE Trans. Commun.*, 67 (11): 7573–7586, 2019.
 - [35] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. learning Syst.*, 32 (1): 4–24, 2020.
 - [36] Yunchou Xing and Theodore S Rappaport. Propagation measurement system and approach at 140 GHz-moving to 6G and above 100 GHz. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2018.
 - [37] Wei Yang, Rafael F Schaefer, and H Vincent Poor. Wiretap channels: Nonasymptotic fundamental limits. *IEEE Trans. Inf. Theory*, 65 (7): 4069–4093, 2019.
 - [38] Alessio Zappone, Marco Di Renzo, and M  rouane Debbah. Wireless networks design in the era of deep learning: Model-based, AI-based, or both? *IEEE Trans. Commun.*, 67 (10): 7331–7376, 2019.

Higher-order Statistics 7, 61–63, 73
 Hyperparameters 32, 37, 44–50
 Hybrid Automatic Repeat Request (HARQ) 294

i

Immediate resource
 orchestration 216
 incomplete gamma function, 27
 independent and identically distributed (i.i.d.) 64, 66, 134
 Index Modulation (IM) 170
 Information Set Reduction (IISR) 141
 Information of Mode Selection (IMS) 270
 Industrial Internet-of-Things (IIoT) 321
 Instantaneous normalized capacity 22
 Integrating Knowledge into Learning Algorithms 6
 Internet of Things (IoT) 64, 119
 Internet of Vehicles (IoV) 259
 Internet Protocol (IP) 218
 International Telecommunication Union (ITU) 215

k

K-means Clustering 328, 329, 331
 Karush-Kuhn-Tucker (KKT) 90
 Kullback–Leibler (KL) 69
 Key Performance Indicators (KPIs) 1, 2

l

large-scale fading 19, 313
 Lambert function 301, 302
 Light-tailed 67, 68
 line-of-sight (LoS) 19, 264, 268, 279
 linear time-invariant(LTI) 18
 Low Density Parity Check (LDPC) 119
 lognormal 19
 Long-term evolution (LTE) 122

m

maximum ratio
 transmission 14, 231
 maximum a posteriori probability (MAP) 126
 Markov Chain Monte Carlo (MCMC) 33
 Machine Learning 221
 Maximum-Likelihood Decoding (MLD) 138
 medium access control (MAC) 247
 Minislot 7, 219
 minimum mean-square-error (MMSE) 43, 203
 mmWave 4, 190
 mobile-edge computing (MEC) 3, 295
 mobile robots 16, 323
 Mobility Prediction 291
 Modulation and Coding Scheme (MCS) 169
 multipath fading 241
 multiple-input-multiple-output 278

n

narrowband 16, 18, 39–40
 Network Slicing 215, 217
 Network Slice Management (NSM) 217
 Next Generation Mobile Network Alliance (NGMA) 215
 normal equations 45
 non-convex optimization problems 8, 247
 Non-Orthogonal Multiple Access (NOMA) 322
 Non-line-of-sight (NLOS) 19, 269

o

Open Base Station Architecture Initiative (OBSAI) 218
 Outage Probability 20–27, 35, 37, 50, 53, 55, 61–63, 293
 Ordered Statistics Decoding (OSD) 7, 131
 Orthogonal Frequency Division Multiplexing (OFDM) 170
 Orthogonal Frequency Division Multiple Access (OFDMA) 241

p

Packet Error Rate (PER) 199
 Packet Data Convergence Protocol (PDCP) 218
 Pickands–Balkema–de Haan theorem 67
 power allocation 88, 95, 101–109, 267, 322, 333, 336
 Position Modulation (PM) 170
 posterior predictive 33, 50
 predictive mean 43

prior information 44, 56
 probability density function (PDF) 19, 68, 90, 132, 180
 Probably Approximately Correct (PAC) 26
 Probably Correct Reliability (PCR) 26–31
 Probabilistic Sufficient Condition (PSC) 151
 physical layer security (PLS) 4
 Physical Downlink Control Channel (PDCCH) 171

q

Quality-of-Service (QoS) 6, 85–87, 91, 95, 99, 102, 104, 106–108, 110, 218–222, 235, 241, 246, 257, 322

r

radio access network 216
 Radio map 38–41
 Radio Network Temporary Identifier (RNTI) 171
 Radio Resource Manage and Control (RRMC) 217
 Radio Resource Control Function (RRCF) 218
 random variable 19, 20, 24, 217
 Rayleigh distribution 4, 6, 260
 Rayleigh fading 19, 24–26, 33, 36, 199
 received signal strength (RSS) 39
 Reduced Round Trip time (RTT) 218
 reinforcement learning 9
 Relay Robot Deployment 328

Remote Radio Heads

(RRHs) 219

Resource efficiency 216

Rician distribution 49, 52

Rician fading 20, 47

S

Sample Average Approximate

(SAA) 227

Shadowing 47–49, 55, 96, 234,

310, 322, 323, 325

Second-Order Reed–Muller

(SORM) 190

SegmentationDiscarding

Decoding (SDD) 161

Self-Interference

Cancellation 264

Semi-Definite Programming

(SDP) 248

Sequential Convex approximation

(SCA) 253

sequential-decision making

problem 9

signal-to-interference-plus-noise

ratio (SINR) 196, 245, 249

Single-Input and MultipleOutput

(SIMO) 198

signal-to-noise ratio (SNR)

199–200

shadowing 47, 48, 55, 96, 234

Short Block-length 119

Short-sized Packet 169, 171

Short-tailed 67

shuttle carriers 16

Slater's condition 93

small-scale fading 19, 39, 47–49,

55

Soft Discarding Rule (SDR) 151

software-defined networking

(SDN) 216

Software-Defined RAN

Coordinator (SDRAN-C) 217

spatial multiplexing 246

spatial correlation 38

Sparse Vector Coding 7, 169,

170, 172, 196, 201

statistical analysis 2, 30, 55

statistical radio maps 18, 38

Statistical Learning 26, 37–38,

69, 71, 78

Stopping Rule (SR) 147

successive convex approximation

(SCA) 8, 248, 253

Successive Cancellation List

(SCL) 125

Supervised Learning 7, 87, 90,

107

Symbol Error Rate (SER) 205

symbol-embedded SVC

(SE-SVC) 201

t

Test Error Patterns (TEP) 131

Transmission Time Interval

(TTI) 218

Tail distribution 61, 62, 68, 71

tail-biting convolutional code

(TB-CC) 123

u

ultra-reliable 9, 291, 321

Ultra Dense Network

(UDN) 195

Ultra-Reliable and Low-Latency

Communications (URLLC) 1,

85, 119, 169, 291

Unmanned Aerial Vehicle
(UAV) 3

Unsupervised Deep Learning 7,
9, 85–89, 94, 99, 111

unsupervised learning 7, 86, 95,
105–107, 207, 215–216

uplink (UL) 97, 120, 199, 244,
263, 266, 293

User Association 3

User Experienced Delay
295–297, 312

user equipment (UE) 73, 171,
217, 323

V

Vehicle-to-Vehicle (V2V)
75, 79

Vehicular User Equipment
(VUEs) 73

Virtual Reality (VR)
9, 291, 321

Virtualized Network Function
(VNF) 218

von Mises conditions 68

W

Weighted Hamming Distance
(WHD) 134