Performance Prediction of URLLC in Interference-Limited Networks Using Transfer Learning

by

Mujtaba GHOUS

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
WITH THESIS IN ELECTRICAL ENGINEERING
M.A.Sc.

MONTREAL, MARCH 04, 2026

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Prof. George Kaddoum, Thesis supervisor
Département de génie électrique, École de technologie supérieure

Prof. Rami Langar, Chair, President
Département du génie logiciel et des TI (LOGTI), École de technologie supérieure

Prof. Bassant Selim, Board of Examiner
Département de génie des systèmes (GSYS), École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON FEBRUARY 09, 2026

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

FOREWORD

This thesis is conducted within the context of the rapid evolution of wireless communication networks, where reliability and latency requirements are becoming increasingly stringent. The choice of this topic is motivated by a strong interest in ultra-reliable and low-latency communications (URLLC), as well as by the challenges associated with short-packet communication in interference-prone environments. These issues are central to the design of next-generation communication systems.

The primary objective of this work is to analyze the performance of URLLC systems operating in clustered networks under the finite blocklength regime, and to investigate approaches that can improve performance prediction. In particular, this thesis highlights the limitations of conventional analytical models in heterogeneous environments and examines the contribution of deep learning and transfer learning techniques to enhance predictive capabilities.

The scope of this study is intentionally restricted to specific configurations and assumptions in order to ensure a rigorous and well-controlled analysis. Although these choices introduce certain limitations, they enable interpretable results and provide a solid foundation for future research. Overall, this thesis combines analytical modeling and data-driven methods to contribute to the understanding and design of reliable, low-latency communication systems.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all the individuals and institutions who contributed, directly or indirectly, to the completion of this thesis.

First and foremost, I would like to thank **Professor Georges Kaddoum**, my research supervisor, for his rigorous scientific guidance, insightful advice, and continuous support throughout this work. His availability and the quality of his guidance played a decisive role in the successful completion of this research.

This work was supported financially through the **ULTRA TCS Research Chair on Intelligent Tactical Radio Networks for Constrained Environments**, funded by the **Natural Sciences and Engineering Research Council of Canada (NSERC)** under Grant **CRDPJ 538896–19**. This support contributed significantly to the research activities presented in this thesis.

I would also like to acknowledge **Dr. Tri Nhu Do** for his scientific contributions during the initial phases of the project, as well as **Thanh Luan** for his collaboration, technical discussions, and support throughout the research activities.

I would like to thank the École de technologie supérieure (ÉTS), along with the administrative staff and research support services, for providing the resources, infrastructure, and academic environment that supported my graduate studies.

On a more personal note, I am deeply grateful to my family for their moral support, patience, and constant encouragement throughout my academic journey. Their support was a fundamental element in the successful completion of this thesis.

Finally, I would like to thank my friends for their understanding and support during this demanding period, which helped maintain an essential balance throughout the completion of this work.

Prédiction des performances de l'URLLC dans des réseaux soumis à interférences à l'aide de l'apprentissage par transfert

Mujtaba GHOUS

RÉSUMÉ

Dans le contexte émergent des réseaux de sixième génération (6G), garantir une fiabilité extrêmement élevée pour des transmissions de courte longueur constitue un défi majeur, les limites classiques de capacité de Shannon n'étant plus adaptées au régime de blocs finis. Ce mémoire s'intéresse à cette problématique et présente une analyse des performances d'un système de communication ultra-fiable et à faible latence (URLLC) déployé dans un réseau en grappes utilisant la communication par paquets courts (SPC). Dans cette architecture, plusieurs utilisateurs terrestres soumis à diverses sources d'interférence sont regroupés autour d'un chef de cluster (CH) assurant le rôle de relais sans fil entre la station de base et les utilisateurs.

Une expression analytique en forme fermée du taux d'erreur de bloc (BLER) global est d'abord établie en tenant compte de la taille du paquet, de la longueur de bloc et du débit maximal atteignable. Afin de compléter cette analyse, un cadre de transfert d'apprentissage est proposé pour la prédiction en temps réel des performances du système dans des environnements statistiquement indépendants mais non nécessairement identiquement distribués (Non-IID). L'approche consiste à exploiter un modèle source pré-entraîné puis à l'affiner sur des données propres au domaine d'intérêt, permettant ainsi d'améliorer significativement la capacité de généralisation face à différents scénarios d'interférence.

Les résultats obtenus mettent en évidence l'intérêt de combiner modélisation analytique et apprentissage profond pour la caractérisation précise des performances URLLC dans le régime de blocs finis. L'approche proposée offre une solution efficace pour la prédiction dynamique du BLER et ouvre la voie à un déploiement plus fiable des systèmes URLLC dans des applications pratiques où l'évaluation instantanée des performances demeure essentielle.

Mots-clés: Communication ultra-fiable et à faible latence (URLLC), apprentissage par transfert (TL), taux d'erreur de bloc (BLER), réseau neuronal convolutionnel (CNN), analyse de performances, réseau d'interférences non identiquement distribuées (non-IID), prédiction des performances basée sur l'apprentissage automatique

Performance Prediction of URLLC in Interference-Limited Networks Using Transfer Learning

Mujtaba GHOUS

ABSTRACT

In the emerging context of sixth-generation (6G) networks, a major challenge is ensuring extremely high reliability for short-packet transmissions. This challenge is linked to the fact that classical Shannon capacity limits are no longer applicable in the finite-blocklength regime. The present thesis addresses this issue by presenting a comprehensive performance analysis of a URLLC system deployed in a clustered wireless network using short-packet communication (SPC). In this architecture, multiple ground users affected by heterogeneous interference sources are grouped around a cluster head (CH), which acts as a wireless relay between the base station and the users.

A closed-form analytical expression for the global block error rate (BLER) is first derived by considering key constraints such as packet size, blocklength, and achievable rate. In addition, a transfer-learning framework is proposed for real-time performance prediction in environments statistically independent, but not necessarily identically distributed (Non-IID). The approach uses a pre-trained source model that is subsequently fine-tuned using domain-specific data, significantly improving generalization capability under diverse interference scenarios.

The results highlight the benefits of combining analytical modeling with deep learning for accurate characterization of URLLC performance in the finite-blocklength regime. The proposed method provides an effective solution for dynamic BLER prediction and paves the way for a more reliable deployment of URLLC systems in practical applications requiring instantaneous performance evaluation.

Keywords: Ultra-reliable and low-latency communications (URLLC), transfer learning (TL), block error rate (BLER), convolutional neural network (CNN), performance analysis, non-identically distributed (Non-IID) interference networks, machine learning based performance prediction

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

URLLC	Ultra-Reliable Low Latency Communication
QoS	Quality of Service
SPC	Short Packet Communication
FBL	Finite Blocklength
BLER	Block Error Rate
PER	Packet Error Rate
SNR	Signal-to-Noise Ratio
SINR	Signal-to-Interference-plus-Noise Ratio
INID	Independent but Not Identically Distributed
IID	Independent and Identically Distributed
PDF	Probability Density Function
CDF	Cumulative Distribution Function
MCS	Modulation and Coding Scheme
BER	Bit Error Rate
LLR	Log-Likelihood Ratio
CH	Cluster Head
BS	Base Station
UE	User Equipment
AWGN	Additive White Gaussian Noise

RV	Random Variable
RIS	Reconfigurable Intelligent Surface (if mentioned in references)
NOMA	Non-Orthogonal Multiple Access
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
DF	Decode-and-Forward
AF	Amplify-and-Forward
LOS	Line-of-Sight
NLOS	Non-Line-of-Sight
ML	Machine Learning
AI	Artificial Intelligence
DL	Deep Learning
DNN	Deep Neural Network
CNN	Convolutional Neural Network
TL	Transfer Learning
FT	Fine-Tuning
SGD	Stochastic Gradient Descent
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
MAE	Mean Absolute Error

5G	Fifth-Generation
6G	Sixth-Generation
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CSIR	Channel State Information at the Receiver

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

P	Transmission power (W)
N_0	Noise power spectral density (W/Hz)
σ^2	Noise variance
γ	Instantaneous signal-to-noise ratio (SNR)
Γ	Average SNR
ρ	Interference-to-noise ratio (INR)
h	Channel fading coefficient
$\|h\|^2$	Channel power gain
I	Total interference power
n	Blocklength (number of channel uses)
k	Number of information bits
R	Transmission rate (bits/channel use)
ε	BLER (block error rate)
$Q(\cdot)$	Gaussian Q-function
$f_X(x)$	Probability density function (PDF) of X
$F_X(x)$	Cumulative distribution function (CDF) of X
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
ω	Weight parameter in neural network
b	Bias parameter in neural network

$\hat{y}$	Predicted output of the model
y	Ground truth label
$L(\cdot)$	Loss function (e.g., MSE loss)
η	Learning rate
θ	Set of model parameters
Δ	Domain shift between datasets (Non-IID)
S	Source domain dataset
T	Target domain dataset
$\phi(\cdot)$	Feature-extraction function of the network
τ	Latency (ms)
d	Distance between transmitter and receiver (m)

INTRODUCTION

Modern wireless communication networks are evolving rapidly to meet the increasing demands of mission-critical applications that require extremely reliable and low-latency communication. Among these requirements, Ultra-Reliable Low-Latency Communications (URLLC) has emerged as a key component of 5G networks and is expected to play an even more crucial role in future 6G architectures. URLLC aims to ensure end-to-end latency on the order of a few milliseconds while maintaining an extremely low error probability, typically below 10^{-5} or even 10^{-9}.

However, achieving such stringent performance targets poses significant challenges. In practical scenarios, communication systems operate under finite blocklength constraints, where short packets, dynamic interference conditions, and heterogeneous radio environments significantly affect system performance. Under these conditions, classical asymptotic results from information theory are no longer sufficient to accurately characterize real system behavior.

In real-world wireless environments, users are subjected to interference originating from multiple independent sources with diverse statistical characteristics. These interference components are often independent but not identically distributed (Non-IID), reflecting realistic dense network deployments. When interference becomes dominant, the reliability of URLLC systems strongly depends on a detailed analysis of the Block Error Rate (BLER), which is influenced by key physical parameters such as channel gains, transmit power, blocklength, modulation schemes, and coding strategies. Nevertheless, deriving closed-form mathematical expressions for BLER under such complex conditions often leads to highly involved analytical formulations that are difficult to utilize in real-time system design and decision-making processes. Therefore, a rigorous investigation of these effects remains essential for accurate system dimensioning and for ensuring compliance with URLLC reliability constraints.

To address these limitations, machine learning–based approaches have gained considerable attention as efficient tools for predicting the performance of wireless communication systems. In particular, deep neural networks offer the capability to capture complex nonlinear relationships among multiple system parameters without requiring explicit analytical expressions for every network configuration. In the URLLC context, such models can learn to predict BLER directly from measured or simulated data, enabling fast and adaptive performance estimation in dynamically changing environments. However, training reliable predictive models typically requires large volumes of representative data, which may become costly when network conditions or interference characteristics vary. Transfer learning provides an effective solution to this challenge by allowing knowledge learned from a source domain to be reused and adapted to a new target domain with different interference statistics.

In this thesis, we investigate a downlink URLLC communication system in which a base station transmits short packets to multiple ground users through a cluster head acting as a relay. The system operates under independent but non-identically distributed (Non-IID) interference, representing realistic dense network scenarios. We first develop a finite blocklength analytical framework to characterize BLER as a function of system parameters. Next, we design a deep learning model capable of predicting system performance efficiently. Finally, we propose a transfer learning strategy to enable rapid adaptation of the predictive model across varying interference environments.

Motivation

URLLC plays a critical role in a wide range of emerging applications, including smart energy systems, collaborative robotics, industrial automation, telemedicine, and vehicular communications. In these scenarios, even minor transmission delays or errors can lead to severe operational consequences. Although classical analytical models remain valuable for understanding fundamental system behavior, their practical applicability becomes limited as system complexity increases and interference patterns exhibit Non-IID characteristics.

In this context, a hybrid approach that combines rigorous analytical modeling with data-driven predictive techniques becomes highly desirable. This work is motivated by the need for a comprehensive framework that offers both theoretical insight and fast, accurate performance prediction under realistic operating conditions.

Objectives

The main objective of this thesis is to develop a comprehensive framework for evaluating and predicting the performance of a URLLC system operating under Non-IID interference conditions. This objective includes the analytical modeling of BLER in the finite blocklength regime, the design of a deep learning–based predictive model capable of estimating BLER from physical system parameters, and the integration of transfer learning techniques to enhance model generalization across diverse interference environments. The overall goal is to provide a robust solution that satisfies the stringent requirements of accuracy, computational efficiency, and adaptability imposed by URLLC applications.

Methodology

The proposed methodology is structured around three interrelated components. First, a rigorous theoretical analysis is conducted to derive analytical expressions for BLER under finite blocklength constraints while accounting for Non-IID interference. Second, a deep neural network model is developed to learn the complex mapping between system parameters and BLER outcomes. Finally, a transfer learning framework is implemented to enable efficient adaptation of the predictive model to new environments without requiring full retraining. This integrated approach ensures both accuracy and flexibility in performance evaluation.

Contributions

The main contributions of this thesis can be summarized as follows:

- **Comprehensive finite blocklength BLER analysis under Non-IID interference.** A key contribution of this work is the development of a rigorous analytical framework that characterizes the block error rate across different communication links in the considered URLLC system. Unlike conventional approaches based on asymptotic assumptions, this analysis explicitly accounts for finite blocklength effects and the presence of statistically independent but non-identically distributed interference sources, providing deeper insights into system reliability under realistic conditions.
- **Design of a deep neural network for efficient BLER prediction.** A second contribution is the development of a deep learning model capable of accurately predicting BLER based on system parameters such as channel conditions, transmit power, blocklength, and interference levels. This approach circumvents the complexity of direct analytical computation while offering rapid and precise performance estimation across diverse network scenarios.
- **Development of a transfer learning framework for rapid adaptation.** Recognizing the high data requirements of deep learning models, a transfer learning strategy is proposed to adapt a pretrained model from a source domain to a target domain with different interference characteristics. This significantly reduces data collection effort and training time while maintaining strong predictive performance.
- **Extensive numerical validation of analytical and predictive approaches.** Finally, a comprehensive validation study is conducted to compare analytical results, simulation outcomes, and neural network predictions. This evaluation confirms the accuracy of the analytical derivations, the effectiveness of the predictive model, and the benefits of transfer learning in realistic Non-IID interference environments.

Publication

The research presented in this thesis has led to the following publication: M. Ghous, T. L. Nguyen, T. N. Do, and G. Kaddoum, “Deep Transfer Learning-Based Performance Prediction of URLLC in Independent and Not Necessarily Identically Distributed Interference Networks,” *IEEE Access*, vol. 12, pp. 99071–99093, 2024.

Thesis Organization

This thesis is organized as follows:

- **Chapter 1** presents the general introduction and literature review on URLLC performance analysis and machine learning–based prediction methods;
- **Chapter 2** provides theoretical performance analysis and finite blocklength BLER derivations;
- **Chapter 3** introduces the machine learning and transfer learning–based BLER prediction framework;
- **Chapter 4** presents numerical results and performance comparisons;
- **Chapter 5** concludes the thesis and outlines future research directions.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

In recent years, short-packet communication (SPC) has become a central enabler for ultra-reliable low-latency communications (URLLC) in beyond sixth-generation (B5G) and sixth-generation (6G) wireless networks. SPC enables the transmission of very small payloads under strict reliability and latency constraints, making it essential for mission-critical applications such as industrial automation, autonomous systems, medical robotics, high-speed transportation, and massive Internet-of-Things (IoT) deployments (Vaezi *et al.*, 2022). Early research, such as (Aijaz *et al.*, 2018), demonstrated practicality of SPC by examining the challenges associated with transmitting small packets and providing mechanisms for achieving stringent reliability requirements. Later studies expanded SPC analysis to 6G systems (Zhang *et al.*, 2022), highlighting the rising demand for network architectures capable of delivering sub-millisecond latency with extremely low block error rates (BLER). An overview of SPC and URLLC relationship .

URLLC introduces design constraints that significantly depart from traditional Shannon-capacity assumptions. Instead of relying on asymptotically large blocklengths, URLLC performance must be analyzed in the finite blocklength (FBL) regime. In this regime, a reduction in blocklength lowers transmission latency, but leads to increased BLER, thus creating a fundamental trade-off that must be carefully quantified. Several previous studies e.g., (Chen *et al.*, 2019b), (Ju *et al.*, 2019) showed that classical infinite-blocklength assumptions cannot accurately characterize reliability in URLLC scenarios, particularly when systems operate with strict latency budgets. Therefore, FBL-based reliability metrics are needed to evaluate the true performance limits of modern URLLC systems. Figure 1.1 illustrated the fundamental relationship between SPC and URLLC.

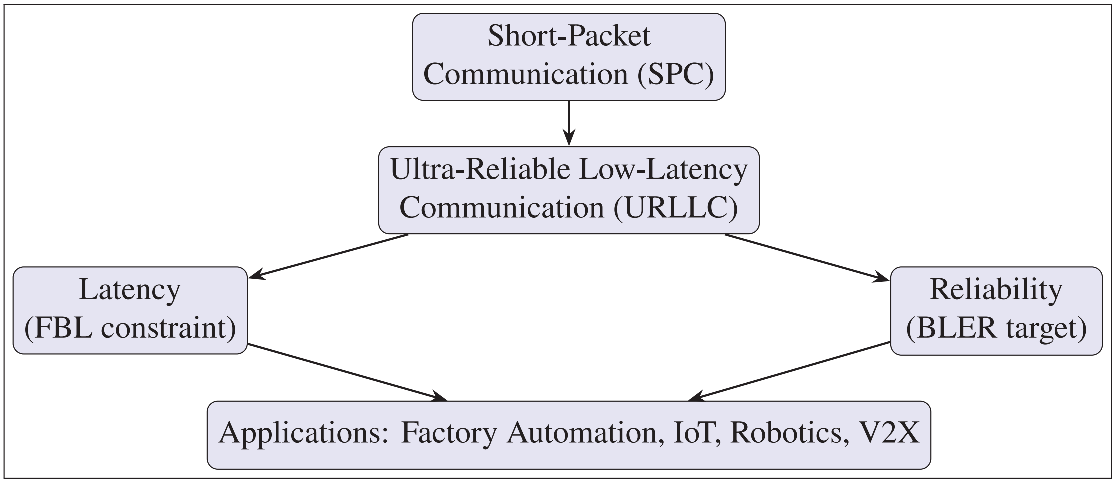

Figure 1.1 Overview of SPC and URLLC relationships

A significant challenge for URLLC arises from heterogeneity in propagation environments. Many wireless systems operate under statistically independent, but not identically distributed (non-IID) fading due to spatially separated transmitters, distributed antenna systems, or heterogeneous interferers. (Tao, 2009) highlighted that, in distributed networks, channel gains substantially differ across links, leading to non-IID fading conditions that directly affect reliability. For systems requiring broadcasting or multicasting, such as the cluster-head architecture studied in this thesis, non-IID effects create non-uniform reliability constraints and complicate BLER modeling.

URLLC's strict latency constraints further challenge analytical modeling. Transmission latency, processing delay, and queuing contribute to end-to-end delay; however, transmission delay remains the most dominant component in FBL analysis (Chen *et al.*, 2018). When networks experience strong interference, as in dense deployment scenarios, analytical models become increasingly difficult to evaluate due to the presence of multiple summations and special functions in closed-form FBL expressions.

Machine learning (ML) approaches have emerged as promising solutions to approximate wireless performance under complex and dynamic environments. Deep neural networks (DNNs) have demonstrated strong predictive capabilities for wireless metrics such as secrecy outage (Xu *et al.*, 2022), outage probability (Xu *et al.*, 2021), channel quality, and performance under imperfect or rapidly time-varying channels. These models can learn nonlinear relationships between channel conditions, interference levels, and reliability metrics without requiring analytically tractable models. However, ML training typically requires large datasets, and retraining for every new interference scenario is computationally expensive. Transfer learning (TL) provides a remedy by enabling adaptation of pretrained models to new environments with minimal training.

In this chapter, we provide a comprehensive overview of the foundational concepts, classical analytical approaches, and modern ML-based prediction frameworks relevant to URLLC. We also identify research gaps and motivates the hybrid analytical ML TL framework developed in later chapters.

1.2 Literature Review

1.2.1 Short-Packet URLLC and Physical-Layer Reliability

The foundations of SPC and its role in URLLC were established in a number of works, where the main focus was enabling small-packet transmission under very stringent latency and reliability constraints (Vaezi *et al.*, 2022; Aijaz *et al.*, 2018; Zhang *et al.*, 2022; Prathyusha & Sheu, 2022). These studies highlighted that applications such as factory automation, intelligent transportation, and mission-critical IoT cannot rely on conventional best-effort communication, but instead require strict guarantees on delay and error probability.

Most early physical-layer analyses assumed infinite blocklength and relied on Shannon capacity to characterize performance (Chen *et al.*, 2019b; Ju *et al.*, 2019). However, in URLLC scenarios, such assumptions break down because packets are necessarily short. Finite blocklength (FBL) information theory provides more accurate bounds on achievable rates and BLER when blocklength is limited (Chen *et al.*, 2018). Recent work continued to refine FBL-based URLLC performance analysis. For instance, the authors in (Lam *et al.*, 2023) investigated short-packet URLLC transmission schemes in quasi-static channels, showing how different coding and diversity techniques affect BLER under practical delay constraints. Similarly, (Interdonato *et al.*, 2024) developed a unified FBL framework for URLLC in cell-free massive MIMO systems, emphasizing that naive use of outage capacity can significantly underestimate error probability in the ultra-reliable regime.

These contributions confirm that FBL modeling is not only theoretically necessary, but also practically essential for accurately capturing URLLC performance in realistic scenarios where packets are short and reliability targets are extremely stringent.

1.2.2 Blocklength Optimization and Resource Allocation

Since blocklength directly influences the trade-off between delay and reliability, it has been the subject of several optimization studies. In (Pala *et al.*, 2024), a multi-RIS-assisted rate-splitting multiple access (RSMA) system was proposed to improve URLLC performance. The results showed that it can achieve higher throughput than NOMA and SDMA while satisfying FBL reliability constraints. Furthermore, (Chen *et al.*, 2019a) considered a wireless-powered IoT network where a hybrid access point first transfers energy and then receives short packets, formulating optimization problems that jointly adjust transmission time and packet error rates to maximize effective throughput. Figure 1.2 presents the finite blocklength reliability evaluation flow adopted in (Chen *et al.*, 2019a).

More recent research investigated FBL URLLC in various network architectures. For instance, (Zhang *et al.*, 2024) analyzed ultra-reliable and low-latency wireless hierarchical federated learning (URLLC-WHFL) using multi-antenna transceivers and a secure FBL design, revealing trade-offs between privacy, utility, and physical-layer security. In another relevant study, (Huang *et al.*, 2024), short-packet communications were examined in wireless energy transfer scenarios, and a deep-learning-based framework was proposed to predict BLER and throughput under relay selection strategies. Figure 1.3 depicts a representative concept of blocklength and resource optimization. Energy-aware URLLC design also attracted attention; for instance, (Iqbal *et al.*,

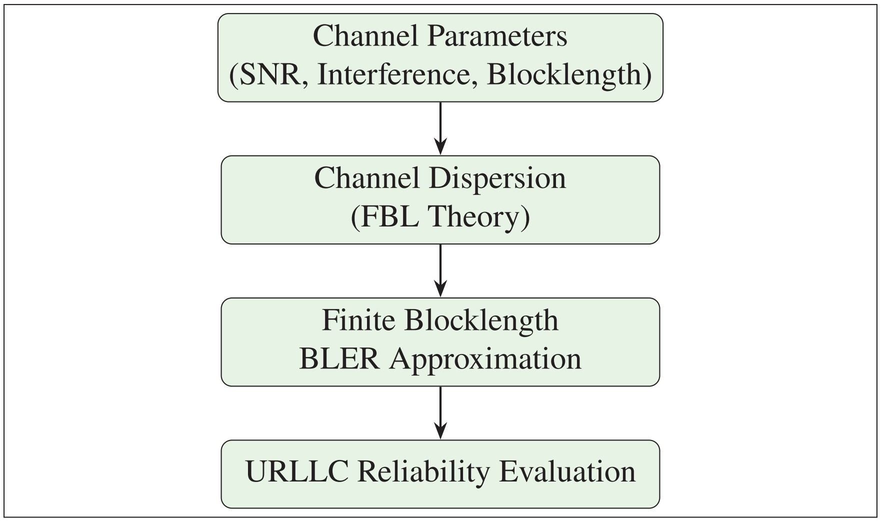

Figure 1.2 Finite blocklength reliability evaluation flow

2025) proposed an energy-efficient URLLC scheme for beyond-5G and 6G-enabled hybrid IoT systems, highlighting the joint impact of reliability, latency, and energy constraints.

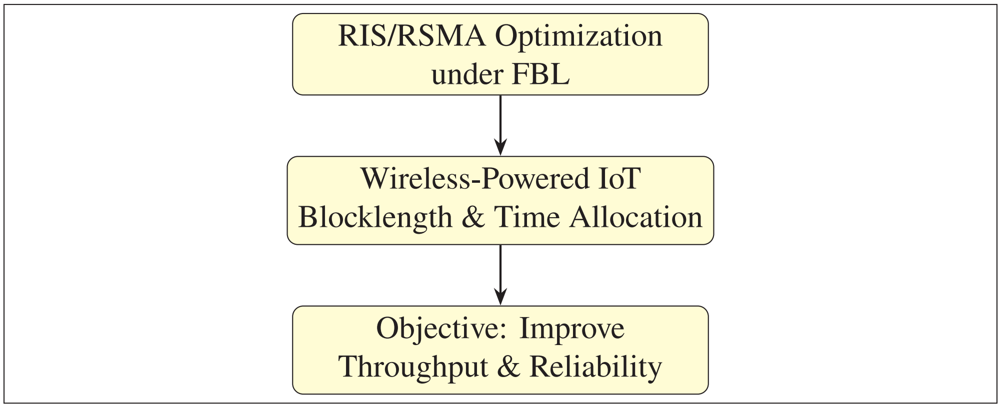

Figure 1.3 Representative blocklength and resource optimization concept

TThe studies briefly reviewed above show that blocklength, power allocation, and time allocation must be jointly optimized for modern URLLC systems, particularly in the presence of additional constraints such as energy harvesting, security, or federated learning.

1.2.3 Non-IID Interference Environments

In many realistic deployments, wireless links experience statistically independent, but not identically distributed fading due to different propagation conditions, antenna locations, or shadowing environments. As discussed in (Tao, 2009), this non-IID behavior becomes particularly relevant in distributed antenna systems or clustered networks, where each link may have distinct large-scale and small-scale fading characteristics.

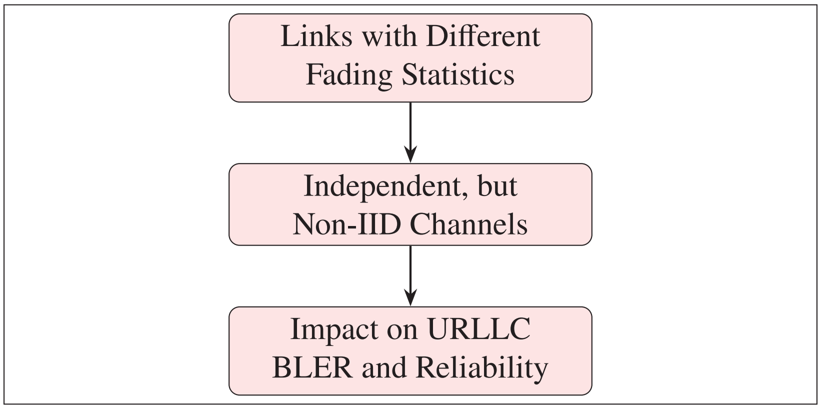

Figure 1.4 Characterization of Non-IID wireless environments

Recent work extended FBL analysis to more realistic topologies. For example, (Zhang *et al.*, 2023) studied multidevice downlink cell-free systems under FBL and strict deadline constraints, showing how spatially distributed access points and heterogeneous channel conditions affect packet error probabilities. A more recent unified framework proposed in (Interdonato *et al.*, 2024) considered the coexistence of URLLC and enhanced mobile broadband (eMBB) users in cell-free massive MIMO systems, again under FBL constraints. These studies demonstrate that non-IID fading and interference patterns significantly influence URLLC performance and must be explicitly modeled rather than approximated by IID assumptions. Figure 1.4 characterized non-IID wireless environments, with varied channel statistics and interference conditions across users and links.

1.2.4 Machine Learning for Wireless Reliability Prediction

Machine learning has become a powerful tool for performance prediction in wireless systems where analytical modeling is difficult or intractable. CNNs and other deep architectures have been extensively used for resource allocation, queue management, and congestion control in IoT and wireless networks (Mao *et al.*, 2018). For example, (Xu *et al.*, 2022) proposed a CNN-based approach for predicting secrecy performance in 6G wireless networks, while (Xu *et al.*, 2021) employed CNNs to estimate outage probability in mobile IoT systems with cooperative communication and antenna selection. Complexity-focused work such as (Oikonomou *et al.*, 2024) analyzed how CNN depth and architecture impact computational cost, providing guidance for designing models capable of operating under real-time constraints.

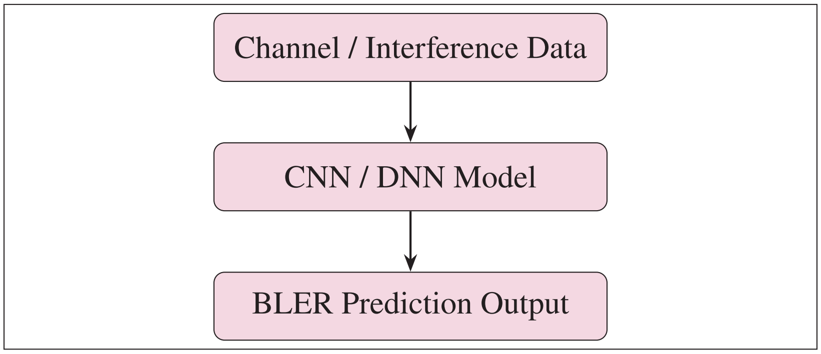

Figure 1.5 ML-based wireless reliability prediction pipeline

More broadly, recent surveys (Liu *et al.*, 2023a; Chen *et al.*, 2024) reviewed the use of ML and deep learning at the physical layer, highlighting applications in channel estimation, signal detection, and link adaptation for 6G networks. Figure 1.5 illustrated the ML-based wireless reliability prediction pipeline. Other contributions, such as (Davey *et al.*, 2024), explored over-the-air training methods that do not require a feedback channel and allow models to be trained directly on the physical channel. These works indicate that deep learning is increasingly viewed as a core component of future PHY-layer designs, particularly in regimes like URLLC where complex channel behaviors and hardware imperfections make closed-form analysis challenging.

1.2.5 URLLC in 6G and Time-Sensitive Communication

Several previous studies particularly examined URLLC in the context of future 6G systems. For instance, (Nguyen *et al.*, 2021) identified the main challenges associated with high-mobility scenarios and time-sensitive data collection, while (Liu *et al.*, 2023b) considered the coexistence of Wi-Fi and NR-U systems and used centralized and federated deep reinforcement learning to optimize URLLC channel access. High-mobility channel prediction was analyzed in (Glinskiy *et al.*, 2023), where autoregressive and neural network-based predictors were compared under realistic channel models.

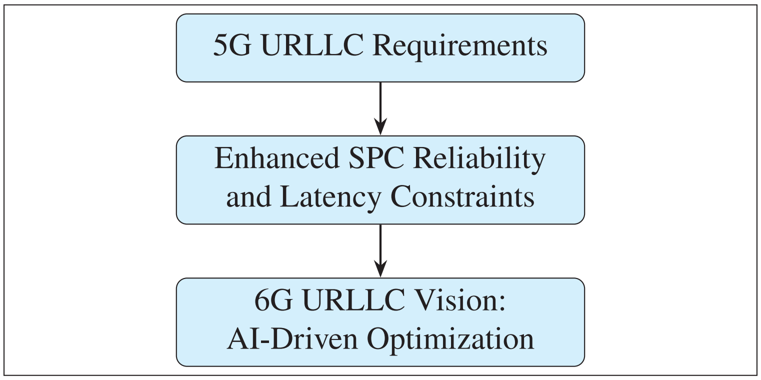

Figure 1.6 Evolution from 5G URLLC to 6G URLLC

Concurrently, new frameworks are being developed to address delay and reliability jointly. The CCSND model proposed in (Cai *et al.*, 2024) introduced a contrastive convolutional structure for network delay prediction, achieving significantly lower mean squared error than several benchmark time-series models. Moreover, a recent 2024 finite-blocklength framework for IRS-aided MIMO URLLC proposed by (Kumar & Singh, 2024) derived tight upper and lower bounds on error probability and proposed gradient-based phase shift optimization. The work in (Maghsoudnia *et al.*, 2024) critically examined the gap between theoretical URLLC promises and practical deployments, arguing that protocol, scheduling, and implementation constraints still limit URLLC performance in real networks. The evolution from 5G URLLC to 6G URLLC is illustrated in Figure 1.6.

Furthermore, surveys such as (Shaik & Talukdar, 2024) discussed AI/ML-based capacity maximization strategies for URLLC, including resource allocation, multi-access design, and beamforming under stringent reliability and latency constraints. More recently, (Sharma & Agrawal, 2025) investigated the role of digital twins in 6G-based URLLC, showing how virtual replicas of physical systems can be exploited to anticipate network behavior and improve reliability.

1.2.6 Multihop, Cognitive, and IoT URLLC Systems

Multihop and cooperative relay architectures were previously investigated to support large URLLC coverage areas and improve reliability. In (Tu *et al.*, 2023), short-packet transmissions in underlay cognitive multihop MIMO relay networks were studied, with ML-based estimators used to predict performance metrics such as throughput, latency, and reliability. Full-duplex energy-harvesting IoT networks for URLLC applications were analyzed in (Nguyen *et al.*, 2023), where deep learning models were used for real-time BLER and throughput prediction under relay selection policies. Figure 1.7 presented the general workflow of transfer learning for wireless BLER prediction.

More generally, recent work on short-packet communications in wireless energy transfer systems, such as (Huang *et al.*, 2024), proposed deep learning frameworks for system performance prediction in real-time configurations.

These contributions indicate that URLLC performance prediction in heterogeneous and energy-constrained systems can benefit from a combination of analytical FBL modeling and data-driven estimation.

1.2.7 Transfer Learning for Wireless Systems

Recently, transfer learning has emerged as a key technique to reduce the data and computational requirements of deep learning models in wireless communication. Several recent surveys e.g., (Niu *et al.*, 2020; Zhou *et al.*, 2023; Wang *et al.*, 2021; Coutinho & Boukerche, 2021) describe how TL can be categorized into inductive, transductive, and unsupervised approaches, and how

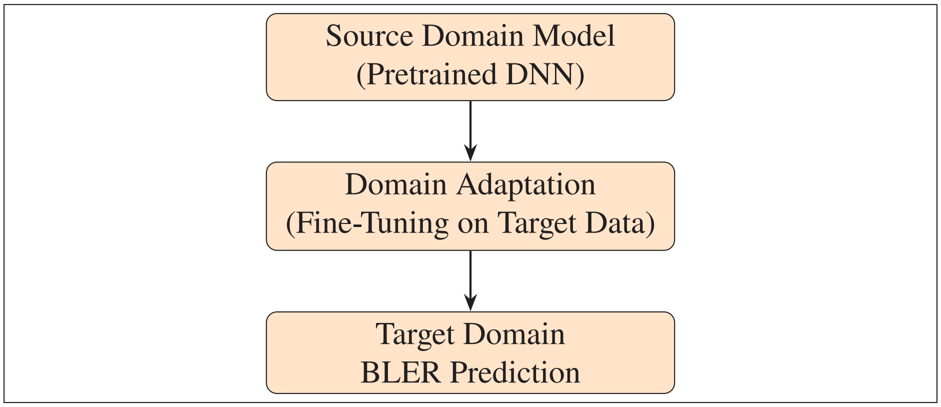

Figure 1.7 Workflow of transfer learning for wireless BLER prediction

it can be used to adapt models across domains or tasks. In cognitive radio and spectrum sensing, several other studies (Peng *et al.*, 2019; Zheng *et al.*, 2020) proposed TL-based CNN frameworks that transfer models trained on one set of signals or environments to new sensing tasks with limited data.

More recent works explicitly explored TL in physical-layer tasks. For instance, (Uyoata *et al.*, 2024) investigated TL for fully convolutional networks used in wireless signal detection, finding that models trained under one channel condition can be efficiently adapted to another without full retraining. A broader perspective on TL for signal detection and channel estimation was discussed in (Li *et al.*, 2022). These results highlight the potential of TL for adapting wireless ML models to new configurations, which is particularly relevant for URLLC scenarios where interference statistics and channel conditions may change frequently.

However, despite these advances, previous studies have not yet addressed transfer-learning-based BLER prediction in URLLC systems with non-IID interference and clustered broadcasting architectures. In most prior works, either FBL-based analytical performance was considered without ML, or ML models were applied without exploiting transfer learning in non-IID interference scenarios. This gap motivates the hybrid analytical deep learning transfer learning framework proposed in this thesis, which combines closed-form BLER expressions and TL-enhanced prediction models to provide accurate, scalable, and adaptable performance prediction in realistic URLLC networks.

1.3 Existing Research Gap

Available literature on SPC, URLLC, and FBL provided strong analytical foundations, while ML-based prediction frameworks demonstrated excellent performance in dynamic and complex environments. However, prior studies either relied purely on mathematical modeling or applied ML methods in scenarios unrelated to FBL-based URLLC reliability. Another limitation is that, previous ML works generally assumed IID interference or fading distributions and did not incorporate non-IID characteristics or broadcast-relay structures.

Although TL was applied to classification, sensing, and cognitive radio systems, prior work did not examine TL-enabled BLER prediction for URLLC systems with non-IID multiuser interference. To fill this gap in the literature, in the present thesis, we propose a method that combines FBL theory with TL-driven reliability prediction under realistic interference conditions.

while previous research highlighted important advances in short-packet transmission, finite blocklength reliability analysis, and machine learning–based performance prediction, none of the reviewed studies addressed the combined challenges posed by clustered URLLC architectures, heterogeneous non-IID interference patterns, and the need for real-time BLER prediction across diverse system configurations. This lack of unified analytical and learning-based models motivates the methodology developed in the present thesis.

To address this gap, in Chapter 2, we outline a detailed system model for a clustered URLLC downlink scenario with aggregated interference, followed by the derivation of new finite-blocklength analytical expressions for hop-wise and end-to-end BLER.

CHAPTER 2

ANALYTICAL CHARACTERIZATION OF URLLC PERFORMANCE IN INTERFERENCE-LIMITED SCENARIOS

In this chapter, we introduce the system model and the analytical principles that form the basis of our study. We outline the motivations behind the performance analysis and highlight the main challenges in predicting BLER under interference-limited URLLC conditions. The next sections present the mathematical formulation and the contributions of this analysis.

2.1 Motivations and Contributions

Analytical formulas for calculating the average BLER are generally complex due to the involvement of complex functions, e.g., Q-function, render them infeasible for time-sensitive scenarios, such as real-time, low-latency, and URLLC systems. On the other hand, while CNNs are excellent at providing accurate predictions, they often require a large amount of training data, with potentially high network complexity for accurate predictions. However, having a large dataset is infeasible in dynamic environments. Similarly, complex neural network models are unsuitable for URLLC-based application.

In response, our research aims to find a balance by introducing a more efficient model that avoids the computational challenges of the analytical method and the complexity associated with CNNs, allowing for both speed and accuracy in predicting BLER for real-time applications. For this, we propose a transfer learning method for URLLC-based clustered wireless networks with short packet communication. Transfer learning allows us to leverage knowledge gained from a different but related task or dataset. Instead of starting from scratch, we design machine learning models that exploit pre-learned features from a source model. The designed models require small dataset, and significantly reduces the computational complexity and training time, making it more applicable for real-time applications and for dynamic wireless environments. To the best of the authors' knowledge, such a URLLC system has not been considered in the literature. We summarize our contributions as follows:

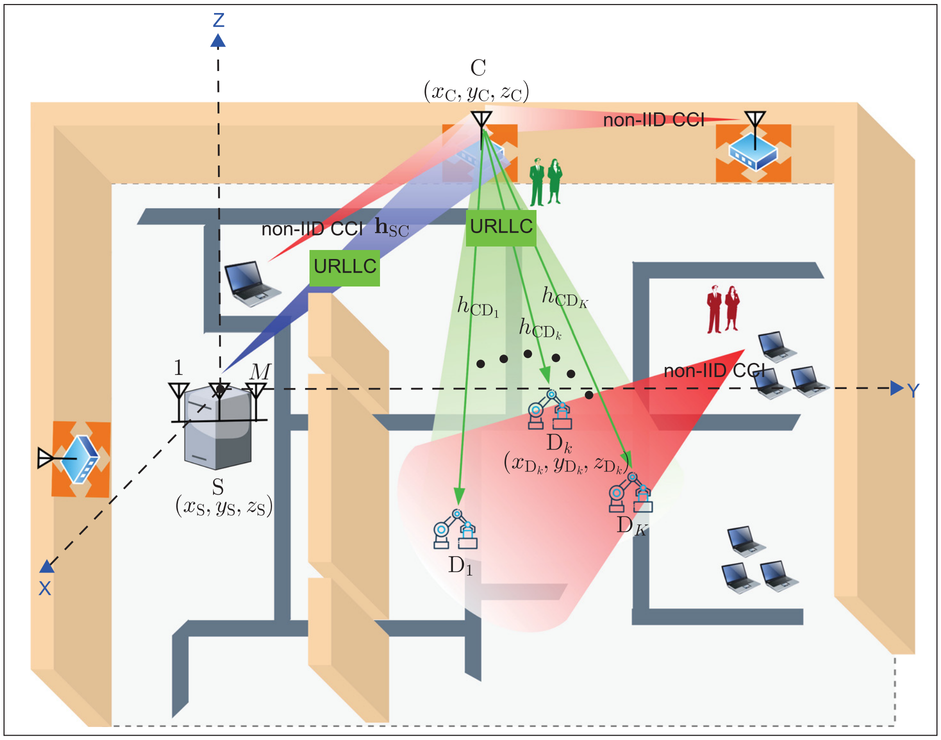

Figure 2.1 Illustration of the URLLC system in the presence of non-IID interferers, where a M-antenna base station (S) communicate with the user equipments (D_k) with the help of a cluster head (C)

- Utilizing the Q-function approximation, we derive closed-form expressions for the average per-hop BLERs and the average e2e BLER of the considering SPC system consisting of multiple users operating on a shared frequency resource with Non-IID co-channel interference (CCI) at both the cluster head and individual users. It is noted that the presence of CCI increases the computational complexity of the resulting expressions due to the presence of a bivariate Meijer G-function.
- To mitigate the complexity of the derived expressions, we propose a novel approximation to express the average e2e BLER in terms of the average per-hop per-user BLER using elementary functions, effectively addressing the analytical complexity challenge. Using these derived expressions, we assess system reliability through the error floor, asymptotic performance, and the diversity order.

Table 2.1 List of mathematical notations

Notation	Description
S	Source
C	Cluster head
D_K	K-th ground user
$\mathcal{I}$	Interferer
$\mathbf{h}_{\mathrm{SC}}$	Channel vector from source-to-cluster head
$\mathbf{h}_{\mathbf{I_i}\mathrm{C}}$	Channel vector from interferer-to-cluster head
N	Number of antennas at transmitter
P_{S}	Transmit power of the source
P_{I_i}	Transmit power of the interferer
$x_{I_i}.$	Normalized energy signal of the interferer
$\sqrt{l}_{\mathrm{SC}}$	Large scale fading of the S-to-C link
$\mathbf{w}_{\mathrm{S}}$	Beamforming vector at S
$m_{I_i\mathrm{C}}$	Shape parameter of the interfere, $i \in \mathcal{I}$
γ_{SC}	Instantaneous SNR from S-to-C
γ_{SC}	SINR at the C
$f_{I_{\mathrm{C}}}(x)$	PDF of the aggregated interference power at C
$F_{I_{\mathrm{C}}}(x)$	CDF of the aggregated interference power at C
y_{CD_k}	Received signal at D_k
Γ_{sys}	SINR of the system
ϕ_{SC}	Event during S-to-C transmission
ϕ_{CD}	Event during C-to-D transmission

- We propose a transfer learning-based performance prediction approach for real-time accurate predictions of the overall average BLER of the system. Numerical results show that the proposed fine-tuned target model requires less training data while achieving comparable prediction accuracy compared to the source model.
- Through numerical and simulation results, we validate the use of short packet communication in a clustered URLLC system. Simulation results show that the derived closed-form expressions provide a reasonable approximation for the downlink average e2e BLER of the system. In addition, the results show that IID interference yields lower average e2e BLER compared to Non-IID interference. We also investigate the asymptotic analysis and discuss the error floor results, where the average e2e BLER found unaffected by the interference power distribution and rather the total Interference-to-Noise Ratio (INR).

- By incorporating domain adaptation techniques, we improve the accuracy of our system model. Specifically, a discrepancy between the source and target domain decreases, the accuracy of the proposed transfer learning model increases.
- By investigating the BLER, exploring the impact of block length (M), investigating convergence and performance of transfer learning, and evaluating the impact of domain adaptation and transferred knowledge, we provide readers with invaluable insights into the potential performance and robustness of our proposed approach. This comprehensive analysis emphasizes the practical relevance and applicability of our research in real-world scenarios.

2.2 System Model

We consider a system model consisting of a base station (S) as a gateway, a tentative cluster head (C), multiple interferers $\mathcal{I} \in \{1, 2, ..., I\}$, where I denotes the number of ambient interferers, and $K \geq 1$ ground users, denoted as $\mathrm{D}_1, \mathrm{D}_2, \ldots, \mathrm{D}_K$. Due to the presence of multiple interferers, there is aggregated interference at C, and hence C experiences aggregated INR during the communication from S to C. This considered clustering strategy is depicted in Fig. 2.1, where the cluster head forwards the information from S to all ground users. Direct connectivity between S and D_k, for $k \in \{1, 2, \ldots, K\}$, is missing due to several obstacles. Therefore, it is supposed that S delivers short packets to D_k with the help of C, which functions as a relay and utilizes a decode-and-forward method while operating in half-duplex mode. C broadcasts the same information to each user. When significant message broadcasting delays are encountered, delivering urgent information requires URLLC. This need becomes particularly crucial in scenarios involving downlink multi-casting and broadcasting, where communication channels are Non-IID phases, i.e., phase I: S-to-C and phase II: C-to-D_k. During phase I, S delivers messages to C with transmit power P_{S}. If C successfully decodes the received signals, in phase II, it delivers the decoded signals to the D_k with transmit power P_{C}.

2.2.1 S-to-C Communication

The received signal at C from S is given as

$$y_{\mathrm{SC}} = \sqrt{P_{\mathrm{S}}}\sqrt{\ell_{\mathrm{SC}}}\mathbf{h}_{\mathrm{SC}}^{H}\mathbf{w}_{\mathrm{S}}x_{\mathrm{S}} + \sum_{i\in\mathcal{I}} \sqrt{\ell_{\mathrm{I}_i\mathrm{C}}}h_{\mathrm{I}_i\mathrm{C}}\sqrt{P_{\mathrm{I}_i}}x_{\mathrm{I}_i} + n_{\mathrm{C}}, \tag{2.1}$$

where x_{S} is the normalized signals transmitted by the source node S for C, n_{C} is the complex additive white Gaussian noise (AWGN) with zero-mean and variance σ^2, $\mathbf{w}_{\mathrm{S}} \in \mathbb{C}^{N\times 1}$ is the beamforming vector at S. Herein, $\sqrt{\ell_{\mathrm{SC}}}$ is the large-scale fading of the S-to-C link, $h_{\mathrm{SC}}^{(1)}$, $h_{\mathrm{SC}}^{(2)}$, $\ldots$, $h_{\mathrm{SC}}^{(N)}$ are the complex small-scale fading, $h_{\mathrm{I}_i\mathrm{C}}$ and $\ell_{\mathrm{I}_i\mathrm{C}}$ denote the complex small-scale fading and the path loss from interferer $i \in \mathcal{I}$ to C, respectively. P_{I_i} and x_{I_i} are the transmit power and the normalized unit energy signal of the interferer $i \in \mathcal{I}$, respectively. We assume that $|h_{\mathrm{I}_i\mathrm{C}}|^2$'s are Non-IID Gamma random variables, each with unit mean and respective shape parameter $m_{\mathrm{I}_i\mathrm{C}}$. The system model's large-scale fading is presented as follows.

2.2.1.1 Channel Model of the S-to-C link

We consider the 3GPP Urban Micro (UMi) path loss model In particular, the path loss from a transmitter $\mathrm{Tx} \in \{\mathrm{S}, \mathrm{C}\}$ to a receiver $\mathrm{Rx} \in \{\mathrm{C}, \mathrm{D}_1, \mathrm{D}_2, \ldots, \mathrm{D}_K\}$ is formulated as

$$\ell_{\text{Tx-Rx}}\ [\mathrm{dB}] = G_{\mathrm{Tx}} + G_{\mathrm{Rx}} - 22.7 - 26\log_{10}(f_c) - 36.7\log_{10}\left(\frac{d_{\text{Tx-Rx}}}{d_0}\right), \tag{2.2}$$

where $d_0 = 1$ [m] is the reference distance, f_c [GHz] is the carrier frequency, $d_{\text{Tx-Rx}}$ [m] is the distance from the Tx to Rx, G_{Tx} and G_{Rx} are the antenna gains (in dBi) at the Tx and the Rx, respectively.

Moreover, for decoding x_{S}, the instantaneous SNR γ_{SC}, is expressed as $\gamma_{\mathrm{SC}} = \bar{\gamma}_{\mathrm{S}}\ell_{\mathrm{SC}}|\mathbf{h}_{\mathrm{SC}}^{H}\mathbf{w}_{\mathrm{S}}|^2$, where $\bar{\gamma}_{\mathrm{S}} \triangleq \frac{P_{\mathrm{S}}}{\sigma^2}$, $(\cdot)^H$ denotes the conjugate transpose operator and $\mathbf{w}_{\mathrm{S}} = \frac{\mathbf{h}_{\mathrm{SC}}}{\|\mathbf{h}_{\mathrm{SC}}\|}$ is the beamforming vector, where $\|\mathbf{h}_{\mathrm{SC}}\|^2 = \sum_{n=1}^{N} \left|h_{\mathrm{SC}}^{(n)}\right|^2$. Hence, γ_{SC} is written as follows

$$\gamma_{\mathrm{SC}} = \bar{\gamma}_{\mathrm{S}}\ell_{\mathrm{SC}} \left|\frac{\mathbf{h}_{\mathrm{SC}}^{H}\mathbf{h}_{\mathrm{SC}}}{\|\mathbf{h}_{\mathrm{SC}}\|}\right|^2 = \bar{\gamma}_{\mathrm{S}}\ell_{\mathrm{SC}} \sum_{n=1}^{N} \left|h_{\mathrm{SC}}^{(n)}\right|^2, \tag{2.3}$$

where the first equality is obtained by substituting the value of $\mathbf{w}_{\mathrm{S}}$ and adopting the identity $\mathbf{h}_{\mathrm{SC}}^{H}\mathbf{h}_{\mathrm{SC}} = \|\mathbf{h}_{\mathrm{SC}}\|^2$.

Hence, the SINR at the cluster head is formulated as

$$\Gamma_{\mathrm{SC}} = \frac{\bar{\gamma}_{\mathrm{S}}\ell_{\mathrm{SC}}|h_{\mathrm{SC}}|^2}{\sum_{i\in \mathrm{I}} \bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}|h_{\mathrm{I}_i\mathrm{C}}|^2 + 1} = \frac{\gamma_{\mathrm{SC}}}{\mathrm{I}_{\mathrm{C}} + 1}. \tag{2.4}$$

where $\mathrm{I}_{\mathrm{C}} = \sum_{i\in\mathcal{I}} \bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}|h_{\mathrm{I}_i\mathrm{C}}|^2$ denotes the aggregated INR at C and $\bar{\gamma}_{\mathrm{I}_i} = P_{\mathrm{I}_i}/\sigma^2$. It is noted that γ_{SC} and I_{C} are independent RVs.

2.2.1.2 Statistically Non-IID to model CCI

Let us first explain the difference between the statistically independent and identically distributed (IID) and Non-IID assumptions. Specifically, IID RVs follow the same distribution with similar statistical characteristics Thus, it is apparent that the Non-IID assumption is more general than the IID assumption. On the other hands, the concept of Non-IID fading is essential and necessary in spatially distributed systems, and has been widely considered in the literature

Lemma 1. *The PDF and CDF of the aggregated interference power at the cluster head, denoted as* $f_{\mathrm{I_C}}(x)$ *and* $F_{\mathrm{I_C}}(x)$ *for* $x > 0$*, respectively, can be obtained as follows*

$$f_{\mathrm{I_C}}(x) = \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \frac{a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}^{j}}{(j-1)!} x^{j-1} e^{-x a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}}, \tag{2.5}$$

$$F_{\mathrm{I_C}}(x) = 1 - \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) e^{-x a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}} \sum_{r=0}^{j-1} \frac{(x a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}})^r}{r!}, \tag{2.6}$$

where $\mathcal{A} = \mathrm{diag}([\underbrace{a_{\mathrm{I}_1\mathrm{C}}, a_{\mathrm{I}_1\mathrm{C}}, \ldots, a_{\mathrm{I}_1\mathrm{C}}}_{m_{\mathrm{I}_1\mathrm{C}}\ \textit{times}}, \ldots, \underbrace{a_{\mathrm{I}_L\mathrm{C}}, \ldots, a_{\mathrm{I}_L\mathrm{C}}}_{m_{\mathrm{I}_L\mathrm{C}}\ \textit{times}}])$, $a_{\mathrm{I}_i\mathrm{C}} \triangleq m_{\mathrm{I}_i\mathrm{C}}/(\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}})$, $\varrho(\mathcal{A})$ *is the number of distinct diagonal elements of* $\mathcal{A}$, $a_{\mathrm{I}_{\langle 1\rangle}\mathrm{C}} > a_{\mathrm{I}_{\langle 2\rangle}\mathrm{C}} > \cdots > a_{\mathrm{I}_{\langle \varrho(\mathcal{A})\rangle}\mathrm{C}}$ *are the distinct diagonal elements in decreasing order,* $\tau_i(\mathcal{A})$ *is the multiplicity of* $a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}$, *and* $\chi_{i,j}(\mathcal{A})$ *is the* $(i, j)^{\text{th}}$ *characteristic coefficient of* $\mathcal{A}$.

Proof: The proof of Lemma 1 is given in Appendix I. ■

2.2.1.3 PDF and CDF of γ_{SC}

From (2.3), the PDF of γ_{SC}, which is a Gamma variate with shape parameter Nm_{SC} and scale parameter $\frac{\bar{\gamma}_{\mathrm{S}}\ell_{\mathrm{SC}}}{m_{\mathrm{SC}}}$, is obtained as

$$f_{\gamma_{\mathrm{SC}}}(z) = \frac{(a_{\mathrm{SC}})^{Nm_{\mathrm{SC}}}}{\Gamma(Nm_{\mathrm{SC}})} z^{Nm_{\mathrm{SC}}-1} e^{-a_{\mathrm{SC}}z},\ z > 0, \tag{2.7}$$

where $a_{\mathrm{SC}} \triangleq \frac{m_{\mathrm{SC}}}{\bar{\gamma}_{\mathrm{S}}\ell_{\mathrm{SC}}}$. In order to obtain the corresponding CDF, denoted as $F_{\gamma_{\mathrm{SC}}}(z)$, we integrate (2.7) as follows

$$F_{\gamma_{\mathrm{SC}}}(z) = \int_0^z f_{\gamma_{\mathrm{SC}}}(z)\mathrm{d}z = \frac{(a_{\mathrm{SC}})^{Nm_{\mathrm{SC}}}}{\Gamma(Nm_{\mathrm{SC}})} \int_0^z z^{Nm_{\mathrm{SC}}-1} e^{-a_{\mathrm{SC}}z}\mathrm{d}z, \tag{2.8}$$

where $\gamma(s, x)$ is the lower incomplete Gamma function Hence, we obtain

$$\begin{aligned} F_{\gamma_{\mathrm{SC}}}(z) &= \frac{1}{\Gamma(Nm_{\mathrm{SC}})}\gamma(Nm_{\mathrm{SC}}; a_{\mathrm{SC}}z) \\ &= 1 - e^{-a_{\mathrm{SC}}z} \sum_{k=0}^{Nm_{\mathrm{SC}}-1} \frac{(a_{\mathrm{SC}}z)^k}{k!}, z > 0. \end{aligned} \tag{2.9}$$

Lemma 2. *The CDF of* Γ_{SC} *can be expressed as*

$$F_{\Gamma_{\mathrm{SC}}}(\gamma) = 1 - e^{-a_{\mathrm{SC}}\gamma} \sum_{m=0}^{N_{\mathrm{SC}}-1} \frac{(a_{\mathrm{SC}}\gamma)^m}{m!} \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \sum_{r=0}^{m} \binom{m}{r} \frac{(j)_r\, a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}^{j}}{\left(a_{\mathrm{SC}}\gamma + a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}\right)^{r+j}}, \quad \gamma > 0. \tag{2.10}$$

Proof: The CDF of γ_{SC}, defined as $F_{\Gamma_{\mathrm{SC}}}(\gamma) = \Pr\left\{\frac{\gamma_{\mathrm{SC}}}{\mathrm{I_C}+1} < \gamma\right\}$, is written as

$$F_{\Gamma_{\mathrm{SC}}}(\gamma) = 1 – e^{-a_{\mathrm{SC}}\gamma} \sum_{m=0}^{Nm_{\mathrm{SC}}-1} \frac{(a_{\mathrm{SC}}\gamma)^m}{m!} \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A})$$
$$\times \frac{a^{j}_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}}{(j-1)!} \int_0^\infty x^{j-1}\,(x+1)^m\, e^{-(a_{\mathrm{SC}}\gamma + a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}})x}\mathrm{d}x. \tag{2.11}$$

Using the definition of the Confluent hypergeometric function of the second kind, where $U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt}\, t^{a-1}(1+t)^{\,b-a-1}\,\mathrm{d}t$ and with the fact that $U(a, a+b+1, z) = \sum_{r_k=0}^{b}\binom{b}{r_k}(a_{r_k})(z)^{-r_k-a}$, we obtain (2.10). This completes the proof of Lemma 2. ■

2.2.2 C-to-D_k Communication

The received signal at D_k from C is written as

$$y_{\mathrm{CD_k}} = \sqrt{\ell_{\mathrm{CD}_k}}h_{\mathrm{CD_k}}\sqrt{P_{\mathrm{C}}}x_{\mathrm{C}} + \sum_{i\in\mathcal{I}}\sqrt{\ell_{\mathrm{I}_i\mathrm{D}_k}}h_{\mathrm{I}_i\mathrm{D}_k}\sqrt{P_{\mathrm{I}_i}}x_{\mathrm{I}_i} + n_{\mathrm{D}_k}, \tag{2.12}$$

where x_{C} is the normalized short packet signal dedicated to D_k and n_{D_k} is the complex AWGN with zero-mean and variance σ^2. $h_{\mathrm{I}_i\mathrm{D}_k}$ and $\ell_{\mathrm{I}_i\mathrm{D}_k}$ denote the complex small-scale fading and the path loss from interferer $i \in \mathcal{I}$ to D_k, respectively. Additionally, we assume that $|h_{\mathrm{I}_i\mathrm{D}_k}|^2$s for $k \in \mathcal{K}$, where $\mathcal{K} = 1, 2, \ldots, K$, are Non-IID Gamma random variables, each with unit mean and corresponding shape parameter $m_{\mathrm{I}_i\mathrm{D}_k}$.

2.2.2.1 Channel Model of the C-to-D_k links

The path loss from C to each D_k, denoted as ℓ_{CD_k} for $k \in \mathcal{K}$, are not necessarily identical and follow the model in (2.2). In other words, $\ell_{\mathrm{CD}_k} = G_{\mathrm{C}} + G_{\mathrm{D}_k} - 22.7 - 26\log_{10}(f_c) - 36.7\log_{10}(\frac{d_{\mathrm{CD}_k}}{d_0})$ [dB], where d_{CD_k} is the distance from C to D_k. Moreover, the SINR of the second phase at each user $k \in \mathcal{K}$, can be written as

$$\Gamma_{\mathrm{CD}_k} = \frac{\bar{\gamma}_{\mathrm{C}}\ell_{\mathrm{CD}_k}|h_{\mathrm{CD}_k}|^2}{\sum_{i\in\mathrm{I}}\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{D}_k}|h_{\mathrm{I}_i\mathrm{D}_k}|^2 + 1} = \frac{\gamma_{\mathrm{CD}_k}}{\mathrm{I}_{\mathrm{D}_k} + 1}. \tag{2.13}$$

where $\bar{\gamma}_{\mathrm{C}} \triangleq P_{\mathrm{C}}/\sigma^2$, $\gamma_{\mathrm{CD}_k} \triangleq \bar{\gamma}_{\mathrm{C}}\ell_{\mathrm{CD}_k}|h_{\mathrm{CD_k}}|^2$ and $\mathrm{I}_{\mathrm{D}_k} \triangleq \sum_{i\in\mathcal{I}} \bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{D}_k}|h_{\mathrm{I}_i\mathrm{D}_k}|^2$ are the instantaneous SNR and INR at each D_k, where $k \in \mathcal{K}$, respectively.

In this work, we primarily focus on the e2e error performance, as determined via the user with the lowest e2e SINR. In other words, the e2e SINR used for evaluated the error performance is formulated as

$$\Gamma_{\mathrm{sys}} = \min_{k\in\mathcal{K}} \left\{ \min\{\Gamma_{\mathrm{SC}}, \Gamma_{\mathrm{CD}_k}\} \right\} = \min \left\{ \Gamma_{\mathrm{SC}}, \min_{k\in\mathcal{K}}\{\Gamma_{\mathrm{CD_k}}\} \right\}. \tag{2.14}$$

Lemma 3. *The PDF and CDF of the aggregated interference power at* D_k*, denoted as* $f_{\mathrm{I}_{\mathrm{D}_k}}(x)$ *and* $F_{\mathrm{I}_{\mathrm{D}_k}}(x)$ *for* $x > 0$*, respectively, are obtained as*

$$f_{\mathrm{I}_{\mathrm{D}_k}}(x) = \sum_{i=1}^{\varrho(\mathcal{B}_k)} \sum_{j=1}^{\tau_i(\mathcal{B}_k)} \chi_{i,j}(\mathcal{B}_k) \frac{a_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}^{j}}{(j-1)!} x^{j-1} e^{-x a_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}}, \tag{2.15}$$

$$F_{\mathrm{I}_{\mathrm{D}_k}}(x) = 1 - \sum_{i=1}^{\varrho(\mathcal{B}_k)} \sum_{j=1}^{\tau_i(\mathcal{B}_k)} \chi_{i,j}(\mathcal{B}_k) e^{-\frac{x}{\mu_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}}} \sum_{r=0}^{j-1} \frac{(x a_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k})^r}{r!}, \tag{2.16}$$

where $\mathcal{B}_k = \mathrm{diag}([\underbrace{a_{\mathrm{I}_1\mathrm{D}_k}, \ldots, a_{\mathrm{I}_1\mathrm{D}_k}}_{m_{\mathrm{I}_1\mathrm{D}_k}\ \textit{times}}, \ldots, \underbrace{a_{\mathrm{I}_L\mathrm{D}_k}, \ldots, a_{\mathrm{I}_L\mathrm{D}_k}}_{m_{\mathrm{I}_L\mathrm{D}_k}\ \textit{times}}])$*, are the distinct diagonal elements in decreasing order,* $\tau_i(\mathcal{B}_k)$ *is the multiplicity of* $a_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}$*, and* $\chi_{i,j}(\mathcal{B}_k)$ *is the* $(i, j)^{\mathrm{th}}$ *characteristic coefficient of* $\mathcal{B}_k$*.*

Proof: The proof of Lemma 3 is given in Appendix II. ■

Lemma 4. *The CDF of* Γ_{CD_k} *can be expressed as*

$$\begin{aligned} F_{\Gamma_{\mathrm{CD}_k}}(\gamma) = 1 - e^{-a_{\mathrm{CD}_k}\gamma} \sum_{m_k=0}^{m_{\mathrm{CD}_k}-1} \frac{(a_{\mathrm{CD}_k}\gamma)^{m_k}}{m_k!} \sum_{i_k=1}^{\varrho(\mathcal{B}_k)} \sum_{j_k=1}^{\tau_{i_k}(\mathcal{B}_k)} \\ \times \chi_{i_k,j_k}(\mathcal{B}_k) \sum_{r_k=0}^{m_k} \binom{m_k}{r_k} \frac{a_{\mathrm{I}_{\langle i_k\rangle}\mathrm{D}_k}^{j_k} (j_k)_{r_k}}{(a_{\mathrm{CD}_k}\gamma + a_{\mathrm{I}_{\langle i_k\rangle}\mathrm{D}_k})^{r_k+j_k}}, \ \gamma > 0. \end{aligned} \tag{2.17}$$

Proof: The CDF of γ_{CD_k}, defined as $F_{\Gamma_{\mathrm{CD}_k}}(\gamma) = \Pr\left\{\frac{\gamma_{\mathrm{CD}_k}}{\mathrm{I}_{\mathrm{D}_k}+1} < \gamma\right\}$, is written as

$$\begin{aligned} F_{\Gamma_{\mathrm{CD}_k}}(\gamma) &= 1 - e^{-a_{\mathrm{CD}_k}\gamma} \sum_{m=0}^{Nm_{\mathrm{CD}_k}-1} \frac{\left(a_{\mathrm{CD}_k}\gamma\right)^m}{m!} \sum_{i=1}^{\varrho(\mathcal{B}_{\boldsymbol{k}})} \sum_{j=1}^{\tau_i(\mathcal{B}_{\boldsymbol{k}})} \\ &\times \chi_{i,j}(\mathcal{B}_{\boldsymbol{k}}) \frac{a^j_{\mathrm{I}_{\langle i \rangle}\mathrm{C}}}{(j-1)!} \int_0^\infty x^{j-1} (x+1)^m \times e^{-(a_{\mathrm{CD}_k}\gamma + a_{\mathrm{I}_{\langle i \rangle}\mathrm{C}})x} \mathrm{d}x. \end{aligned} \tag{2.18}$$

Using the Confluent hypergeometric function of the second kind, we obtain (2.17). This completes the proof of Lemma 4. ■

Lemma 5. *The CDF of* Γ_{sys} *can be written as*

$$F_{\Gamma_{sys}}(\gamma) = 1 - \left[1 - F_{\Gamma_{\mathrm{SC}}}(\gamma)\right] \prod_{k \in \mathcal{K}} \left[1 - F_{\Gamma_{\mathrm{CD}_k}}(\gamma)\right], \ \gamma > 0. \tag{2.19}$$

Proof: The SINR of the system is defined as the minimum SINR obtained by all users

$$\Gamma_{\mathrm{sys}} = \min\left\{\Gamma_{\mathrm{SC}}, \min_{k \in \mathcal{K}}\{\Gamma_{\mathrm{CD_k}}\}\right\}, \tag{2.20}$$

where $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$. Hence, the CDF of Γ_{sys}, defined as $F_{\Gamma_{\mathrm{sys}}}(\gamma) = \Pr\left\{\Gamma_{\mathrm{sys}} < \gamma\right\}$, is obtained as

$$\begin{aligned} F_{\Gamma_{\mathrm{sys}}}(\gamma) &= 1 - \Pr\left\{\Gamma_{\mathrm{SC}} > \gamma, \bigcap_{k \in \mathcal{K}} \left[\Gamma_{\mathrm{CD}_k} > \gamma\right]\right\} \\ &= 1 - \Pr\left\{\Gamma_{\mathrm{SC}} > \gamma\right\} \prod_{k \in \mathcal{K}} \Pr\left\{\Gamma_{\mathrm{CD}_k} > \gamma\right\}, \end{aligned} \tag{2.21}$$

which eventually yields (2.19). Note that Γ_{SC} is a function of h_{SC} and Γ_{CD_k} is a function of h_{CD_k} and $h_{\mathrm{I}_i\mathrm{D}_k}$, $k = 1, \ldots, K$ and $i = 1, \ldots, I$. Since h_{SC}, h_{CD_k}, and $h_{\mathrm{I}_i\mathrm{D}_k}$, for $k = 1, \ldots, K$ and $i = 1, \ldots, I$, are statistically independent, specifically non-IID, we conclude that Γ_{SC} and Γ_{CD_k}, $k = 1, \ldots, K$, are also statistically independent. Hence, the intersection (joint event) in the second equality in Eq. (21) occurs due to the property $\Pr\{E_1 E_2 \ldots\} = \prod_i \Pr\{E_i\}$ when E_i are mutually independent random events. This completes the proof of Lemma 5. ■

2.3 Performance Analysis

To reduce the physical-layer transmission latency, URLLC often uses short-packet communications. The best design for short-packet communications is based on a detailed examination of the relationship between the BLER, maximum attainable rate (MAR), and blocklength. Given a blocklength of size M symbol/block, the MAR $R(M, \gamma, \epsilon)$ (in bits per channel usage (BPCU)) is

$$R(M, \gamma, \epsilon) = C(\gamma) - \sqrt{\frac{V(\gamma)}{M}} Q^{-1}(\epsilon) + O\left(\frac{\log_2 M}{M}\right), \tag{2.22}$$

where $C(\gamma) = \log_2(1+\gamma)$ is the Shannon capacity, $V(\gamma) = (1 - \frac{1}{(1+\gamma)^2})(\log_2 e)^2$ is the channel dispersion, $Q^{-1}(x)$ is the Gaussian inverse Q-function, and $O\left(\frac{\log_2 M}{M}\right)$ is the remainder term which can be omitted when $M \geq 100$ Hence, the MAR is expressed as $R = \frac{F}{M}$, where F denotes the number of data bits for D_k.

We are using a half duplex relaying (HDR) decode-and-forward (DF) system. Each transmission block in HDR is divided into two equal-length time slots. The source transmits information to C during the first time slot, then C decodes and forwards the received signal to the destination in the second time slot. Loop interference at C is fully avoided in HDR.

2.3.1 BLER Analysis

In this subsection, we study the average e2e BLER. Using DF relaying at C, errors can arise from two possible scenarios: (i) C identifies an error in the received data and successfully corrects it before relaying the data to the destination, (ii) C correctly relays the data, but the destination still detects errors in the received data.

Consider a transmission from S to C, let ϕ_{SC} and ϕ_{CD_k} represent the events related to the transmission process between S-to-C and C-to-D_k, respectively. The instantaneous BLER at C can be written as

$$\Pr(\phi_{\mathrm{SC}}) = \tilde{\epsilon}_{\mathrm{SC}} \approx Q\left(\frac{C(\Gamma_{\mathrm{SC}}) - R}{\sqrt{V(\Gamma_{\mathrm{SC}})/M}}\right). \tag{2.23}$$

Provided that C successfully decodes and forwards that signal to D$_k$, the the instantaneous BLER at D$_k$ $\tilde{\epsilon}_{\mathrm{CD}_k|\mathrm{Csucceeds}}$ is written as

$$\Pr(\phi_{\mathrm{CD}_k}|\bar{\phi}_{\mathrm{SC}}) = \tilde{\epsilon}_{\mathrm{CD}_k|\mathrm{SC\ succeeds}} \approx Q\left(\frac{C(\Gamma_{\mathrm{CD}_k}) - R}{\sqrt{V(\Gamma_{\mathrm{CD}_k})/M}}\right), \tag{2.24}$$

where $\Pr(\phi_{\mathrm{CD}_k}|\phi_{\mathrm{SC}})$ is the conditional probability of ϕ_{CD_k} given ϕ_{SC}. It is noted that D$_k$'s signal i.e., x_{CD_k} can be received successfully at D$_k$ when the signal at C i.e., x_{SC} is decoded correctly. Meanwhile, due to the high level of interference, x_{CD_k} is decoded incorrectly when x_{SC} is erroneously decoded. As a result, it is assumed that $\Pr(\phi_{\mathrm{CD}_k}|\phi_{\mathrm{SC}})$ is equal to one Finally, since $\Pr(\phi_{\mathrm{CD}_k}) = \Pr(\phi_{\mathrm{CD}_k}|\phi_{\mathrm{SC}})\Pr(\phi_{\mathrm{SC}}) + \Pr(\phi_{\mathrm{CD}_k}|\bar{\phi}_{\mathrm{SC}})\Pr(\bar{\phi}_{\mathrm{SC}})$, the e2e instantaneous BLER ($\tilde{\epsilon}_{e2e}$), where $\tilde{\epsilon}_{\mathrm{CD|SC\ succeeds}} = \tilde{\epsilon}_{\mathrm{CD}_k}$ is written as

$$\begin{aligned} \tilde{\epsilon}_{\mathrm{e2e}} &= Q\left(\frac{C(\Gamma_{\mathrm{e2e}}) - R}{\sqrt{V(\Gamma_{\mathrm{e2e}})/M}}\right) && (2.25) \\ &\approx 1 \times \tilde{\epsilon}_{\mathrm{SC}} + \tilde{\epsilon}_{\mathrm{CD}_k} \times (1 - \tilde{\epsilon}_{\mathrm{SC}}) = \tilde{\epsilon}_{\mathrm{SC}} + (1 - \tilde{\epsilon}_{\mathrm{SC}})\tilde{\epsilon}_{\mathrm{CD}_k}. && (2.26) \end{aligned}$$

2.3.2 Average BLER of each transmission (each hop)

For SPC, the average BLER can be approximated as

$$\bar{\epsilon}_{\chi} \approx \mathbb{E}\left[Q\left(\frac{C(\gamma_{\chi}) - R}{\sqrt{V(\gamma_{\chi})/M}}\right)\right], \tag{2.27}$$

where $\chi \in \{\mathrm{SC}, \mathrm{CD}, \mathrm{e2e}\}$, $\mathbb{E}[\cdot]$ is the expectation operator, and $Q(\cdot)$ denotes the Gaussian Q-function. It is intractable to derive analytically closed-form expression of (2.27) due to the Gaussian Q-function. To approximate the expression, the estimated linear form of $Q\left(\frac{C(\gamma_{\chi})-R}{\sqrt{V(\gamma_{\chi})/M}}\right)$ is invoked as

$$Z(\gamma_{\chi}) = \frac{1}{2} - \tau\left(\gamma_{\chi} - \omega\right), \quad \psi < \gamma_{\chi} < \eta, \tag{2.28}$$

where $\tau \triangleq \frac{\sqrt{M}}{\sqrt{2\pi(4^R-1)}}$, $\eta \triangleq \omega + \frac{1}{2\tau}$, and $\omega \triangleq 2^R - 1$. Hence, substituting (2.28) in (2.27), the average BLER is tightly approximated as

$$\bar{\epsilon}_{\chi} \approx \int_{0}^{\infty} Z(\gamma_{\chi}) f_{\gamma_{\chi}}(x)dx = \tau \int_{\psi}^{\eta} F_{\gamma_{\chi}}(x)dx, \tag{2.29}$$

where $F_{\gamma_{\chi}}(x)$ is the CDF of γ_{χ}, and second equality is due to the partial integration method.

2.3.2.1 Average BLER of the first hope (S-to-C)

We derive the closed-form expression for the average BLER at C, i.e., $\bar{\epsilon}_{\mathrm{SC}}$, in Theorem 1.

Theorem 1. *Given a number of transmit data bits, the average BLER* $\bar{\epsilon}_{\mathrm{SC}}$ *at* C *can be written in closed-form expression as*

$$\bar{\epsilon}_{\mathrm{SC}} = 1 - \tau \sum_{m=0}^{Nm_{\mathrm{SC}}-1} \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \sum_{r=0}^{m} \frac{a_{\mathrm{I}_{\langle i \rangle}\mathrm{C}}^{j} a_{\mathrm{SC}}^{r-1}}{r!} \Big[\Theta(r,\psi) - \Theta(r,\eta)\Big], \tag{2.30}$$

where $\Theta(r,\psi) \triangleq \psi^r e^{-a_{\mathrm{SC}}\psi} U\left(j, r+j, a_{\mathrm{I}_{\langle i \rangle}\mathrm{C}} + a_{\mathrm{SC}}\psi\right)$.

Proof: The proof of Theorem 1 is given in Appendix III. ■

Corollary 1. *Let us assume an interference-free scenario, where the number of interferers is zero, then we can write the average BLER without interference at* C *i.e.,* $\hat{\epsilon}_{\mathrm{SC}}$, *as*

$$\hat{\epsilon}_{\mathrm{SC}} = 1 - \tau \sum_{k=0}^{Nm_{\mathrm{SC}}-1} \sum_{q=0}^{k} \frac{(a_{\mathrm{SC}})^{q-1}}{q!} (e^{-a_{\mathrm{SC}}\psi}\psi^q - e^{-a_{\mathrm{SC}}\eta}\eta^q). \tag{2.31}$$

Proof: The proof of Corollary 1 is obtained following the same approach in Appendix III by taking zero interferers. ■

2.3.2.2 Average BLER of the second hope for each individual user k (C-to-D$_k$)

In the C-to-D$_k$ (Cluster head to k^{th} user) links, the small scale fading follows IID Nakagami-m distributions and h_{CD_k} follows IID Gamma distributions. We derive the closed-form expression of the average BLER at D$_k$, i.e., $\bar{\epsilon}_{\mathrm{CD}_k}$, in Theorem 2.

Theorem 2. *The average BLER at* D$_k$*, denoted as* $\bar{\epsilon}_{\mathrm{CD}_k}$*, is obtained as*

$$\bar{\epsilon}_{\mathrm{CD}_k} = 1 - \tau \sum_{m_k=0}^{Nm_{\mathrm{CD}_k}-1} \sum_{i_k=1}^{\varrho(\mathcal{B}_k)} \sum_{j_k=1}^{\tau_{ik}(\mathcal{B}_k)} \chi_{i_k,j_k}(\mathcal{B}_k) \sum_{r_k=0}^{m} \frac{a_{\mathrm{I}_{\langle i_k \rangle}\mathrm{C}}^{j_k} (a_{\mathrm{CD}_k})^{r_k-1}}{r_k!} \Big[\Theta_1(r_k,\psi) - \Theta_1(r_k,\eta)\Big], \tag{2.32}$$

where $\Theta_1(r_k,\psi) \triangleq \psi^{r_k} e^{-a_{\mathrm{CD}_k}\psi} U(j_k, r_k + j_k, a_{\mathrm{I}_{\langle i_k \rangle}\mathrm{C}} + a_{\mathrm{CD}_k}\psi)$.

Proof: The proof of Theorem 2 is given in Appendix IV. ■

Corollary 2. *Let us assume an interference-free scenario, where the number of interferers is zero, then we can write the average BLER at* D$_k$ *i.e.,* $\hat{\epsilon}_{\mathrm{CD}_k}$*, as*

$$\hat{\epsilon}_{\mathrm{CD}_k} = 1 - \tau \sum_{k=0}^{Nm_{\mathrm{CD}_k}-1} \sum_{q=0}^{k} (a_{\mathrm{CD}_k})^{q-1} \frac{e^{-a_{\mathrm{CD}_k}\psi}\psi^q - e^{-a_{\mathrm{CD}_k}\eta}\eta^q}{q!}. \tag{2.33}$$

Proof: The proof of Corollary 2 is obtained following the same approach in Appendix IV by taking zero interferers. ■

2.3.2.3 Average e2e BLER

The average e2e BLER of the system is given in Theorem 3.

Theorem 3. *The average e2e BLER* $\bar{\epsilon}_{\mathrm{e2e}}$ *for the system can be written as*

$$\bar{\epsilon}_{\mathrm{e2e}} \approx 1 - \sum_{m=0}^{Nm_{\mathrm{SC}}-1} \sum_{m_1=0}^{m_{\mathrm{CD}_1}-1} \cdots \sum_{m_k=0}^{m_{\mathrm{CD}_k}-1} \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{i_1=1}^{\varrho(\mathcal{B}_1)} \cdots \sum_{i_k=1}^{\varrho(\mathcal{B}_k)} \\ \times \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \sum_{j_1=1}^{\tau_1(\mathcal{B}_1)} \cdots \sum_{j_k=1}^{\tau_k(\mathcal{B}_k)} \left\{ \prod_{k=1}^{K} \chi_{i_k,j_k}(\mathcal{B}_k) \right\} \Lambda_1, \tag{2.34}$$

$$\Lambda_1 = \sum_{r=0}^{m} \binom{m}{r} \frac{(j+r-1)!}{(j-1)!} \frac{a_{\mathrm{I}\langle i\rangle\mathrm{C}}^{j} a_{\mathrm{SC}}^{m}}{m!} \sum_{r_1=0}^{m_1} \cdots \sum_{r_k=0}^{m_k} (\Lambda_2 + \Lambda_3) \prod_{k=1}^{K} \binom{m_k}{r_k} \frac{(j_k + r_k - 1)!}{(j_k - 1)!} \frac{a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}^{j_k} a_{\mathrm{CD}_k}^{m_k}}{m_k!}, \tag{2.35}$$

$$\Lambda_2 = \sum_{k=1}^{K} \sum_{l_k=1}^{r_k+j_k} \frac{\chi_{k,l_k}}{a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}^{l_k}} \left\{ \Lambda\left(\eta, m+m_\Sigma, a_\Sigma, \frac{a_{\mathrm{CD}_k}}{a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}}\eta\right) - \Lambda\left(\psi, m+m_\Sigma, a_\Sigma, \frac{\psi a_{\mathrm{CD}_k}}{a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}}\right) \right\}, \tag{2.36}$$

$$\chi_{k,l_k} = \frac{\left(\frac{a_{\mathrm{CD}_k}}{a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}}\right)^{r_k+j_k-l_k}}{(r_k+j_k-l_k)!} \left[\frac{\mathrm{d}^{r_k+j_k-l_k}}{\mathrm{d}s^{r_k+j_k-l_k}} \left(\frac{a_{\mathrm{SC}}}{a_{\mathrm{I}\langle i\rangle\mathrm{C}}}\gamma + 1 \right)^{r_k+j_k-l_k} \right] \Bigg|_{s\to -a_{\mathrm{CD}_k}/a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}}, \tag{2.37}$$

$$\Lambda_3 = \sum_{l=1}^{r+j} \frac{\chi_l}{a_{\mathrm{I}\langle i\rangle\mathrm{C}}^{l}} \left\{ \Lambda\left(\eta, m+m_\Sigma, a_\Sigma, \frac{\eta a_{\mathrm{SC}}}{a_{\mathrm{I}\langle i\rangle\mathrm{C}}}\right) - \Lambda\left(\psi, m+m_\Sigma, a_\Sigma, \frac{\psi a_{\mathrm{SC}}}{a_{\mathrm{I}\langle i\rangle\mathrm{C}}}\right) \right\}, \tag{2.38}$$

$$\chi_l = \frac{\left(\frac{a_{\mathrm{SC}}}{a_{\mathrm{I}\langle i\rangle\mathrm{C}}}\right)^{r+j-l}}{(r+j-l)!} \left[\frac{\mathrm{d}^{r+j-l}}{\mathrm{d}s^{r+j-l}} \left(\prod_{k=1}^{K} (a_{\mathrm{CD}_k}\gamma + a_{\mathrm{I}\langle i_k\rangle\mathrm{D}_k}) \right)^{r_k+j_k} \right] \Bigg|_{s\to -a_{\mathrm{SC}}/a_{\mathrm{I}\langle i\rangle\mathrm{C}}}. \tag{2.39}$$

---

where $\Lambda_1, \Lambda_2, \chi_{k,l_k}, \Lambda_3$, *and* χ_l *are given in* (2.35), (2.36), (2.37), (2.38), *and* (2.39), *respectively. Herein,* $m_\Sigma = \sum_{k=1}^{K} m_k$ *and* $a_\Sigma = a_{\mathrm{SC}} + \sum_{k=1}^{K} a_{\mathrm{CD}_k}$.

Proof: The proof of Theorem 3 is given in Appendix V. ■

2.3.2.4 Approximation for $\bar{\epsilon}_{\mathrm{e2e}}$

The $\bar{\epsilon}_{\mathrm{e2e}}$ can be approximated as follow

$$\begin{aligned} \bar{\epsilon}_{\mathrm{e2e}} &\approx \bar{\epsilon}_{\mathrm{SC}} + (1-\bar{\epsilon}_{\mathrm{SC}}) \left(1 - \prod_{k=1}^{K} (1-\bar{\epsilon}_{\mathrm{CD}_k})\right) \\ &= 1 - (1-\bar{\epsilon}_{\mathrm{SC}}) \prod_{k=1}^{K} (1-\bar{\epsilon}_{\mathrm{CD}_k}). \end{aligned} \tag{2.40}$$

2.3.3 Asymptotic, Error Floor, and Diversity Gain Analyses

2.3.3.1 Average BLER of the First Phase

In the high transmit SINR regime, where $\bar{\gamma}_{\mathrm{S}}$ is relatively large, we have $F_{\gamma_{\mathrm{SC}}(\gamma)} \rightarrow (a_{\mathrm{SC}}\gamma)^{Nm_{\mathrm{SC}}}/(Nm_{\mathrm{SC}})!$, for $\gamma > 0$. Hence, the CDF of the SINR at C can be simplified as

$$F_{\Gamma_{\mathrm{SC}}}(\gamma) \rightarrow \alpha_{\mathrm{SC}} \frac{(a_{\mathrm{SC}}\gamma)^{Nm_{\mathrm{SC}}}}{(Nm_{\mathrm{SC}}+1)!}, \quad \gamma > 0, \tag{2.41}$$

where $\alpha_{\mathrm{SC}} \triangleq \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) a^{j}_{\mathrm{I}_{\langle i \rangle}\mathrm{C}} U(j, Nm_{\mathrm{SC}} + j + 1, a_{\mathrm{I}_{\langle i_k \rangle}\mathrm{C}})$.

As a result, the average BLER at C in the high SINR regime, denoted as $\bar{\epsilon}^{\infty}_{\mathrm{SC}}(P_{\mathrm{S}})$, is obtained as

$$\bar{\epsilon}^{\infty}_{\mathrm{SC}}(P_{\mathrm{S}}) = \frac{\tau \alpha_{\mathrm{SC}} (a_{\mathrm{SC}})^{Nm_{\mathrm{SC}}}}{(Nm_{\mathrm{SC}}+1)!} (\eta^{Nm_{\mathrm{SC}}+1} - \psi^{Nm_{\mathrm{SC}}+1}). \tag{2.42}$$

2.3.3.2 Average BLER of the Second Phase

In the high transmit SINR regime, where $\bar{\gamma}_{\mathrm{CD}_k}$ is relatively large, we have $F_{\gamma_{\mathrm{CD}_k}(\gamma)} \rightarrow (a_{\mathrm{CD}_k}\gamma)^{m_{\mathrm{CD}_k}}/(Nm_{\mathrm{CD}_k})!$, for $\gamma > 0$, Hence, the CDF of the SINR at the cluster head can be simplified as

$$F_{\Gamma_{\mathrm{CD}_k}}(\gamma) \rightarrow \beta_{\mathrm{CD}_k} \frac{(a_{\mathrm{CD}_k}\gamma)^{m_{\mathrm{CD}_k}}}{(m_{\mathrm{CD}_k}+1)!}, \quad \gamma > 0, \tag{2.43}$$

where $\beta_{\mathrm{CD}_k} \triangleq \sum_{i=1}^{\varrho(\mathcal{B})} \sum_{j=1}^{\tau_i(\mathcal{B})} \chi_{i,j}(\mathcal{B}_k) \frac{(j+m_{\mathrm{CD}_k}-1)!}{(j-1)!} (a_{\mathrm{I}_{\langle i \rangle}\mathrm{D}_k})^{-m_{\mathrm{CD}_k}}$.

Hence, the average BLER at D$_k$ in the high SINR regime, denoted as $\bar{\epsilon}^{\infty}_{\mathrm{CD}_k}(P_{\mathrm{C}})$, is obtained as

$$\bar{\epsilon}^{\infty}_{\mathrm{CD}_k}(P_{\mathrm{C}}) = \frac{\tau \beta_{\mathrm{CD}_k}}{(m_{\mathrm{CD}_k}+1)!} (a_{\mathrm{CD}_k})^{m_{\mathrm{CD}_k}} (\eta^{m_{\mathrm{CD}_k}+1} - \psi^{m_{\mathrm{CD}_k}+1}). \tag{2.44}$$

2.3.3.3 average e2e BLER

Accordingly, the average BLER of the system in the high SINR regime, denoted as $\bar{\epsilon}_{\text{e2e}}^{\infty}$, is given by

$$\bar{\epsilon}_{\text{e2e}}^{\infty} = 1 - (1 - \bar{\epsilon}_{\text{SC}}^{\infty}(P_{\text{S}})) \prod_{k=1}^{K} (1 - \bar{\epsilon}_{\text{CD}_k}^{\infty}(P_{\text{C}})) \tag{2.45}$$

$$= 1 - \left\{ 1 - \frac{\tau \alpha_{\text{SC}}}{(N m_{\text{SC}} + 1)!} (a_{\text{SC}})^{N m_{\text{SC}}} (\eta^{N m_{\text{SC}}+1} - \psi^{N m_{\text{SC}}+1}) \right\} \times \prod_{k=1}^{K} \left\{ 1 - \frac{\tau \beta_{\text{CD}_k}}{(m_{\text{CD}_k} + 1)!} (a_{\text{CD}_k})^{m_{\text{CD}_k}} (\eta^{m_{\text{CD}_k}+1} - \psi^{m_{\text{CD}_k}+1}) \right\}. \tag{2.46}$$

2.3.3.4 Error Floor Analysis

In this subsection, we determine the error floor of the average e2e BLER. The error floor analysis provides useful system design guidelines. There is a region in which performance flattens is called the error floor region. We consider the error floor to be the lower limit achieved by the average e2e BLER when either P_{S} or P_{C} tends towards infinity while the counterpart (P_{C} or P_{S}, respectively) remains within a low-to-middle transmission regime. The error floor of the average e2e BLER when only P_{S} or $P_{\text{C}} \to \infty$ is obtained as

$$\bar{\epsilon}_{\text{e2e}}^{\infty} \to \begin{cases} \prod_{k=1}^{K} \bar{\epsilon}_{\text{CD}_k}, & P_{\text{S}} \to \infty, \\ \bar{\epsilon}_{\text{SC}}, & P_{\text{C}} \to \infty. \end{cases} \tag{2.47}$$

From (2.47), we draw the insights that by increasing the transmission power of the BS or CH alone does not lead to further improvement in the BLER. This is because the end-to-end instantaneous BLER is determined by the minimum BLER equation, which implies that the system's performance is limited by the weakest link, such as the user with the worst channel conditions.

2.3.3.5 Diversity Gain Analysis

Unlike the error floor analysis, the diversity gain is calculated at $P_\mathrm{S} = P_\mathrm{C} \triangleq P \longrightarrow \infty$. In other words, the diversity gain of the considering URLLC system can be defined as

$$\begin{aligned} D_\mathrm{e2e} &= -\lim_{P} \frac{\log(\bar{\epsilon}_\mathrm{e2e}(P))}{\log(P)} = -\lim_{P} \frac{\log(\bar{\epsilon}^{\infty}_\mathrm{e2e}(P))}{\log(P)} \\ &= \min\{N m_\mathrm{SC}, m_{\mathrm{CD}_1}, \ldots, m_{\mathrm{CD}_K}\}, \end{aligned} \tag{2.48}$$

where $\bar{\epsilon}^{\infty}_\mathrm{e2e}(P) = 1-(1-\bar{\epsilon}^{\infty}_\mathrm{SC}(P))\prod_{k=1}^{K}(1-\bar{\epsilon}^{\infty}_{\mathrm{CD}_k}(P))$ is merely the average e2e BLER expressed as a function of P.

Remark 1. *Specifically, when setting the transmit powers of the source (P_S) and the cluster head (P_C) equal ($P_S = P_C = P$), the average end-to-end BLER ($\bar{\epsilon}^{\infty}_{e2e}(P)$) can be modeled as a summation of terms $C_i P^{-b_i}$, where C_i and b_i are positive constants with respect to P. In this scenario, the diversity order of the system (D_{e2e}) is determined by the minimum value of b_i across all terms, denoted as $\min_i\{b_i\}$. As a result, when both BS and CH transmission powers are increased by r [dB] in the high transmit SINR regime, the average end-to-end BLER decreases by $r \min_i\{b_i\}$ [dB]. This relationship highlights the beneficial impact of increasing transmit powers on improving system performance, particularly in environments characterized by high SNR levels.*

With such a communication network, locations of the S, C, and users are treated as random variables (RVs), which make the performance analysis intractable. Thus, relying on the data driven techniques to develop a CNN model based on transfer learning that can be trained to predict the average overall BLER using pre-collected CSI and associated outage probability data. The multivariate confluent hypergeometric function having complex BLER expression renders the calculation of the system average overall BLER specified in (2.40) unfeasible for real-time application, as distances between nodes in the proposed system are now also RVs. We approach the challenge of determining the system average overall BLER as a non-linear regression problem in order to get over this hurdle. Specifically, we produce a data set that thoroughly describes the considered system. The created CNN can predict the average overall BLER accurately under different system settings.

To deal with unpredictable performance at ground users, S and C can leverage a CNN to predict the user performance and adjust parameters to satisfy the fairness and quality of service at each user. Motivated by the aforementioned performance analysis, this work offers a methodology for assessing system performance and then proceeds to evaluate performance prediction through transfer learning in next chapter.

CHAPTER 3

TRANSFER LEARNING BASED PERFORMANCE PREDICTION

In this work, we leverage ML for real-time performance prediction. We specifically use transfer learning, employing a pre-trained model initially trained on a larger source domain, to handle a target task on a limited dataset. This method applies the source model's knowledge to smaller datasets, facilitating efficient real-time predictions.

3.1 Universal Approximation Function for Performance Prediction

The average e2e BLER can be mathematically represented as a function $\Upsilon : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ represent the input and output spaces, respectively. Our goal is to develop a transfer learning-based universal approximation function $\widetilde{\Upsilon} : \mathcal{X} \to \hat{\mathcal{Y}}$, where $\hat{\mathcal{Y}}$ is an accurate approximation of $\mathcal{Y}$ based on a limited dataset that does not fully capture the input-output relationship in Υ. Hence, the developed ML models utilize regression analysis to accurately characterize the continuous function Υ. These models are designed to exhibit high generalization capabilities, ensuring accurate predictions even with new inputs. Here, regression analysis is a statistical technique used to estimate the relationship between a dependent variable y and an input feature vector, denoted as $\mathbf{x} = [x_1, \ldots, x_m]^\mathsf{T}$, where m denotes the number of featuresThe goal of regression analysis is to determine an objective predictive function Υ that maps the feature space $\mathcal{X}$ to the output space $\mathcal{Y}$, i.e., $\Upsilon : \mathcal{X} \to \mathcal{Y}$, such that

$$y = \Upsilon(\mathbf{x}; \boldsymbol{\theta}), \tag{3.1}$$

where $\mathbf{x} \in \mathcal{X}$, $y \in \mathcal{Y}$, and $\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_m]^\mathsf{T}$ denotes the unknown parameters with θ_0 being the bias that allows for an offset in the prediction model. In this chapter, we focus on non-linear Regression (NLR) models since the average e2e BLER is a non-linear function of input system parameters $\mathbf{x}$. In addition, using NLR models provide more flexible relationships between y and $\mathbf{x}$ using non-linear functions of the parameters and/or the independent variables. We denote the average e2e BLER as $y = \Upsilon(\mathbf{x})$ with

$$\mathbf{x} = [\mathbf{p}_{\mathsf{S}}^{\mathsf{T}}, \mathbf{p}_{\mathrm{D}_1}^{\mathsf{T}}, \ldots, \mathbf{p}_{\mathrm{D}_K}^{\mathsf{T}}, \mathbf{p}_{\mathrm{I}_1}^{\mathsf{T}}, \ldots, \mathbf{p}_{\mathrm{I}_I}^{\mathsf{T}}, \mathbf{p}_{\mathrm{C}}^{\mathsf{T}}, F, M, m_{\mathrm{SC}}, m_{\mathrm{CD}_1}, \ldots, m_{\mathrm{CD}_K}, \ell_{\mathrm{SC}}, \ell_{\mathrm{CD}_1}, \ldots, \ell_{\mathrm{CD}_K}, N]^{\mathsf{T}}, \tag{3.2}$$

where $\mathbf{p}_{\mathsf{X}} \triangleq [x_{\mathsf{X}}, y_{\mathsf{X}}, z_{\mathsf{X}}]^{\mathsf{T}}$ denotes the 3D Cartesian coordinates of node X. It is noted that the total number of input parameters $L = 11 + 5K + 3I$. As a result, predicting the average e2e BLER of a URLLC system is a complex task that involves modeling the relationship between these factors and the average e2e BLER.

3.2 Domain Adaptation-based Transfer Learning Solution

In domain adaptation-based transfer learning, a *domain* $\mathcal{D} = \{\mathcal{X}, p(\mathbf{x})\}$ consists of a feature space $\mathcal{X}$ and a marginal probability distribution $p(\mathbf{x})$. A *task* $\mathcal{T} = \{\mathcal{Y}, f\}$ consists of an output space $\mathcal{Y}$ and an objective predictive function f which is not observed but can be learned from the training data. Here, the data consist of pairs $(\mathbf{x}, y)$, where $\mathbf{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$.

In this chapter, we consider a pre-trained (source) model with one source domain $\mathcal{D}_{\mathsf{s}}$ and one source target $\mathcal{T}_{\mathsf{s}}$, and a fine-tuned (target) model with one target domain $\mathcal{D}_{\mathsf{t}}$ and one target task $\mathcal{T}_{\mathsf{s}}$. We can define $\mathcal{D}_{\mathsf{s}}$, $\mathcal{T}_{\mathsf{s}}$, $\mathcal{D}_{\mathsf{t}}$, and $\mathcal{T}_{\mathsf{t}}$ as

$$\begin{aligned} \mathcal{D}_{\mathsf{s}} &\triangleq \{\mathcal{X}_{\mathsf{s}}, p(\mathbf{x}_{\mathsf{s}})\}, \mathbf{x}_{\mathsf{s}} \in \mathcal{X}_{\mathsf{s}}, \mathcal{T}_{\mathsf{s}} \triangleq \{\mathcal{Y}_{\mathsf{s}}, f_{\mathsf{s}}\}, \\ &f_{\mathsf{s}} : \mathcal{X}_{\mathsf{s}} \to \mathcal{Y}_{\mathsf{s}}, y_{\mathsf{s}} = f_{\mathsf{s}}(\mathbf{x}_{\mathsf{s}}) \in \mathcal{Y}_{\mathsf{s}}, \end{aligned} \tag{3.3}$$

$$\begin{aligned} \mathcal{D}_{\mathsf{t}} &\triangleq \{\mathcal{X}_{\mathsf{t}}, p(\mathbf{x}_{\mathsf{t}})\}, \mathbf{x}_{\mathsf{t}} \in \mathcal{X}_{\mathsf{t}}, \quad \mathcal{T}_{\mathsf{t}} \triangleq \{\mathcal{Y}_{\mathsf{t}}, f_{\mathsf{t}}\}, \\ &f_{\mathsf{t}} : \mathcal{X}_{\mathsf{t}} \to \mathcal{Y}_{\mathsf{t}}, y_{\mathsf{t}} = f_{\mathsf{t}}(\mathbf{x}_{\mathsf{t}}) \in \mathcal{Y}_{\mathsf{t}}, \end{aligned} \tag{3.4}$$

respectively, where $\mathcal{X}_{\mathsf{s}}$, $\mathcal{Y}_{\mathsf{s}}$, $\mathcal{X}_{\mathsf{t}}$, and $\mathcal{Y}_{\mathsf{t}}$ denote the feature space of source domain, output space of source domain, feature space of target domain, and output space of target domain, respectively. Similarly, $\mathbf{x}_{\mathsf{s}}$ and $\mathbf{x}_{\mathsf{t}}$ denote the input features vectors of the source and target domains, respectively. It is noted that the source model is designed with a complex architecture and high computational costs to accurately capture the average e2e BLER represented by the function Υ. In this context, the source task is characterized as $f_{\mathtt{s}} = \Upsilon$, and the output $y_{\mathtt{s}}$ is obtained by applying Υ to the input $\mathbf{x}_{\mathtt{s}}$, where $\mathbf{x}_{\mathtt{s}}$ corresponds to the system features $\mathbf{x}$ in (3.2). However, training the source model requires a substantial dataset and extensive training time. To overcome this challenge, we aim to develop a target model with less complex architecture by fine-tuning the source model with $\mathcal{T}_{\mathtt{t}} = \mathcal{T}_{\mathtt{s}}$. By opting for a fine-tuned target model on a small dataset, benefits such as faster training, reduced data requirements, and comparable accuracy to the source model can be achieved.

In practice, exact knowledge of $p(\mathbf{x}_{\mathtt{s}})$ and $p(\mathbf{x}_{\mathtt{t}})$ is unknown; instead, these probabilities are empirically estimated using the available dataset, which can vary depending on the dataset size. When the dataset size is sufficiently large, such as in the source model, the empirical probabilities $p(\mathbf{x}_{\mathtt{s}})$ can effectively represent the true probabilities. However, in cases where the dataset is small, the resulting empirical marginal probabilities in the target model may differ due to the limited representation of the true probabilities, i.e., $\mathbf{p}(\mathbf{x}_{\mathtt{t}}) \neq \mathbf{p}(\mathbf{x}_{\mathtt{s}})$, where $\mathbf{x}_{\mathtt{t}}, \mathbf{x}_{\mathtt{s}} \in \mathcal{X}$. Considering the constraint of limited data available in the target domain, our objective is to design a target model that can be effectively trained and yield comparable accuracy to the source model.

Remark 2. *Since $\mathcal{T}_{\mathtt{s}} = \mathcal{T}_{\mathtt{t}}$, $\mathcal{X}_{\mathtt{s}} = \mathcal{X}_{\mathtt{t}}$ and $p(\mathbf{x}_{\mathtt{s}}) \neq p(\mathbf{x}_{\mathtt{t}})$, our problem is identified as a closed set domain adaptation in transductive transfer learning Specifically, some part of the dataset belong only to the source domain or to the target domain, while others belong to both datasets.*

3.3 Data

- The number of samples of our source dataset, denoted by n_{src}, is 1×10^5, while the number of samples of our target dataset, denoted by n_{T}, is 1×10^2.
- The target dataset contains a relatively smaller number of samples (10^2), as it represents a specific domain or task for which labeled data may be limited or expensive to obtain. In such scenarios, transfer learning offers a compelling advantage by leveraging knowledge acquired from the source domain to enhance learning and inference in the target domain. By

fine-tuning the pretrained model with the smaller target dataset, we effectively transfer the learned representations and knowledge from the source domain to the target domain, thereby mitigating the challenges posed by limited labeled data availability.

- The choice of the number of samples in our source and target datasets is based on several factors. The source dataset comprises a large number of samples (10^5) to ensure a comprehensive coverage of diverse data instances, thus facilitating robust learning of underlying patterns and features. This abundance of data allows our model to capture a wide range of variations and complexities present in the source domain, which enhances its ability to effectively generalize and perform across different scenarios.

3.3.1 Source Dataset

The source dataset is the dataset that is used to pre-train the original model. This dataset is typically large and diverse, and is used to learn a general feature representation of the data. All the features of our source dataset are represented in (3.2). In our case, for K users, the source dataset is in the form of a feature matrix and target vector.

3.3.2 Feature matrix

In our work, we consider the following feature matrix

$$\mathbf{X} = \left[\mathbf{x}_1^\mathsf{T}; \cdots ; \mathbf{x}_K^\mathsf{T}\right]^\mathsf{T}, \quad \mathbf{x}_k \triangleq [\mathbf{p}_\mathrm{S}^\mathsf{T}, \mathbf{p}_{\mathrm{D}_k}^\mathsf{T}, \mathbf{p}_{\mathrm{I}_1}^\mathsf{T}, \ldots, \mathbf{p}_{\mathrm{I}_I}^\mathsf{T}, \mathbf{p}_\mathrm{C}^\mathsf{T}, F, M, m_\mathrm{SC}, m_{\mathrm{CD}_k}, \ell_\mathrm{SC}, \ell_{\mathrm{CD}_k}, N]^\mathsf{T}. \tag{3.5}$$

In addition, the feature set and target set of the source dataset are written as $\mathcal{X}_\mathsf{s} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(N)}\}$ and $\mathcal{Y}_\mathsf{s} = \{\bar{\epsilon}_\mathrm{e2e}^{(1)}, \bar{\epsilon}_\mathrm{e2e}^{(2)}, \ldots, \bar{\epsilon}_\mathrm{e2e}^{(N)}\}$, respectively, where $\bar{\epsilon}_\mathrm{e2e}^{(n)} = \Upsilon(\mathbf{X}^{(n)})$ and n denotes the instance number. Hence, the structure of the dataset is $[\mathcal{X}_\mathsf{s}, \mathcal{Y}_\mathsf{s}]$, in which each data point is $[\mathbf{X}^{(n)}, \bar{\epsilon}_\mathrm{e2e}^{(n)}]$, where $\mathbf{X}^{(n)} \in \mathcal{X}_\mathsf{s}$, $\bar{\epsilon}_\mathrm{e2e}^{(n)} \in \mathcal{Y}_\mathsf{s}$, and N is the number of samples.

3.3.3 Target Dataset

The target dataset is the dataset for which the source model is adapted to perform a new task. This dataset is much smaller and more specific than the source dataset. The feature set and target set of the whole data are written as $\mathcal{X}_{\mathtt{t}} = [\mathbf{X}'_1, \mathbf{X}'_2, ..., \mathbf{X}'_M]^\mathsf{T}$ and $\mathcal{Y}_{\mathtt{t}} = [\bar{\epsilon}_{\text{e2e},1}, \bar{\epsilon}_{\text{e2e},2}, ..., \bar{\epsilon}_{\text{e2e},M}]^\mathsf{T}$, respectively. Hence, the structure of the target dataset is $[\mathcal{X}_{\mathtt{t}}, \mathcal{Y}_{\mathtt{t}}]$. Here the important point to note is that the number of samples used in the source model dataset (N) is much greater than the number of samples used in the target model dataset (M) i.e., $M < N$.

3.4 The Proposed CNN-based TL Algorithm

First, we design the source model $\mathcal{PT}(\mathbf{X}; \boldsymbol{\theta}^{\mathtt{src}})$ with $\mathcal{F}^{\mathtt{src}}$ being the set of all layers in the source model. Specifically, the source model is a CNN model designed for training on a large dataset to achieve a suitable robust performance in terms of accurate BLER predictions. This well-trained source model serves as our starting point. After training, layers $\mathcal{F}^{\mathtt{freeze}} \subseteq \mathcal{F}^{\mathtt{src}}$ are freezed to preserve these learned features.

Our second CNN model, i.e., the NFT target model, has fewer layers and less training data. We customize the source model for our task by adding a new fully-connected layer with appropriate output neurons, trained on our dataset. Since the source model has pre-trained features, we only train the newly added layers. We then fine-tune the source model by unfreezing certain layers in the source model, transferring its trained parameters to the non-fine-tuned (NFT) target model to better adapt to our task. After training and fine-tuning, the model's performance is evaluated on a validation set. If unsatisfactory, we modify the hyperparameters for improved performance and subsequently use this fine-tuned model for predictions. Details of the transfer learning algorithm are presented in Algorithm 3.1, and the descriptions of the source, NFT, and fine-tuned target models are explained in subsequent sections.

Algorithm 3.1 Proposed adaptation algorithm based on transfer learning

```
Initialize: BATCH_SIZE = $S_{\mathtt{batch}}$, NUM_EPOCH = 30, SPLIT_SIZE = 0.1;
Load: Source dataset $\mathcal{S}_{\mathtt{s}} = \{(\mathbf{X}_{\mathtt{s}}^{(i)}, y_{\mathtt{s}}^{(i)}) : \mathbf{X}_{\mathtt{s}}^{(i)} \in \mathcal{X}_{\mathtt{s}}, y_{\mathtt{s}}^{(i)} \in \mathcal{Y}_{\mathtt{s}}, i \in [1, L_{\mathtt{s}}]\}$;
Load: Target dataset $\mathcal{S}_{\mathtt{t}} = \{(\mathbf{X}_{\mathtt{t}}^{(i)}, y_{\mathtt{t}}^{(i)}) : \mathbf{X}_{\mathtt{t}}^{(i)} \in \mathcal{X}_{\mathtt{t}}, y_{\mathtt{t}}^{(i)} \in \mathcal{Y}_{\mathtt{t}}, i \in [1, L_{\mathtt{t}}]\}$;
Phase 1: Source model training phase,
    input Source training dataset $\mathcal{S}_{\mathtt{s}}^{\mathtt{train}} \subset \mathcal{S}_{\mathtt{s}}$, where $|\mathcal{S}_{\mathtt{s}}^{\mathtt{train}}| = L_{\mathtt{s}}^{\mathtt{train}}$ and
    $\frac{L_{\mathtt{s}}^{\mathtt{train}}}{L_{\mathtt{s}}} = 1 - \mathtt{SPLIT\_SIZE}$;
    build Source CNN model following structure in (3.6);
    $i \leftarrow 0$;
    while $i < \mathtt{NUM\_EPOCH}$ do
        $i \leftarrow i + 1$; $j \leftarrow 0$; $\boldsymbol{\theta}_{(i,j)}^{\mathtt{src}} \leftarrow \boldsymbol{\theta}_{(i-1,j)}^{\mathtt{src}}$;
        while $j < \mathtt{NO\_BATCH}$ do
            $j \leftarrow j + 1$; $\boldsymbol{\theta}_{(i,j)}^{\mathtt{src}} \leftarrow \mathtt{Optimizer}(\boldsymbol{\theta}_{(i,j-1)}^{\mathtt{src}}, \nabla \mathcal{L}_{\mathtt{batch}}(\mathcal{B}_{\mathtt{s}}^{(i,j)}; \boldsymbol{\theta}_{(i,j-1)}^{\mathtt{src}}))$;
    return: trained parameters $\tilde{\boldsymbol{\theta}}^{\mathtt{src}} \leftarrow \boldsymbol{\theta}_{(i,j)}^{\mathtt{src}}$ and pre-trained source model $\mathcal{PT}(\mathbf{X}; \tilde{\boldsymbol{\theta}}^{\mathtt{src}})$;
Phase 2: FT model training phase
    input Target training dataset $\mathcal{S}_{\mathtt{t}}^{\mathtt{train}} \subset \mathcal{S}_{\mathtt{t}}$, where $|\mathcal{S}_{\mathtt{t}}^{\mathtt{train}}| = L_{\mathtt{t}}^{\mathtt{train}}$ and
    $\frac{L_{\mathtt{t}}^{\mathtt{train}}}{L_{\mathtt{t}}} = 1 - \mathtt{SPLIT\_SIZE}$;
    build FT CNN model with structure in (3.15);
    $i \leftarrow 0$;
    while $i < \mathtt{NUM\_EPOCH}$ do
        $i \leftarrow i + 1$; $j \leftarrow 0$; $\boldsymbol{\theta}_{(i,j)}^{\mathtt{ft}} \leftarrow \boldsymbol{\theta}_{(i-1,j)}^{\mathtt{ft}}$;
        while $j < \mathtt{NO\_BATCH}$ do
            $j \leftarrow j + 1$; $\boldsymbol{\theta}_{(i,j)}^{\mathtt{ft}} \leftarrow \mathtt{Optimizer}(\boldsymbol{\theta}_{(i,j-1)}^{\mathtt{ft}}, \nabla \mathcal{L}_{\mathtt{batch}}(\mathcal{B}_{\mathtt{t}}^{(i,j)}; \boldsymbol{\theta}_{(i,j-1)}^{\mathtt{ft}}))$;
    return: trained parameters $\tilde{\boldsymbol{\theta}}^{\mathtt{ft}} \leftarrow \boldsymbol{\theta}_{(i,j)}^{\mathtt{ft}}$ and fine-tuned model $\mathcal{FT}(\mathbf{X}; \tilde{\boldsymbol{\theta}}^{\mathtt{ft}})$;
Phase 3: Inference phase,
    Input: Test dataset
    $\mathcal{S}_{\mathtt{test}} = \{(\mathbf{X}_{\mathtt{test}}^{(i)}, y_{\mathtt{test}}^{(i)}) : \mathbf{X}_{\mathtt{test}}^{(i)} \in \mathcal{X}_{\mathtt{test}}, y_{\mathtt{test}}^{(i)} \in \mathcal{Y}_{\mathtt{test}}, i \in [1, L_{\mathtt{test}}]\}$;
    Compute: RMSE = $\sqrt{\frac{1}{L^{\text{test}}} \sum_{i=1}^{L_{\mathtt{test}}} \left(\hat{y}_{\mathtt{test}}^{(i)} - y_{\mathtt{test}}^{(i)}\right)^2}$, where $\hat{y}_{\mathtt{test}}^{(i)} = \mathcal{FT}(\mathbf{X}_{\mathtt{test}}^{(i)}; \tilde{\boldsymbol{\theta}}^{\mathtt{ft}})$;
    if RMSE > criterion then
        Go to Phase 2;
    else
        Output: Trained FT model $\mathcal{FT}(\mathbf{X}; \tilde{\boldsymbol{\theta}}^{\mathtt{ft}})$;
```

3.4.1 Source CNN Model for Source Task

The source model, which consists of convolutional layers, pooling layers, fully connected layers, and output layers, is written as

$$\begin{aligned}\mathcal{PT}(\mathbf{X};\boldsymbol{\theta}^{\texttt{src}}) = \big[&f^{(11)}_{\texttt{fc}_6}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_6}) \circ f^{(10)}_{\texttt{fc}_5}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_5}) \circ f^{(9)}_{\texttt{fc}_4}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_4}) \\ &\circ f^{(8)}_{\texttt{fc}_3}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_3}) \circ f^{(7)}_{\texttt{fc}_2}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_2}) \circ f^{(6)}_{\texttt{fc}_1}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_1}) \circ f^{(5)}_{\texttt{flat}} \\ &\circ f^{(4)}_{\texttt{p}_2} \circ f^{(3)}_{\texttt{c}_2}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{c}_2}) \circ f^{(2)}_{\texttt{p}_1} \circ f^{(1)}_{\texttt{c}_1}(\boldsymbol{\theta}^{\texttt{src}}_{\texttt{c}_1})\big](\mathbf{X}),\end{aligned} \tag{3.6}$$

where $(f_2 \circ f_1(\boldsymbol{\theta}))(\mathbf{x}) \triangleq f_2(f_1(\mathbf{x};\boldsymbol{\theta}))$, $(f_2(\boldsymbol{\theta}) \circ f_1)(\mathbf{x}) \triangleq f_2(f_1(\mathbf{x});\boldsymbol{\theta})$ and $(f_2(\boldsymbol{\theta}_2) \circ f_1(\boldsymbol{\theta}_1))(\mathbf{x}) \triangleq f_2(f_1(\mathbf{x};\boldsymbol{\theta}_1);\boldsymbol{\theta}_2)$, $\boldsymbol{\theta}_{\texttt{lay}}$ represents the weights of layer $\texttt{lay}$, $f^{(l)}_{\texttt{lay}_t}(\boldsymbol{\theta}_{\texttt{lay}_t})$ represent the operation at the layer $\texttt{lay} \in \{\texttt{c}, \texttt{p}, \texttt{flat}, \texttt{fc}\}$, $t \geq 1$ denotes the index of the layer $\texttt{lay}$, and $\boldsymbol{\theta}^{\texttt{src}} = \{\boldsymbol{\theta}^{\texttt{src}}_{\texttt{c}_1}, \boldsymbol{\theta}^{\texttt{src}}_{\texttt{c}_2}, \ldots, \boldsymbol{\theta}^{\texttt{src}}_{\texttt{fc}_3}, \ldots\}$.

3.4.1.1 Architecture

Our CNN is composed two main parts, a convolution feature extraction part and a regression part. Specifically, the feature extraction part consists of several convolutional (`Conv2D`) layers and max-pooling (`MaxPooling2D`) layers, followed by the regression part, which consists of multiple fully-connected (`Dense`) layers. To ensure proper handling of regression tasks, it is essential to position a fully connected layer before the regression layer in the network's output.

3.4.1.2 2D Convolutional layer (`Conv2D`-t)

First, p_t padding is applied to the input $\mathbf{A}^{(l-1)}$ of size $H^{(l-1)} \times W^{(l-1)} \times C^{(l-1)}$ to increase its spatial dimension, which involves adding extra rows and columns of zeros around the input. Let $\mathbf{A}^{(l-1)}_t$ be the input after padding, the relationship between $\mathbf{A}^{(l-1)}_t$ and the input $\mathbf{A}^{(l-1)}$ is $[\mathbf{A}^{(l-1)}_t]_{i,j,c} = [\mathbf{A}^{(l-1)}]_{i-p_t,j-p_t,c}$ if $1 + p_t \leq i \leq H^{(l-1)} + p_t$ and $1 + p_t \leq j \leq W^{(l-1)} + p_t$, otherwise $[\mathbf{A}^{(l-1)}_t]_{i,j,c} = 0$. It is noted that $\mathbf{A}^{(0)} = \mathbf{X}$ and $\mathbf{A}^{(0)}_t$ is the padded matrix.

The next step consists of applying a CNN operation using a kernel that applies a convolutional operation to the input data. Given $C^{(l)}$ kernels, we define each l-th kernel as a matrix $[\mathbf{W}_t^{(l)}]c$ with $P^{(l)}$ rows and $Q^{(l)}$ columns, where $[\mathbf{W}_t^{(l)}]p, q, c = w_{p,q,c}^{(l)}$. Each kernel yields a feature map $[\mathbf{Z}^{(l)}]_c$ of size $H^{(l)} \times W^{(l)}$, where $H^{(l)} = \frac{H^{(l-1)}+2p_t-P^{(l)}}{s} + 1$ and $W^{(l)} = \frac{W^{(l-1)}+2p_t-Q^{(l)}}{s} + 1$ with s being the stride. The feature map $[\mathbf{Z}_t^{(l)}]_c$ is presented as

$$[\mathbf{Z}^{(l)}]_c = b_c^{(l)} + [\mathbf{W}_t^{(l)}]_c * [\mathbf{A}_t^{(l-1)}]_c, \tag{3.7}$$

$$[\mathbf{Z}^{(l)}]_{i,j,c} = b_c^{(l)} + \sum_{u=1}^{P} \sum_{v=1}^{Q} w_{v,u,c}^{(l)} \bar{a}_{s(i-1)+u-p_t, s(j-1)+v-p_t, c}^{(l-1)}, \tag{3.8}$$

$\forall c \in [1, C^{(l)}]$, where $\bar{a}_{i,j,c}^{(l)} = [\mathbf{A}_t^{(l)}]_{i,j,c}$ and $\mathbf{b}_t^{(l)} = [b_1, b_2, \ldots, b_{C^{(l)}}]^\mathsf{T}$ is the bias of the t-th convolutional layer.

After the convolution operation, an element-wise non-linear activation function $\phi_{\mathsf{c}_t}^{(l)}$ is applied to introduce non-linearity into the network. The output of `Conv2D-1` is $\mathbf{A}^{(l)} \triangleq f_{\mathsf{c}_t}^{(l)}(\mathbf{A}^{(l-1)}; \boldsymbol{\theta}_{\mathsf{c}_t}^{\mathtt{src}})$, that is a 3D matrix of size $H^{(l)} \times W^{(l)} \times C^{(l)}$. The c-th channel of $\mathbf{A}^{(l)}$, denoted as $[\mathbf{A}^{(l)}]_c$, is formulated as

$$\mathbf{A}^{(l)} = \phi_{\mathsf{c}_t}^{(l)}(\mathbf{Z}^{(l)}). \tag{3.9}$$

Herein, the parameters of the `Conv2D-`t layer are denoted as $\boldsymbol{\theta}_{\mathsf{c}_t}^{\mathtt{src}} = \{\mathbf{b}_t^{(l)}, \mathbf{W}_t^{(l)}\}$.

3.4.1.3 Max pooling layer (`MaxPooling2D-`t)

Maxpooling is a pooling operation that reduces the spatial dimension of the feature map while retaining the most important features. Let $\mathbf{A}^{(l)} = f_{\mathsf{p}_t}^{(l)}(\mathbf{A}^{(l-1)})$ be the output of the t-th maxpooling layer, we have

$$[\mathbf{A}^{(l)}]_c = \max_{1 \le u \le P^{(l)}} \max_{1 \le v \le Q^{(l)}} \bar{a}_{s(i-1)+u, s(j-1)+v, c}^{(l-1)}, \tag{3.10}$$

where $P^{(l)}$ and $Q^{(l)}$ specify the kernel size and s is the stride. It is noted that given the input $\mathbf{A}^{(l-1)}$ of size $H^{(l-1)} \times W^{(l-1)} \times C^{(l-1)}$, the size of $\mathbf{A}^{(l)}$ of the `MaxPooling2D-`t layer is obtained with $H^{(l)} = \left\lfloor \frac{H^{(l-1)} - P^{(l)}}{s} \right\rfloor + 1$, $W^{(l)} = \left\lfloor \frac{W^{(l-1)} - Q^{(l)}}{s} \right\rfloor + 1$, and $C^{(l)} = C^{(l-1)}$.

3.4.1.4 Flatten layer (`flat`)

Before transforming our CNN into a multi-layer FC CNN, we flatten the input matrix $\mathbf{A}^{(l-1)}$ to the flatten output vector $\mathbf{a}^{(l)} \in \mathbb{R}^{d^{(l)}}$ where $d^{(l)} = H^{(l-1)} W^{(l-1)} C^{(l-1)}$.

3.4.1.5 Fully connected (FC) layer (`Dense-`t)

Followed by the flatten layer, FC layers compute a linear transformation followed by an activation function $\phi^{(l)}_{\mathtt{fc}_t}$ on the preceding layer's flattened output. Let $\mathbf{a}^{(l-1)} \in \mathbb{R}^{d^{(l-1)}}$, the output vector $\mathbf{a}^{(l)} \triangleq f_{\mathtt{fc}_t}(\mathbf{a}^{(l-1)}, \boldsymbol{\theta}^{\mathtt{src}}_{\mathtt{fc}_t})$ is formulated as

$$\mathbf{a}^{(l)} = \phi^{(l)}_{\mathtt{fc}_t}(\mathbf{z}^{(l)}), \tag{3.11}$$

$$\mathbf{z}^{(l)} = \mathbf{b}^{(l)}_t + (\mathbf{W}^{(l)}_t)^{\top} \mathbf{a}^{(l-1)}, \tag{3.12}$$

where $\mathbf{W}^{(l)}_t \in \mathbb{R}^{d^{(l-1)} \times d^{(l)}}$ is the weight and $\mathbf{b}^{(l)}_t \in \mathbb{R}^{d^{(l)}}$ is the bias of the t-th fully connected layer. Herein, the parameters of the `Dense-`t layer are denoted as $\boldsymbol{\theta}^{\mathtt{src}}_{\mathtt{fc}_t} = \{\mathbf{b}^{(l)}_t, \mathbf{W}^{(l)}_t\}$.

3.4.1.6 Loss Functions in Training a Model

We assess the performance of our model in terms of root mean square error (RMSE) function. The RMSE is used to determine the discrepancy between the predicted value and the actual output value across the entire test set. During the training phase, the loss function for each (i, j)-th mini-batch $\mathcal{B}^{(i,j)}$ of each i-th epoch is calculated. This loss function is formulated as

$$\mathcal{L}_{\mathtt{batch}}(\mathcal{B}^{(i,j)}; \boldsymbol{\theta}^{\mathtt{src}}) = \frac{1}{|\mathcal{B}^{(i,j)}|} \sum_{k=1}^{|\mathcal{B}^{(i,j)}|} (\bar{\epsilon}^{(k)}_{\text{e2e}} - \hat{\epsilon}^{(k)}_{\text{e2e}})^2, \tag{3.13}$$

where $\bar{\epsilon}_{\text{e2e}}^{(k)}$ and $\hat{\epsilon}_{\text{e2e}}^{(k)} = \mathcal{PT}(\mathbf{X}^{(k)}; \boldsymbol{\theta}^{\texttt{src}})$ for $\mathbf{X}^{(k)} \in \mathcal{B}^{(i,j)}$ are the true average e2e BLER and the predicted average e2e BLER, respectively. After training each i-th epoch, the loss function over all mini-batches is formulated as

$$\mathcal{L}_{\texttt{epoch}}(\mathcal{B}^{(i,1)}, \ldots, \mathcal{B}^{(i,N_{\texttt{batch}})}; \boldsymbol{\theta}^{\texttt{src}}) = \frac{1}{N_{\texttt{batch}}} \sum_{j=1}^{N_{\texttt{batch}}} \mathcal{L}_{\texttt{batch}}(\mathcal{B}^{(i,j)}; \boldsymbol{\theta}^{\texttt{src}}),$$

where $N_{\texttt{batch}}$ denotes the number of mini-batches in the training dataset.

3.4.2 NFT **model for target task**

We utilize a reduced version of the original dataset in this part. This allows us to run the models more quickly, which is beneficial for applications with limited compute power. The mathematical representation of the NFT target model is formulated as

$$\mathcal{NFT}(\mathbf{X}; \boldsymbol{\theta}^{\texttt{nft}}) = \left[f_{\texttt{fc}_1}^{(4)}(\boldsymbol{\theta}_{\texttt{fc}_1}^{\texttt{nft}}) \circ f_{\texttt{flat}}^{(3)}(\boldsymbol{\theta}_{\texttt{flat}}^{\texttt{nft}}) \circ f_{\texttt{c}_1}^{(2)}(\boldsymbol{\theta}_{\texttt{c}_1}^{\texttt{nft}}) \circ f_{\texttt{out}}^{(1)}(\boldsymbol{\theta}_{\texttt{out}}^{\texttt{nft}}) \right](\mathbf{X}), \tag{3.14}$$

where $\boldsymbol{\theta}^{\texttt{nft}} \triangleq \{\boldsymbol{\theta}_{\texttt{c}_1}^{\texttt{nft}}, \boldsymbol{\theta}_{\texttt{flat}}^{\texttt{nft}}, \boldsymbol{\theta}_{\texttt{fc}_1}^{\texttt{nft}}, \boldsymbol{\theta}_{\texttt{out}}^{\texttt{nft}}\}$.

The training set for a NFT target model is a subset of the dataset used for the NFT target model that is used to train the model. We take a smaller subset of the training data by randomly selecting a subset of size M (where $M < N$) from the original dataset of size N. Let $\mathcal{D}_{\texttt{nft}}^{\texttt{train}}$ and $\mathcal{D}_{\texttt{nft}}^{\text{test}}$ denote the train and test sets for the NFT models. The training set $\mathcal{D}_{\texttt{nft}}^{\texttt{train}}$ can be written as $\mathcal{D}_{\texttt{train}} = (\bar{x}_j, \bar{y}_j)_{j=1}^{n}$, $(x_j, y_j) \in \{\mathcal{D}_{\texttt{t}}, \mathcal{T}_{\texttt{t}}\}$, where n is the number of samples in the training set for the NFT target model. The loss function during the training phase of our NFT model, denoted as $\mathcal{L}_{\texttt{batch}}(\mathcal{B}^{(i,j)}; \boldsymbol{\theta}^{\texttt{nft}})$, is similar to (3.13) with $\hat{\epsilon}_{\text{e2e}}^{(k)} = \mathcal{NFT}(\mathbf{X}^{(k)}; \boldsymbol{\theta}^{\texttt{nft}})$.

3.4.3 Fine-tuned (FT) model for the Target Task

We create a new model with the same architecture as the source model, but with the frozen layers excluded. This new model is used for fine-tuning. We begin by transferring the weights from the unfrozen layers in the pre-trained source model to the fine-tuned model, preserving the learned knowledge from the source dataset. Subsequently, we train this fine-tuned model using the fine-tuning dataset. The new fine-tuned target model can be formulated as

$$\mathcal{FT}(\mathbf{X};\boldsymbol{\theta}^{\mathtt{ft}}) = \big[f^{(13)}_{\mathtt{out}} \circ \overbrace{f^{(12)}_{\mathtt{fc}_7}(\boldsymbol{\theta}^{\mathtt{ft}}_{\mathtt{fc}_7})}^{\text{new layer}} \circ \overbrace{f^{(11)}_{\mathtt{fc}_6}(\boldsymbol{\theta}^{\mathtt{src}}_{\mathtt{fc}_6}) \circ \cdots \circ f^{(l_t)}_{\mathtt{fc}_t}(\boldsymbol{\theta}^{\mathtt{src}}_{\mathtt{fc}_t})}^{\text{trainable layers}} \circ f_{\mathtt{freeze}}(\tilde{\boldsymbol{\theta}}^{\mathtt{src}}_{\mathtt{freeze}}) \big](\mathbf{X}), \tag{3.15}$$

where l_t and t are the indices of the last trainable FC layer, $f_{\mathtt{freeze}}$ presents the non-trainable (freezing) layers, $\tilde{\boldsymbol{\theta}}^{\mathtt{src}}_{\mathtt{freeze}}$ are the non-trainable parameters of the source model due to freezing layers, and $\boldsymbol{\theta}^{\mathtt{ft}} \triangleq \{\boldsymbol{\theta}^{\mathtt{ft}}_{\mathtt{fc}_7}, \boldsymbol{\theta}^{\mathtt{src}}_{\mathtt{fc}_6}, \ldots, \boldsymbol{\theta}^{\mathtt{src}}_{\mathtt{fc}_t}\}$ are the trainable parameters of the FT model.

The loss function in our FT model, denoted as $\mathcal{L}_{\mathtt{batch}}(\mathcal{B}^{(i,j)};\boldsymbol{\theta}^{\mathtt{ft}})$, is similar to (3.13) with $\hat{\epsilon}^{(k)}_{\mathrm{e2e}} = \mathcal{FT}(\mathbf{X}^{(k)};\boldsymbol{\theta}^{\mathtt{ft}})$.

3.4.4 Testing of the Proposed TL Algorithm and all Models

3.4.4.1 Performance Comparison of source vs NFT vs Fine-tuned target model

The RMSE performance comparison among source model with 5 hidden layers, the NFT target model with double the number of hidden layers, and the fine-tuned target model (transfer learning model) with double the number of hidden layers of trainable parameters value is given in Table 3.1.

Table 3.1 Test performance comparison (Source vs. NFT Target Model vs. Fine-tuned-based Transfer Learning Target Model)

Models	**Trainable Parameters**	**Non-trainable Parameters**	**Total Parameters**	**Test RMSE**
Source	731,457	64	731,521	0.0925
NFT	64,993	0	64,993	0.2040
FT	17.409	731,488	748,897	0.0791

3.4.4.2 Inference on New Data

Deep learning inference is the process of using a fully trained CNN to generate inferences (predictions) based on novel (unknown) data. For inference on new data, in our work, we load the qualified trained models $\mathcal{PT}(\boldsymbol{\theta}^{\mathtt{src}})$, $\mathcal{NFT}(\boldsymbol{\theta}^{\mathtt{nft}})$, and $\mathcal{FT}(\boldsymbol{\theta}^{\mathtt{ft}})$ and feed new inference data

$$\mathbf{x}' = \Big[\mathbf{p}_{\mathrm{S}}'^{\mathsf{T}}, \mathbf{p}_{\mathrm{D}_1}'^{\mathsf{T}}, \ldots, \mathbf{p}_{\mathrm{D}_K}'^{\mathsf{T}}, \mathbf{p}_{\mathrm{I}_1}'^{\mathsf{T}}, \ldots, \mathbf{p}_{\mathrm{I}_I}'^{\mathsf{T}}, \mathbf{p}_{\mathrm{C}}'^{\mathsf{T}}, F', M', m'_{\mathrm{SC}}, m'_{\mathrm{CD}_1}, \ldots, m'_{\mathrm{CD}_K}, \\ \ell'_{\mathrm{SC}}, \ell'_{\mathrm{CD}_1}, \ldots, \ell'_{\mathrm{CD}_K}, N\Big]^{\mathsf{T}}.$$

into these trained networks. The output $\hat{\epsilon}^{(i)}_{\mathrm{e2e}}$ is calculated through the fine-tuned trained model as follows $\bar{\epsilon}^{(i)}_{\mathrm{e2e}} = \mathcal{FT}(\boldsymbol{\theta}^{\mathtt{ft}})(\mathbf{x}')$, and then compared $\hat{\epsilon}^{(i)}_{\mathrm{e2e}}$ with analytical results.

3.4.4.3 Evaluation of the Testing

We can predict our output from the trained model as $\mathcal{FT}(\mathbf{X}^{\mathrm{test}}; \boldsymbol{\theta}^{\mathtt{ft}}) \rightarrow \hat{\epsilon}^{(i)}_{\mathrm{e2e,test}}$. To determine whether or not transfer learning works, we consider the RMSE in order to assess the fine-tuned model. The RMSE is a typical means of quantifying the quality of a model's fit in statistical modelling, particularly regression analysis. It is given as RMSE $= \sqrt{\frac{\sum_{i=1}^{L^{\mathrm{test}}} \left(\hat{\epsilon}^{(i)}_{\mathrm{e2e}} - \bar{\epsilon}^{(i)}_{\mathrm{e2e,test}}\right)^2}{L^{\mathrm{test}}}}$ where, $L^{\mathrm{test}} = L^{\mathrm{data}} * 10\%$ and L^{data} denotes the total number of datapoints of the dataset. In addition, the RMSE of the transfer-model is significantly lower than that of the NFT model and is even comparable to that of the source model. The CNN structure of our proposed method is given in Fig. 3.1, where Eq. (55)) and Eq. (65) represent the output functions of the network.

3.5 Complexity analysis:

For input $\mathbf{A}^{(l-1)}$ with $C^{(l-1)}$ channels, $C^{(l)}$ kernels with size $P^{(l)} \times Q^{(l-1)}$, and feature map of size $H^{(l)} \times W^{(l)}$, the time complexity is given in Tab. 3.2: Complexity (O) of Algorithm 1.

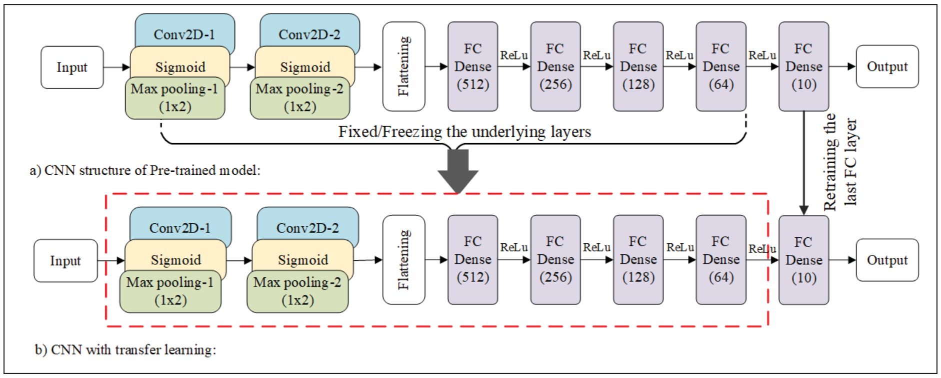

Figure 3.1 Proposed Neural Network structure a) CNN structure of pre-trained model, b) CNN with transfer learning

Table 3.2 Computational complexity ($\mathcal{O}$) of Algorithm 1

Model	**Training complexity**
Source model	$\mathcal{O}\left(1 + \max\limits_{1\le l\le 2}\left\{C^{(l)}H^{(l)}W^{(l)}P^{(l-1)}Q^{(l-1)}C^{(l-1)}\right\} + \max\limits_{6\le l\le 11}\left\{d^{(l-1)}d^{(l)}\right\}\right)$
Non fine-tuned model	$\mathcal{O}\left(1 + C^{(1)}H^{(1)}W^{(1)}P^{(0)}Q^{(0)}C^{(0)} + \max\left\{d^{(3)}d^{(2)},\ d^{(4)}d^{(3)}\right\}\right)$

Fig. 3.1 illustrates the neural network architectures used for complexity analysis, comparing the pre-trained CNN model with the transfer learning–based CNN. In the transfer learning approach, the early convolutional layers are frozen while only the final fully connected layers are retrained to reduce computational complexity.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results and Discussion

In this section, we report illustrative numerical results to show the performances of the proposed schemes in terms of average e2e BLER. In order to evaluate the performance of our system , we first validate our system model and verify the average e2e BLER of the system $(\bar{\epsilon}_{\mathrm{e2e}})$. We then investigate the effect of SNR, block-length, number of bits transmitted, and transmit power on the average overall BLER of the system. Table 4.1 lists the simulation parameters used in this study. Finally, the comparison between simulation-based prediction and analytical results is discussed in this section.

4.1.1 Average Block Error Rate

This subsection explains the average BLER as influenced by the distribution of transmission powers (P_{S} and P_{C}) and interference power distribution. Fig. 4.1 shows a comparison of the performance of the proposed method with a benchmark technique previously reported in (Yuan *et al.*, 2022).

Fig. 4.1 illustrates the average end-to-end BLER as a function of the transmit power P_{S} for two different blocklength values, namely $M = 150$ and $M = 200$, under interference-limited conditions. For each blocklength, both analytical results and Monte-Carlo simulation results are presented to validate the accuracy of the derived expressions. For a fixed transmit power, the curves associated with $M = 200$ consistently achieve a lower BLER compared to those for $M = 150$. This behavior highlights the benefit of using a larger blocklength, as increasing M provides more channel uses per packet, thereby improving decoding reliability in the finite blocklength regime. As expected, this improvement comes at the cost of increased latency, which reflects the fundamental reliability–latency trade-off in URLLC systems. Moreover, the close agreement between the analytical curves and the corresponding simulation results

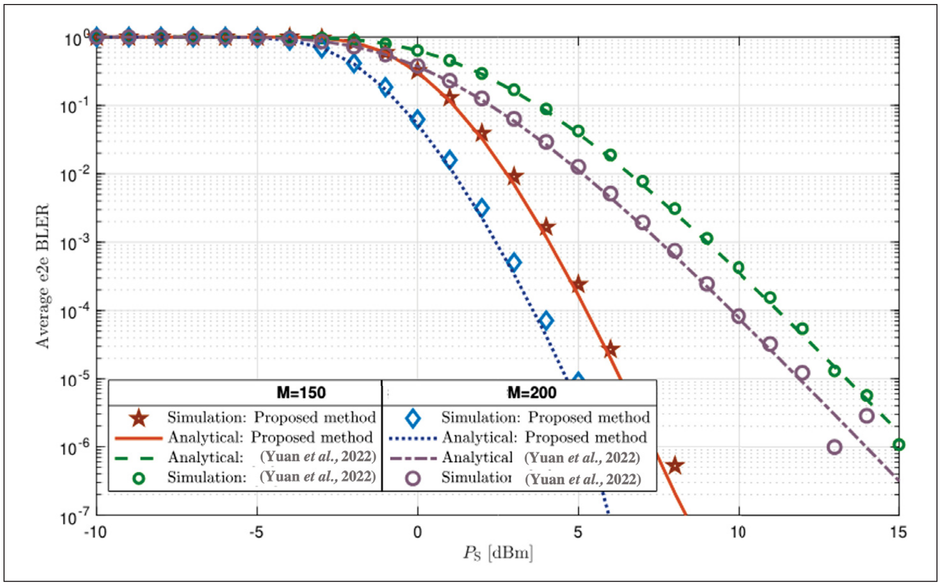

Figure 4.1 Performance comparison of proposed method with a benchmark technique (Yuan *et al.*, 2022) using no interferer

Table 4.1 Simulation Parameters

Parameter	**Value**
Blocklength (M)	[100-500]
Carrier frequency, f_c [GHz]	3
Nakagami-m shape parameters	[3,5]
Noise figure, NF [dBm]	10
No. ground users, K	3
No. information bits, F [bits]	100
No. Kernel per Conv-2D layer	4
No. neuron per `Dense`-t layer	$1024/2^{t-1}$
Transmit power at S, P_S [dBm]	[0,30]
Transmit power at C, P_C [dBm]	[0,30]
Thermal noise power density [dBm/Hz]	-174
No. antenna, N	4
No. interferer, L	3
Noise bandwidth, BW [MHz]	10
Kernel size	(2,3)

for both $M = 150$ and $M = 200$ confirms the accuracy of the proposed analytical model across different blocklength values. Compared with the benchmark scheme from (Yuan *et al.*, 2022), the proposed method demonstrates a noticeable BLER improvement, especially in the medium-to-high transmit power region.

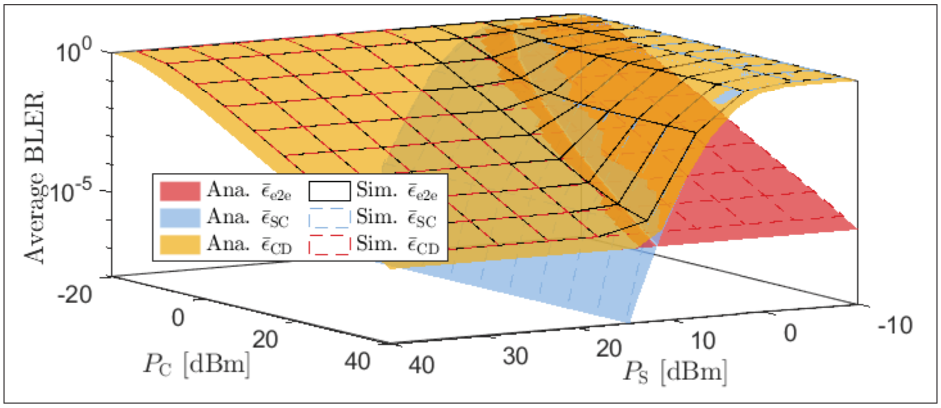

Figure 4.2 Average e2e BLER as a function of transmission powers with $M = 500$ symbol/block

In Fig. 4.2, the average BLER of the first hop, second hop, and e2e BLER were observed under different values of P_S and P_C. First, we found that the analytical results matched well with the simulation results, which showed the high accuracy of the proposed approximation. We also observed that with an increase of both P_S and P_C, the average BLER of each hop and of the system also increased. This is because the system achieves diversity gain as $P_S \to \infty$ and $P_C \to \infty$. However, when fixing either P_S or P_C and varying P_C or P_S, respectively, we observed error floor levels. Specifically, the error floor at $P_C = 15$ dBm is shown in Fig. 4.2.

The insights obtained from Fig. 4.2 regarding the impact of power levels on BLER directly enlighten the analysis presented in Fig. 4.3, where we explored the broader implications of transmission parameters, such as block length, on network performance.

Fig. 4.3, showed the transmit power P_S versus the average e2e BLER with different blocklengths, i.e., $M = 128$, 256, and $M = 512$. In Fig. 4.3, the analytical simulation and asymptotic curves are shown when the number of information bits received by the D_k $(F) = 100$, and $P_C = 15$ dBm. Again, the analytical curves match well with the simulation curves, which confirms

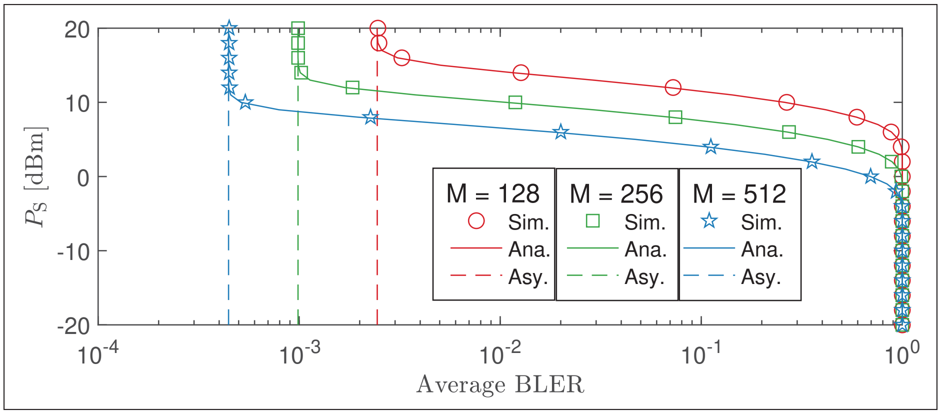

Figure 4.3 Source transmission power as a function of the average e2e BLER with different values of blocklengths

validity of our analysis. Furthermore, as shown in Fig. 4.3, for fixed channel conditions, with an increase of the blocklength M the average e2e BLER and the required P_S decreased. Next, we observed the average e2e BLER when varying the blocklength. The insights gained from Section 4.1.1 regarding the impact of power levels on BLER directly enlighten the analysis presented in Section 4.1.2 where we explored the broader implications of transmission parameters such as block length, on network performance.

4.1.2 Impact of Blocklength

In this section, we explored into the impact of M on the average BLER under varying values of P_S. Building upon the evidence in Fig. 4.3, foundation from Fig. 4.3, where we used both analytical and simulation results to examine the impact of varying BS power levels P_S and M on BLER, we further explored the influence of interference power distribution on the average block length in Fig. 4.4.

In Fig. 4.4, we investigated the impact of the blocklength on the average e2e BLER with a maximum blocklength, M, set to 500 symbol/block and a source power, P_S, of 33 dBm. The observed trend showed a decrease in the average e2e BLER as the blocklength increased. Specifically, at a blocklength of $M = 100$ symbols/block, elevating the source power, P_S, from

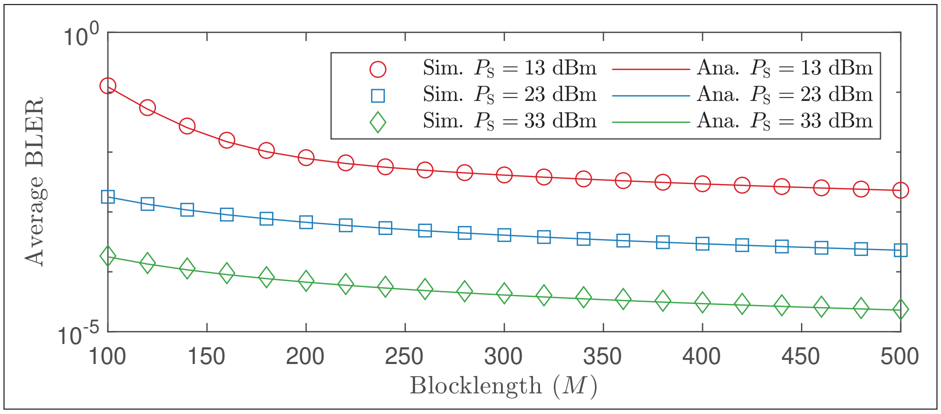

Figure 4.4 Impact of blocklength on the average BLER with different values of source transmission power

13 dBm to 23 dBm induced an approximately two orders of magnitude decrease in the average e2e BLER. However, an additional 10 dBm increase in P_S resulted in only a single order of magnitude reduction in the average e2e BLER. This outcome is due to the presence of an error floor in the system at a fixed cluster head power, P_C, as detailed in our previous discussion of Fig. 4.3.

In Fig. 4.5, we examined the influence of M on the average e2e BLER, with a maximum block M set to 300 symbols per block and a source power P_{src} of 30 dBm.

The observed trend indicated a decrease in the average e2eBLER with an increase of M. However, we observed an increase in the average end-to-end block error rate when the number of users k increased, while P_{src}, P_C, and M remained constant. Fig. 4.5 provides valuable insights into how both block length and the number of users affect reliability and delay of transmissions in the system.

In Fig. 4.6, we focused on the impact of the interference power distribution, denoted as μ, on the average e2e (system) BLER is . The interference power distribution was computed as shown

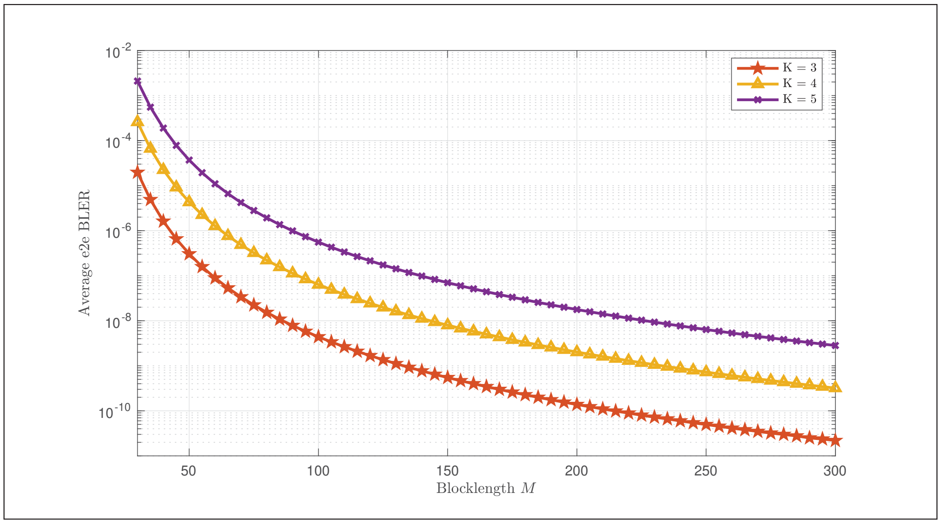

Figure 4.5 Impact of block length M and number of users K with constant P_{src}, P_{C}, and M

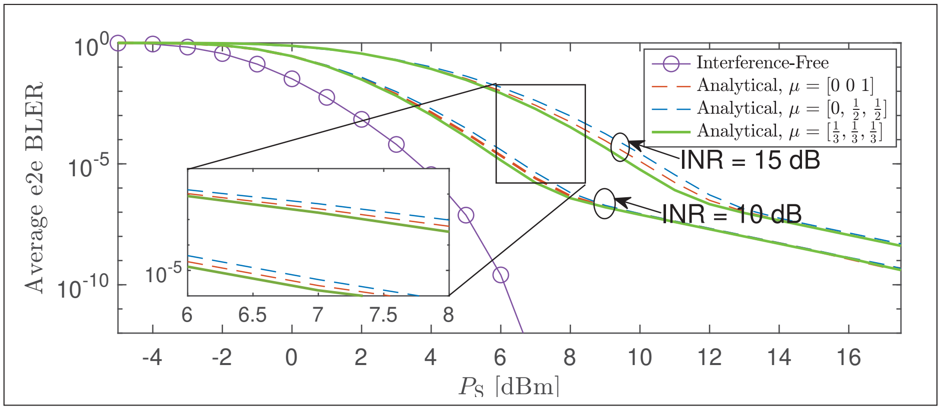

Figure 4.6 Impact of interference power distribution on the average blocklength, where $m_{\text{I}_i\text{C}} = m_{\text{I}_i\text{D}_k}$ for all $i \in [1, I]$ and $k \in [1, K]$

in Eq. (4.1):

$$\mu = \frac{1}{\sum_{i=1}^{I} P_{\text{I}_i\text{C}}\ell_{\text{I}_i\text{C}}}[P_{\text{I}_1\text{C}}\ell_{\text{I}_1\text{C}}, \dots, P_{\text{I}_I\text{C}}\ell_{\text{I}_I\text{C}}] = \frac{1}{\sum_{i=1}^{I} P_{\text{I}_i\text{D}_k}\ell_{\text{I}_i\text{D}_k}}[P_{\text{I}_i\text{D}_k}\ell_{\text{I}_i\text{D}_k}, \dots, P_{\text{I}_I\text{D}_k}\ell_{\text{I}_i\text{D}_k}]; \forall k \in [1, K] \tag{4.1}$$

The interference-free condition, i.e., $I = 0$ interferers, resulted in the lowest average e2e BLER. Interestingly, the system performance was impactednot only by the presence of interference, but also by the variation of μ under the same INR values. Specifically, the lowest average e2e BLER was observed in an IID interference environment.

Building upon the evidence shown in Fig. 4.4, where we used both analytical and simulation results to examine the impact of varying BS power levels P_S and M on BLER, we further explored the influence of interference power distribution on the average block length in Fig. 4.6.

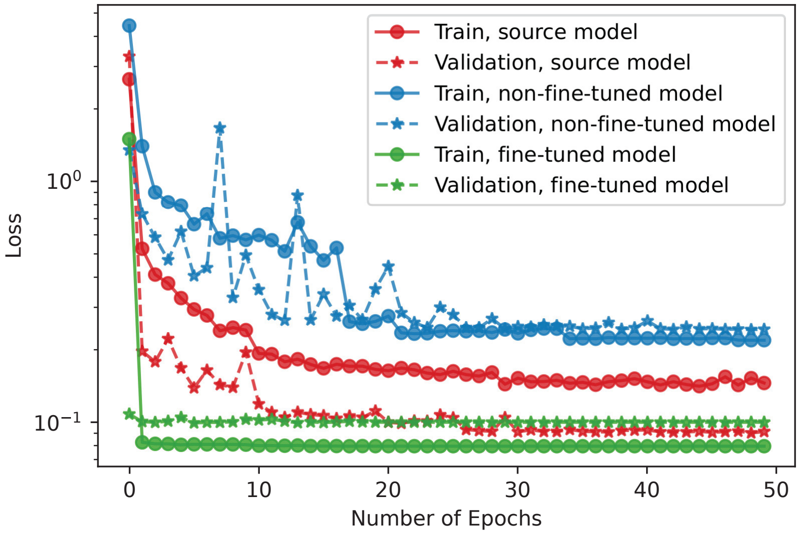

Figure 4.7 Training convergence of the source, non-fine-tune, and fine-tune models

4.1.3 Convergence and Performance of Transfer Learning

In this section, we discuss the convergence of training for the source, non-fine-tuned, and fine-tuned models, along with a comparative analysis between the true (analytical) curve and the predicted results derived from the proposed method.

In Fig. 4.7, the impact of transfer learning on training convergence is illustrated by the green curves, which correspond to the fine-tuned (transfer-learning–enabled) model. Unlike the non-fine-tuned model, which is trained from random initialization, the fine-tuned model leverages pretrained source-domain weights and frozen lower layers. As a result, the fine-tuned model achieves rapid convergence within a few epochs, exhibits reduced training and validation loss, and shows significantly improved stability. This demonstrates that transfer learning substantially reduces convergence time and training complexity.

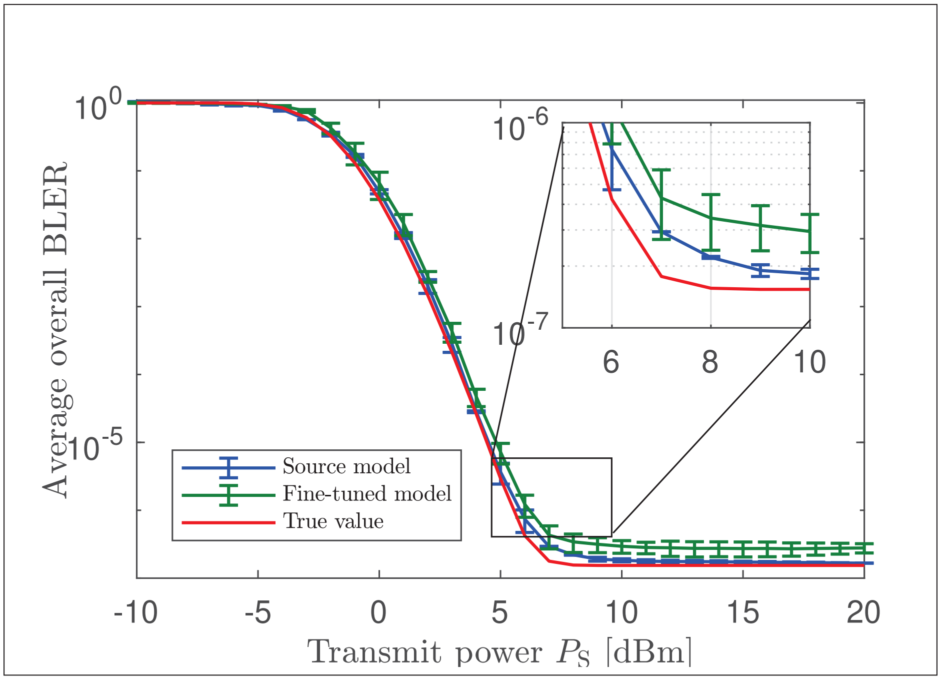

Figure 4.8 Impact of the source's transmission power on the true and predicted BLER of the source and the fine-tuned models

More specifically, in Fig. 4.8, the error bars representing the confidence intervals of the NFT, source, and fine-tuned models are clearly visible. Importantly, We also observed that, with an increase of as the transmission power increased, these confidence intervals (error bars) decreased, implying reduced variability and enhanced precision in the models' predictions. This reduction in the size of the confidence intervals can be attributed to the fact that a higher transmission power typically results in a stronger signal quality, thereby reducing uncertainty in the data and leading to more consistent and reliable model predictions.

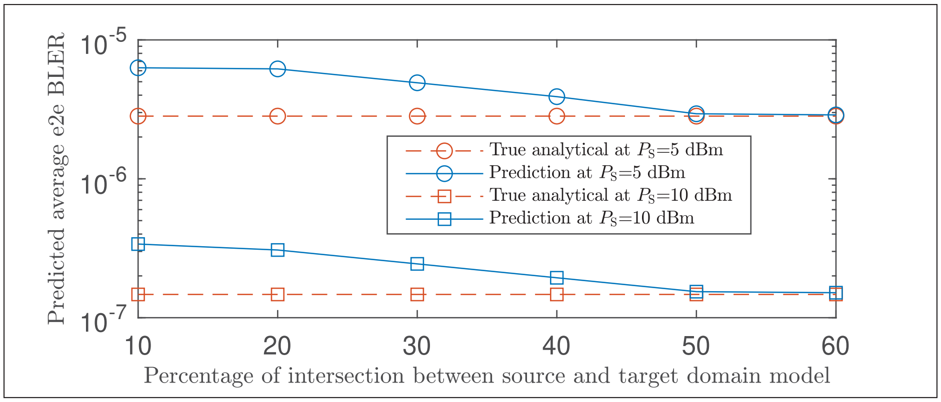

Figure 4.9 Impact of adaptation domain on prediction accuracy

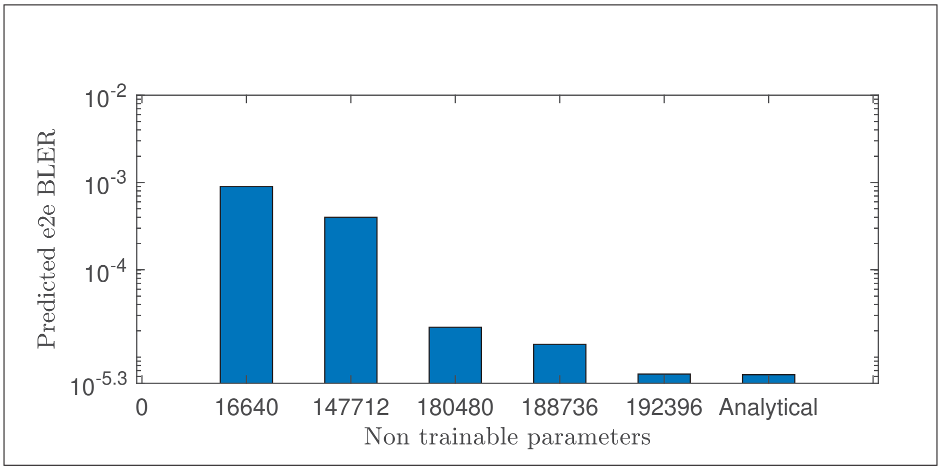

Figure 4.10 Impact of adaptation domain on prediction accuracy

4.1.4 Impact of domain adaptation and transferred knowledge

In this section,, we explain the significance of the intersection between the source domain and target domain models for the overall prediction of our average BLER. We also examine the effect of transferred knowledge, specifically non-trainable parameters, on prediction accuracy.

The impact of the size of the intersection between source domain and target domain can be seen in Fig. 4.9. Some part of source domain dataset was used for target domain dataset, i.e.,

$\mathcal{D}_\mathsf{s} \neq \mathcal{D}_\mathsf{t}$ and $\mathcal{D}_\mathsf{s} \cap \mathcal{D}_\mathsf{t} \neq \emptyset$. Of note, increasing the mutuality from 10% to 60% enhanced the average e2e BLER predictions with the fine-tuned target model, with 50% overlap yielding a notably accurate e2e BLER $\bar{\epsilon}_{\mathrm{e2e}}$.

In Fig. 4.10, the size of the non-trainable parameter implies the amount of transferred knowledge. The higher the number of non-trainable parameters, the greater the transferred knowledge and hence, the better the performance of fine-tuned target model. Specifically, at 192,396 shared parameters, we obtained a high prediction accuracy for the fine-tuned model. This behavior can be attributed to the domain adaptation principle, where increasing shared knowledge, the more the target model can learn from the source, which results in a better prediction accuracy as it more effectively models the target domain's characteristics.

CHAPTER 5

FUTURE WORK AND DIRECTIONS

While the results reported in Chapters 2-4 of the present thesis lay down a complete analytical and ML-driven performance prediction framework for URLLC, several promising extensions can further enhance system reliability and adaptability. In this chapter, we outline the next research direction based on a preliminary system model and early analytical formulations prepared for a future publication.

5.1 Motivation for Future Work

Ultra-Reliable Low-Latency Communication (URLLC) requires extremely high reliability and strict delay constraints, particularly in mission-critical networks such as emergency response, smart manufacturing, remote surgery, and autonomous aerial systems. While the first part of the present thesis is addressed performance prediction for terrestrial URLLC links, real-world deployments increasingly involve multi-hop communication assisted by aerial base stations (ABS) such as UAVs or high-altitude platforms.

Aerial relays can dynamically reposition themselves, establish line-of-sight links, and rapidly extend coverage. However, their mobility, altitude-dependent channel variations, and energy constraints introduce additional uncertainty into the communication model. Classical analytical approaches become mathematically complex when multi-hop relaying and aerial mobility are combined.

Furthermore, URLLC systems frequently operate with limited channel observations, making it difficult to estimate or adapt parameters such as SNR, interference power, link quality, and fading characteristics. A Variational Autoencoder (VAE) provides a promising direction for future work because its latent space can capture meaningful structure in channel parameters and predict unseen configurations.

This motivates a new research direction: namely, that of the integration of multi-hop aerial relaying and VAE-based parameterization for accurate prediction of end-to-end reliability in URLLC networks.

5.2 Proposed System Model for Future Research

In this future research direction, we consider a multi-hop relaying architecture that would incorporate aerial platforms for enhanced URLLC connectivity. The envisioned network, illustrated in Fig. 2.1, consists of a terrestrial base station (TBS), a set of autonomous aerial vehicles (AAVs) acting as relay nodes, and multiple ground users. Throughout this section, the TBS, aerial relays, and users are denoted by S, R_j, and U_k, respectively.

The TBS, equipped with M antennas, initiates the transmission of short-packet URLLC messages. Each AAV operates as a lightweight aerial relay with a single antenna and follows a half-duplex decode-and-forward (DF) protocol, which aligns with the practical design constraints of low-power and size-restricted airborne platforms. The ground users, indexed by $k \in \{1, 2, \ldots, K\}$, are also modeled as single-antenna devices.

In many practical scenarios such as dense urban environments, disaster-response operations, or areas with severe blockage, a direct communication link between the TBS and the ground users may be unreliable or entirely unavailable. To ensure dependable connectivity in such conditions, the network relies on a sequence of aerial relays. The data are forwarded from the TBS to the first AAV, then to subsequent AAVs, and, ultimately, to the destination user in a hop-by-hop manner.

At each hop, the relay attempts to decode the received short-packet transmission and, upon successful decoding, forwards the message to the next node using its available transmit power P_{R_j}. This sequential process continues until the message reaches the intended user. The overall transmission is affected by channel impairments and potential ambient interference. Although the specific channel and interference models are not detailed here, preserving the conceptual components for future publication the general framework remains consistent with the interference-limited characteristics examined in our previous work.

Furthermore, the multi-hop aerial relaying configuration in this thesis establishes a flexible foundation for investigating several advanced URLLC concepts, including reliability analysis under mobility, prediction of performance with aerial assistance, and learning-based parameterization strategies such as VAE-driven latent modeling. These elements constitute the core of the proposed future research direction and will be explored in subsequent stages while maintaining confidentiality of technical details intended for future scientific dissemination.

5.3 Analytical Framework and Reliability Perspective

Based on the multi-hop aerial relaying architecture described in the previous section, the next stage of this research will focus on characterizing reliability of the system in terms of its end-to-end block error rate (BLER). In a traditional analytical setting, the BLER of multi-hop decode-and-forward relaying can be expressed as a function of the instantaneous signal-to-noise ratio (SNR), interference characteristics, and channel conditions across all hops.

Although explicit expressions are not presented here to maintain confidentiality of the future publication, the general idea is to derive the statistical behavior of the hop-by-hop SNR, incorporate the decoding success probability at each relaying node, and aggregate these results into an end-to-end reliability metric. This will involve modeling how impairments such as fading, aerial mobility, and multi-hop interference affect the probability of decoding failure at each relay.

The envisioned analytical component of the work will seek to establish a tractable reliability model that will capture the essential behavior of the system while also serving as a foundation for machine-learning-based parameterization and reliability prediction. The resulting expressions will ultimately be used to validate performance of learning-driven approaches and to support the design of optimized aerial-terrestrial URLLC links.

5.3.1 Rationale for Using Variational Autoencoders

Integrating of a VAE into the proposed framework is motivated by the need to handle complex and high-dimensional channel behaviors arising in aerial-terrestrial multi-hop URLLC systems. Unlike classical analytical models that rely on closed-form assumptions, a VAE enables extracting of meaningful low-dimensional representations of intricate channel statistics, interference patterns, and mobility-driven variations.

A VAE is particularly well suited for this task because it learns a latent distribution rather than a deterministic feature representation. This allows the model to generalize to unseen propagation conditions and to generate parameter samples that reflect realistic channel behaviors. This capability is invaluable for URLLC applications, where exhaustive measurement data are frequently unavailable and where robustness to rare or extreme channel events is essential.

In summary, the VAE serves as a bridge between detailed analytical modeling and scalable data-driven prediction, allowing the system to learn compressed, yet highly informative representations of the parameters governing end-to-end reliability.

5.3.2 Latent Dimension in the Proposed Learning Model

Traditional VAEs employ a latent space designed primarily for generative modeling, where the dimensionality is chosen to balance reconstruction fidelity and sample diversity. In the context of the proposed system, the latent space plays a different and more specialized role: it must encode information that directly affects reliability, such as SNR variations, hop-by-hop degradation, or channel uncertainty.

Therefore, the latent dimension in the new learning architecture may differ from that of a conventional VAE. Instead of focusing on generative richness, the aim is to optimize the latent dimension for relevance, interpretability, and predictive accuracy. A more compact latent space would reduce redundancy and highlights the dominant parameters affecting BLER, whereas a slightly larger latent dimension may be required to capture subtle multi-hop dependencies or aerial mobility effects.

Therefore, the selection of the latent dimension will be guided by both analytical insight into the system model and empirical validation through training and performance evaluation.

CONCLUSION AND RECOMMENDATIONS

This work presented a comprehensive performance analysis of an ultra-reliable low-latency communication (URLLC) system designed for clustered wireless networks operating under finite blocklength constraints. The considered model relies on a cluster head acting as a wireless relay to ensure highly reliable and low-latency transmission between a base station and ground users, despite the presence of multiple interference sources. A closed-form analytical expression for the overall Block Error Rate (BLER) was derived based on a theoretical framework that incorporates key system parameters, including packet size, blocklength, and the achievable coding rate, which collectively determine system performance.

In addition, a transfer learning based prediction mechanism was developed, leveraging domain adaptation and fine-tuning of pretrained models to enable real-time performance estimation. This approach demonstrates significant potential for improving prediction accuracy while facilitating the practical deployment of URLLC solutions in dynamic environments where rapid performance assessment is operationally critical.

The obtained results highlight the effectiveness of combining rigorous analytical modeling with advanced machine learning techniques to characterize and anticipate the performance of URLLC systems subject to stringent reliability and latency constraints. The derived closed-form BLER expression provides precise and interpretable performance bounds, while the proposed transfer learning framework enables adaptive and data-efficient predictive modeling under realistic network conditions. Together, these contributions establish an integrated perspective in which theoretical modeling and artificial intelligence complement each other to enhance the robustness and efficiency of communication systems operating in demanding scenarios.

Future research directions will focus on extending the proposed framework to more complex system configurations, including those involving autonomous aerial relays, multi-hop communication environments, and hybrid aerial–terrestrial network topologies. Such scenarios introduce richer channel dynamics, inter-hop dependencies, and mobility-induced uncertainties that challenge conventional analytical methods. In this context, generative learning techniques, particularly variational autoencoders, represent a promising avenue for capturing and parameterizing the latent structure of these complex environments.

Future investigations will aim to develop a unified analytical and learning-based framework capable of reliably predicting end-to-end BLER while enabling adaptive resource optimization in next-generation URLLC networks. This research direction constitutes a natural extension of the present work and provides a solid foundation for future scientific contributions.

APPENDIX I

PROOF OF LEMMA 1 - CHAPTER 2

In order to evaluate the PDF of $\mathrm{I_C}$, let $\gamma_{\mathrm{I}_i\mathrm{C}} = \bar{\gamma}_{\mathrm{I}_i\mathrm{C}}\ell_{\mathrm{I}_i\mathrm{C}}|h_{\mathrm{I}_i\mathrm{C}}|^2$. Since $|h_{\mathrm{I}_i\mathrm{C}}|^2$ follows a gamma distribution, the PDF and CDF of $\gamma_{\mathrm{I}_i\mathrm{C}}$ are obtained as shown below.

$$\begin{aligned} f_{\gamma_{\mathrm{I}_i\mathrm{C}}}(\gamma) &= \frac{1}{(m_{\mathrm{I}_i\mathrm{C}}-1)!}\left(\frac{m_{\mathrm{I}_i\mathrm{C}}}{\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}}\right)^{m_{\mathrm{I}_i\mathrm{C}}} \\ &\quad \times \gamma^{m_{\mathrm{I}_i\mathrm{C}}-1}\exp\left(-\frac{m_{\mathrm{I}_i\mathrm{C}}}{\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}}\gamma\right), \qquad \text{(A I-1)} \\ F_{\gamma_{\mathrm{I}_i\mathrm{C}}}(\gamma) &= 1-\exp\left(-\frac{m_{\mathrm{I}_i\mathrm{C}}}{\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}}\gamma\right)\sum_{m=0}^{m_{\mathrm{I}_i\mathrm{D}_k}-1}\frac{1}{m!}\left(\frac{m_{\mathrm{I}_i\mathrm{C}}}{\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}}\gamma\right)^{m}, \qquad \text{(A I-2)} \end{aligned}$$

where $\gamma > 0$. The characteristic function (c.f.) of $\gamma_{\mathrm{I}_i\mathrm{C}}$ is given by the following:

$$\begin{aligned} \Phi_{\gamma_{\mathrm{I}_i\mathrm{C}}}(j\omega) &= \mathbb{E}\left\{e^{j\omega\gamma_{\mathrm{I}_i\mathrm{C}}}\right\} \\ &= \frac{\left(\frac{m_{\mathrm{I}_i\mathrm{C}}}{\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}}\right)^{m_{\mathrm{I}_i\mathrm{C}}}}{(m_{\mathrm{I}_i\mathrm{D}_k}-1)!}\int_0^\infty \gamma^{m_{\mathrm{I}_i\mathrm{C}}-1}e^{-\left(\frac{m_{\mathrm{I}_i\mathrm{C}}}{\bar{\gamma}_{\mathrm{I}_i}\ell_{\mathrm{I}_i\mathrm{C}}}-j\omega\right)\gamma} \qquad \text{(A I-3)} \end{aligned}$$

Using the identity [Eq. (3.381.3), R1], the c.f. of $\gamma_{\mathrm{I}_i\mathrm{D}_k}$ can be derived in closed-form as $\Phi_{\gamma_{\mathrm{I}_i\mathrm{C}}} = \left(1-j\omega/a_{\mathrm{I}_i\mathrm{C}}\right)^{-m_{\mathrm{I}_i\mathrm{C}}}$. Hence, the c.f. of $\mathrm{I_C}$, denoted as $\Phi_{\mathrm{I_C}}(j\omega) = \mathbb{E}\left\{e^{j\omega\mathrm{I_C}}\right\}$, is given by the following:

$$\begin{aligned} \Phi_{\mathrm{I_C}}(j\omega) &= \mathbb{E}\left\{\exp\left(j\omega\sum_{i\in\mathcal{I}}\gamma_{\mathrm{I}_i\mathrm{C}}\right)\right\} \\ &= \prod_{i\in\mathcal{I}}\mathbb{E}\left\{e^{j\omega\gamma_{\mathrm{I}_i\mathrm{C}}}\right\} \\ &= \prod_{i\in\mathcal{I}}\Phi_{\gamma_{\mathrm{I}_i\mathrm{C}}}(j\omega) \\ &= \prod_{i\in\mathcal{I}}\left(1-j\omega/a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}\right)^{-m_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}}, \qquad \text{(A I-4)} \end{aligned}$$

where $\gamma_{\mathrm{I}_i\mathrm{C}}$ for $i \in \mathcal{I}$ are mutually independent.

Consider an $n \times n$ matrix $\mathbf{\Lambda}$ with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ in any order. Let $\varrho(\mathbf{\Lambda})$ be the number of distinct eigenvalues, $\lambda_{\langle 1\rangle}, \lambda_{\langle 2\rangle}, ..., \lambda_{\langle \varrho(\mathbf{\Lambda})\rangle}$ be distinct eigenvalues in decreasing order, and $\tau_i(\mathbf{\Lambda})$ the multiplicity of $\lambda_{\langle i\rangle}$. Then, the $(i,j)^{\text{th}}$ characteristics coefficient $\chi_{i,j}(\mathbf{\Lambda})$, $i = 1, 2, ..., \varrho(\mathbf{\Lambda})$, $j = 1, 2, ..., \tau_i(\mathbf{\Lambda})$, is defined as a partial fraction expansion coefficient of $\det(\mathrm{I}_n + \xi\mathbf{\Lambda})^{-1}$ such that

$$\begin{aligned}\det(\mathrm{I}_n + \xi\mathbf{\Lambda})^{-1} &= \prod_{i=1}^{\varrho(\mathbf{\Lambda})} \left(1 + \xi\lambda_{\langle i\rangle}\right)^{-\tau_i(\mathbf{\Lambda})} \\ &= \sum_{i=1}^{\varrho(\mathbf{\Lambda})} \sum_{j=1}^{\tau_i(\mathbf{\Lambda})} \chi_{i,j}(\mathbf{\Lambda}) \left(1 + \xi\lambda_{\langle i\rangle}\right)^{-j}, \end{aligned} \tag{A I-5}$$

where ξ is a scalar constant such that $(\mathrm{I}_n + \xi\mathbf{\Lambda})$ is non-singular and $\chi_{i,j}(\mathbf{\Lambda})$ can be determined by the following equation:

$$\begin{aligned}\chi_{i,j}(\mathbf{\Lambda}) &= \frac{1}{(\tau_i(\mathbf{\Lambda}) - j)!\lambda_{\langle i\rangle}^{\tau_i(\mathbf{\Lambda})-j}} \\ &\times \left[\frac{\mathrm{d}^{\tau_i(\mathbf{\Lambda})-j}}{\mathrm{d}\xi^{\tau_i(\mathbf{\Lambda})-j}} (1 + \xi\lambda_{\langle i\rangle})^{\tau_i(\mathbf{\Lambda})} \det(\mathrm{I}_n + \xi\mathbf{\Lambda})^{-1}\right]\Bigg|_{s \to -\frac{1}{\lambda_{\langle i\rangle}}}, \end{aligned} \tag{A I-6}$$

where $\sum_{i=1}^{\varrho(\mathbf{\Lambda})} \sum_{j=1}^{\tau_i(\mathbf{\Lambda})} \chi_{i,j}(\mathbf{\Lambda}) = 1$. Using a partial fraction decomposition of $\Phi_{\mathrm{I_C}}(j\omega)$, we obtain

$$\Phi_{\mathrm{I_C}}(j\omega) = \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \left(1 + \frac{j\omega}{a_{\mathrm{I}_{\langle i\rangle}\mathrm{C}}}\right)^{-j}. \tag{A I-7}$$

The PDF of $\mathrm{I_C}$, defined as $f_{\mathrm{I_C}}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega x}\Phi_{\mathrm{I_C}}(j\omega)\mathrm{d}\omega$, can be written as follows:

$$f_{\mathrm{I_C}}(x) = \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \frac{\infty}{\in\pi} \int_{-\infty}^{\infty} \rceil^{|\omega\S} \left(\infty + |\omega/\dashv_{\mathrm{I}_{\langle\rangle\rangle}\mathrm{C}}\right)^{-|} \mathrm{d}\omega. \tag{A I-8}$$

Hence we achieve the final expression of the PDF of $\mathrm{I_C}$ given in Eq. (2.5). Now, the CDF of $\mathrm{I_C}$, defined as $F_{\mathrm{I_C}}(x) = \int_0^x f_{\mathrm{I_C}}(t)\mathrm{d}t$, is obtained as follows:

$$F_{\mathrm{I_C}}(x) = 1 - \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \chi_{i,j}(\mathcal{A}) \frac{a^{j}_{\mathrm{I}^{\mathrm{C}}_{\langle i \rangle}}}{(j-1)!} \int_x^\infty t^{j-1} e^{-t\, a_{\mathrm{I}^{\mathrm{C}}_{\langle i \rangle}}} \mathrm{d}t \tag{A I-9}$$

Following (Gradshteyn *et al.*, 2007), we obtain Eq. (2.6). This completes the proof of Lemma 1.

APPENDIX II

PROOF OF LEMMA 2 - CHAPTER 2

In order to evaluate the PDF of $\mathrm{I}_{\mathrm{D}_k}$, let $\gamma_{\mathrm{I}_i\mathrm{D}_k} = \bar{\gamma}_{\mathrm{I}_i\mathrm{D}_k}\ \ell_{\mathrm{I}_i\mathrm{D}_k}|h_{\mathrm{I}_i\mathrm{D}_k}|^2$, since $|h_{\mathrm{I}_i\mathrm{D}_k}|^2$ follows a Gamma distribution, the PDF and CDF of $\gamma_{\mathrm{I}_i\mathrm{D}_k}$ are obtained as follows:

$$\begin{aligned} f_{\gamma_{\mathrm{I}_i\mathrm{D}_k}}(\gamma) &= \frac{(a_{\mathrm{I}_i\mathrm{D}_k})^{m_{\mathrm{I}_i\mathrm{D}_k}}}{(m_{\mathrm{I}_i\mathrm{D}_k}-1)!}\gamma^{m_{\mathrm{I}_i\mathrm{D}_k}-1}e^{-a_{\mathrm{I}_i\mathrm{D}_k}\gamma}, \\ F_{\gamma_{\mathrm{I}_i\mathrm{D}_k}}(\gamma) &= 1 - e^{-a_{\mathrm{I}_i\mathrm{D}_k}\gamma}\sum_{m=0}^{m_{\mathrm{I}_i\mathrm{D}_k}-1}\frac{(a_{\mathrm{I}_i\mathrm{D}_k})^m}{m!}, \end{aligned} \tag{A II-1}$$

where $\gamma > 0$. Now, following the steps used in Lemma 1, the c.f. of $\mathrm{I}_{\mathrm{D}_k}$ is derived as shown below.

$$\Phi_{\mathrm{I}_{\mathrm{D}_k}}(j\omega) = \sum_{i=1}^{\varrho(\boldsymbol{\mathcal{B}_k})}\sum_{j=1}^{\tau_i(\boldsymbol{\mathcal{B}_k})}\chi_{i,j}(\boldsymbol{\mathcal{B}_k})\big(1 + j\omega\mu_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}\big)^{-j}. \tag{A II-2}$$

Hence, the PDF of $\mathrm{I}_{\mathrm{D}_k}$, defined as $f_{\mathrm{I}_{\mathrm{D}_k}}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega x}\Phi_{\mathrm{I}_{\mathrm{D}_k}}(j\omega)\mathrm{d}\omega$, and the CDF of $\mathrm{I}_{\mathrm{D}_k}$, defined as $F_{\mathrm{I}_{\mathrm{D}_k}}(x) = \int_0^x f_{\mathrm{I}_{\mathrm{D}_k}}\mathrm{d}t$, are obtained as follows:

$$f_{\mathrm{I}_{\mathrm{D}_k}}(x) = \sum_{i=1}^{\varrho(\mathcal{B}_k)}\sum_{j=1}^{\tau_i(\mathcal{B}_k)}\chi_{i,j}(\mathcal{B}_k)\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega x}\big(1 + j\omega\mu_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}\big)^{-j}\,\mathrm{d}\omega \tag{A II-3}$$

$$F_{\mathrm{I}_{\mathrm{D}_k}}(x) = 1 - \sum_{i=1}^{\varrho(\mathcal{B}_k)}\sum_{j=1}^{\tau_i(\mathcal{B}_k)}\frac{\chi_{i,j}(\mathcal{B}_k)}{(j-1)!}\Gamma(j, \mu_{\mathrm{I}_{\langle i\rangle}\mathrm{D}_k}), \tag{A II-4}$$

respectively. After some mathematical manipulations, we obtain Eq. (2.15) and Eq. (2.16). This completes the proof of Lemma 3.

APPENDIX III

PROOF OF THEOREM 1 - CHAPTER 2

From Eq. (2.28), we have $Z(\Gamma_{\mathrm{SC}}) \approx \frac{1}{2} - \tau\left(\Gamma_{\mathrm{SC}} - \omega\right)$ for $\psi < \Gamma_{\mathrm{SC}} < \eta$. Hence, the average BLER for the S-to-C, link i.e., $\bar{\epsilon}_{\mathrm{SC}}$, can be written as follows:

$$\bar{\epsilon}_{\mathrm{SC}} \approx \int_0^{\infty} Z(\Gamma_{\mathrm{SC}}) f_{\Gamma_{\mathrm{SC}}}(x)\mathrm{d}x \overset{(a)}{=} \tau \int_{\psi}^{\eta} F_{\Gamma_{\mathrm{SC}}}(x)\mathrm{d}x, \tag{A III-1}$$

where (a) is due to the partial integration method. Plugging Eq. (2.11) into Eq. (A III-1), we obtain

$$\begin{aligned}\bar{\epsilon}_{\mathrm{SC}} = 1 - \tau \sum_{m=0}^{Nm_{\mathrm{SC}}-1} \sum_{i=1}^{\varrho(\mathcal{A})} \sum_{j=1}^{\tau_i(\mathcal{A})} \frac{\chi_{i,j}(\mathcal{A})}{m!} \frac{a_{\mathrm{I}\langle i\rangle\mathrm{C}}}{(j-1)!} \\ \int_0^{\infty} x^{j-1} e^{-a_{\mathrm{I}\langle i\rangle\mathrm{C}}x}\mathrm{d}x \int_{\psi}^{\eta} \left(a_{\mathrm{SC}}\gamma\right)^m e^{-a_{\mathrm{SC}}(x+1)\gamma}\mathrm{d}\gamma.\end{aligned} \tag{A III-2}$$

Importantly, the second integral can be derived in the closed-form expression as shown below.

$$\begin{aligned}\int_{\psi}^{\eta} (\gamma)^m e^{-a_{\mathrm{SC}}(x+1)\gamma} d\gamma = \sum_{r=0}^{m} \frac{m!}{r!} \left(a_{\mathrm{SC}}(x+1)\right)^{r-1} \\ \left[\psi^r e^{-a_{\mathrm{SC}}(x+1)\psi} - \eta^r e^{-a_{\mathrm{SC}}(x+1)\eta}\right].\end{aligned} \tag{A III-3}$$

Hence, substituting the above results and after some mathematical manipulations, we obtain Eq. (2.30). This completes the proof of Theorem 1.

APPENDIX IV

PROOF OF THEOREM 2 - CHAPTER 2

From Eq. (2.28), we have $Z(\Gamma_{\mathrm{CD}_k}) \approx \frac{1}{2} - \tau\left(\Gamma_{\mathrm{CD}_k} - \omega\right)$ for $\psi < \Gamma_{\mathrm{CD}_k} < \eta$. Hence, the average BLER for the C-to-D$_k$ link, i.e., $\bar{\epsilon}_{\mathrm{CD}_k}$, is derived as $\bar{\epsilon}_{\mathrm{CD}_k} \approx \tau \int_{\psi}^{\eta} F_{\Gamma_{\mathrm{CD}_k}}(x)\mathrm{d}x$.

Now, from Eq. (2.17), we obtain the following:

$$\begin{aligned}\bar{\epsilon}_{\mathrm{CD}_k} = 1 - \tau &\sum_{m=0}^{Nm_{\mathrm{CD}_k}-1} \sum_{i_k=1}^{\varrho(\mathcal{B}_{\|})} \sum_{j_k=1}^{\tau_{ik}(\mathcal{B}_{\|})} \frac{\chi_{i_k,j_k}(\mathcal{B}_{\|})}{m_k!} \frac{a_{\mathrm{I}_{\langle i_k\rangle}\mathrm{C}}^{j_k}}{(j_k-1)!} \\ &\int_0^{\infty} x^{j_k-1} e^{-a_{\mathrm{I}_{\langle i_k\rangle}\mathrm{C}}x} \mathrm{d}x \int_{\psi}^{\eta} \left(a_{\mathrm{CD}_k}\gamma\right)^{m_k} e^{-a_{\mathrm{CD}_k}(x+1)\gamma} \mathrm{d}\gamma. \end{aligned} \tag{A IV-1}$$

where the second integral can be derived in closed-form as follows:

$$\begin{aligned}&\int_{\psi}^{\eta} (\gamma)^{m_k} e^{-a_{\mathrm{CD}_k}(x+1)\gamma} \mathrm{d}\gamma \\ &\qquad = \sum_{r_k=0}^{m} \frac{m_k!}{r_k!} \frac{\psi^{r_k} e^{-a_{\mathrm{CD}_k}(x+1)\psi} - \eta^{r_k} e^{-a_{\mathrm{CD}_k}(x+1)\eta}}{(a_{\mathrm{CD}_k}(x+1))^{1-r_k}}. \end{aligned} \tag{A IV-2}$$

Hence, substituting the above results and with the help of the Confluent hypergeometric function of the second kind, we obtain Eq. (2.32). This completes the proof of Theorem 2.

APPENDIX V

PROOF OF THEOREM 3 - CHAPTER 2

The average e2e BLER of the system can be approximated as $\bar{\epsilon}_{\text{e2e}} \approx \tau \int_{\psi}^{\eta} F_{\Gamma_{\text{sys}}}(\gamma)\text{d}\gamma$. Substituting $F_{\Gamma_{\text{sys}}}$ from Eq. (2.19), we obtain the following:

$$\bar{\epsilon}_{\text{e2e}} \approx 1 - \tau \int_{\psi}^{\eta} 1 - \Big[1 - F_{\Gamma_{\text{SC}}}(\gamma)\Big] \prod_{k \in \mathcal{K}} \Big[1 - F_{\Gamma_{\text{CD}_k}}(\gamma)\Big] \text{d}\gamma. \tag{A V-1}$$

Substituting the CDF of Γ_{SC} and Γ_{CD_k} and after some mathematical manipulations, we obtain $\bar{\epsilon}_{\text{e2e}}$ expressed in terms of $\Lambda_2 + \Lambda_3$, where

$$\begin{aligned} &\Lambda_2 + \Lambda_3 \\ &\quad = \int_{\psi}^{\eta} \frac{e^{\gamma(a_{\text{SC}} + \sum_{k=1}^{K} a_{\text{CD}_k})} \gamma^{m + \sum_{k=1}^{K} m_k}}{(a_{\text{SC}}\gamma + a_{\text{I}_{\langle i \rangle}\text{C}})^{r+j} \prod_{k=1}^{K} (a_{\text{CD}_k}\gamma + a_{\text{I}_{\langle i_k \rangle}\text{D}_k})^{r_k + j_k}} \text{d}\gamma. \end{aligned} \tag{A V-2}$$

Performing partial fraction decomposition yields Λ_2 and Λ_3 as shown in Eq. (2.36) and Eq. (2.38), where

$$\begin{aligned} \Lambda(x, a, b, c) &= \int_0^x \gamma^a e^{-b\gamma} (1 + c\gamma)^{-q} \text{d}\gamma \\ &= \frac{1}{(q-1)!} \int_0^x \gamma^a G_{1,1}^{1,1} \left[{}^{1-q}_{0} \middle| c\gamma \right] G_{0,1}^{1,0} \left[{}^{-}_{1} \middle| b\gamma \right] \text{d}\gamma. \end{aligned} \tag{A V-3}$$

We can rewrite $\Lambda(x, a, b, c)$ in terms of bivariate Meijiers's G-function as follows:

$$G^{p,q,s,r,t}_{A,[C,E],B,[D,F]}\left[\begin{array}{c|c} & \mathbf{a}_p, \mathbf{a}_a \\ x & \mathbf{c}_q, \mathbf{c}_c; \mathbf{e}_s, \mathbf{e}_E \\ y & \mathbf{b}_B \\ & \mathbf{d}_r, \mathbf{d}_D; \mathbf{f}_t, \mathbf{f}_F \end{array}\right] = \frac{1}{(2\pi i)^2}\int_{L_1}\int_{L_2}$$
$$\times \frac{\Gamma(1-\mathbf{a}_p+p+\eta)\Gamma(\mathbf{c}_q+p)}{\Gamma(1-\mathbf{c}_C-p)\Gamma(1-\mathbf{d}_D+p)\Gamma(1-\mathbf{e}_E-\eta)}$$
$$\times \frac{\Gamma(\mathbf{d}_r-p)\Gamma(\mathbf{e}_s+\eta)\Gamma(\mathbf{f}_t-\eta)}{\Gamma(1-\mathbf{f}_F+\eta)\Gamma(\mathbf{a}_A-p-\eta)\Gamma(\mathbf{b}_B+p+\eta)} x^p y^\eta \mathrm{d}p\mathrm{d}\eta, \tag{A V-4}$$

where $G[\cdot]$ is the Meijer G-function and $\Gamma(\cdot)$ denotes the Gamma function. The following identity holds

$$\int_0^x t^{\nu-1} G^{m,n}_{p,q}\left[\begin{smallmatrix}\alpha_n,\alpha_p \\ \beta_m,\beta_q\end{smallmatrix}\middle|\mu t\right] G^{M,N}_{P,Q}\left[\begin{smallmatrix}\gamma_N,\gamma_P \\ \delta_M,\delta_Q\end{smallmatrix}\middle|\omega t\right] \mathrm{d}t = x^\nu$$
$$G^{1,n,N,m,M}_{1,[n,N],1,[p+q-n,P+Q-N]}\left[\begin{array}{c|c} & 1-\nu \\ \mu x & 1-\alpha_n; 1-\gamma_N \\ \omega x & \nu+1 \\ & \beta_m, \beta_q, 1-\alpha_p; \delta_M, \delta_Q, 1-\gamma_P \end{array}\right]. \tag{A V-5}$$

Hence, $\bar{\epsilon}_{\mathrm{e2e}}$ can be expressed in terms of Meijer-G function of two variables (Mathai *et al.*, 1978), as given in Eq. (2.34).

BIBLIOGRAPHY

Aijaz, A., Zeadally, S., Shah, M. & Vasilakos, A. (2018). Short packet communication in 5G: Challenges and research directions. *IEEE Commun. Mag.*, 56(12), 186–192. doi: 10.1109/MCOM.2018.1800526.

Cai, Y., Li, W., Meng, X., Zheng, W., Chen, C. & Liang, Z. (2024). Adaptive contrastive learning based network latency prediction in 5G URLLC scenarios. *Computer Networks*, 240, 110185.

Chen, H., Abbas, R., Cheng, P., Shirvanimoghaddam, M., Hardjawana, W., Bao, W., Li, Y. & Vucetic, B. (2018). Ultra-reliable low latency cellular networks: Use cases, challenges and approaches. *IEEE Commun. Mag.*, 56(12), 119–125.

Chen, J., Zhang, L., Liang, Y.-C., Kang, X. & Zhang, R. (2019a). Resource allocation for wireless-powered IoT networks with short packet communication. *IEEE Trans. Wirel. Commun.*, 18(2), 1447–1461.

Chen, R., Li, C., Yan, S., Malaney, R. & Yuan, J. (2019b). Physical layer security for ultra-reliable and low-latency communications. *Wirel. Commun.*, 26(5), 6–11.

Chen, T., Li, W. & Shen, Y. (2024). Data-Driven Intelligent Control for 6G Wireless: A Comprehensive Survey. *IEEE Access*, 12, 145920-145950.

Coutinho, R. W. & Boukerche, A. (2021). Transfer learning for disruptive 5G-enabled industrial internet of things. *IEEE Trans. Ind. Inform.*, 18(6), 4000–4007.

Davey, J., Lorincz, G. & Gabor, L. (2024). Over-the-Air Deep Learning for Channel-Adaptive Wireless Systems. *IEEE Transactions on Wireless Communications*, 23(2), 832-846.

Glinskiy, K., Krasilov, A., Khorov, E. & Kureev, A. (2023). Performance of ML-Based Channel Prediction Algorithms for URLLC: Channel Model Matters. *2023 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 306-311. doi: 10.1109/BlackSeaCom58138.2023.10299788.

Gradshteyn, I. S., Bradshteyn, A. S. & Ryzhik, I. M. (2007). *Tables of Integrals, Series, and Products* (ed. 7th). New York, NY, USA: Academic Press.

Huang, X., Li, Y. & Sun, Z. (2024). Short-Packet Communications for Wireless Energy Transfer Systems Under URLLC Constraints. *IEEE Internet of Things Journal*, 11(5), 10390-10405.

Interdonato, G., Björnson, E. & Frenger, P. (2024). Finite Blocklength Analysis of Coexisting URLLC and eMBB Services in Cell-Free Massive MIMO. *IEEE Transactions on Wireless Communications*, 23(3), 2123-2138.

Iqbal, M., Kim, B. & Kim, B. (2025). Energy-Aware URLLC for 6G-Enabled Hybrid IoT Systems. *IEEE Transactions on Communications*. Early Access.

Ju, Y., Wang, H., Pei, Q. & Wang, H.-M. (2019). Physical layer security in millimeter wave DF relay systems. *IEEE Trans. Wirel. Commun.*, 18(12), 5719–5733.

Kumar, P. & Singh, A. (2024). Finite Blocklength URLLC Performance of IRS-Aided MIMO Systems. *IEEE Transactions on Wireless Communications*, 23(7), 5162-5176.

Lam, H., Nguyen, L. D. & Cao, X.-N. (2023). Short Packet Transmission for Ultra-Reliable and Low-Latency Communications: Analysis and Optimization. *IEEE Transactions on Vehicular Technology*, 72(10), 12845-12860.

Li, D., Yuan, H. & Hao, Y. (2022). Transfer Learning for Deep Signal Detection Under Time-Varying Channels. *Proc. IEEE Globecom*, pp. 1-6.

Liu, W., Zhang, R. & Renzo, M. (2023a). Machine Learning for Physical Layer in 6G Wireless Networks: A Survey. *IEEE Communications Surveys & Tutorials*, 25(4), 2671-2703.

Liu, Y., Deng, Y., Nallanathan, A. & Yuan, J. (2023b). Machine Learning for 6G Enhanced Ultra-Reliable and Low-Latency Services. *IEEE Wireless Communications*, 30(2), 48-54. doi: 10.1109/MWC.006.2200407.

Maghsoudnia, A., Dosti, E. & Durisi, G. (2024). Ultra-Reliable Low-Latency Communication: Theory vs. Practice. *IEEE Communications Magazine*, 62(1), 112-118.

Mao, Q., Hu, F. & Hao, Q. (2018). Deep learning for intelligent wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2595–2621.

Mathai, A. M., Arakaparampil, M. & Saxena, R. K. (1978). *The H-function with applications in statistics and other disciplines*. John Wiley & Sons.

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., Niyato, D., Dobre, O. & Poor, H. V. (2021). 6G Internet of Things: A comprehensive survey. *IEEE Internet Things J.*, 9(1), 359–383.

Nguyen, T.-V., Huynh-The, T. & Bao, V. N. Q. (2023). Performance Analysis and Deep Learning Evaluation of URLLC Full-Duplex Energy Harvesting IoT Networks over Nakagami-m Fading Channels. *2023 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 309–313.

Niu, S., Liu, Y., Wang, J. & Song, H. (2020). A Decade Survey of Transfer Learning (2010–2020). *IEEE Trans. Artif. Intell.*, 1(2), 151-166.

Oikonomou, T. K., Evgenidis, N. G., Nixarlidis, D. G., Tyrovolas, D., Tegos, S. A., Diamantoulakis, P. D., Sarigiannidis, P. G. & Karagiannidis, G. K. (2024). CNN-Based Automatic Modulation Classification Under Phase Imperfections. *IEEE Wireless Communications Letters*, 13(5), 1508-1512.

Pala, S., Katwe, M., Singh, K., Clerckx, B. & Li, C.-P. (2024). Spectral-Efficient RIS-Aided RSMA URLLC: Toward Mobile Broadband Reliable Low Latency Communication (mBRLLC) System. *IEEE Transactions on Wireless Communications*, 23(4), 3507-3524. doi: 10.1109/TWC.2023.3309028.

Peng, Q., Gilman, A., Vasconcelos, N., Cosman, P. C. & Milstein, L. B. (2019). Robust deep sensing through transfer learning in cognitive radio. *IEEE Wirel. Commun. Lett.*, 9(1), 38–41.

Prathyusha, Y. & Sheu, T.-L. (2022). Coordinated Resource Allocations for eMBB and URLLC in 5G Communication Networks. *IEEE Trans. Veh. Technol.*, 71(8), 8717-8728.

Shaik, F. & Talukdar, F. (2024). AI/ML-Driven Capacity Maximization Strategies for URLLC in 6G Networks. *IEEE Open Journal of the Communications Society*, 5, 234–250.

Sharma, N. & Agrawal, R. (2025). Digital Twin–Enabled Reliability Optimization for 6G URLLC. *IEEE Transactions on Industrial Informatics*. Early Access.

Tao, M. (2009). Effects of Non-Identical Rayleigh Fading on Differential Unitary Space-Time Modulation. *IEEE Trans. Commun.*, 57(5), 1359-1369. doi: 10.1109/TCOMM.2009.05.070534.

Tu, N. H., Hoang, T.-D. & Lee, K. (2023). Short-Packet URLLCs for MIMO Underlay Cognitive Multihop Relaying With Imperfect CSI. *IEEE Access*, 11, 81672-81689. doi: 10.1109/ACCESS.2023.3301139.

Uyoata, E., Sutherland, M. & Michael, S. (2024). Transfer Learning for Fully Convolutional Networks in Wireless Signal Classification. *IEEE Transactions on Signal Processing*, 72, 1332-1345.

Vaezi, M., Azari, A. & Khosravirad, S. (2022). Cellular, Wide-Area, and Non-Terrestrial IoT: A Survey on 5G Advances and the Road Toward 6G. *IEEE Commun. Surveys & Tut.*, 24(2), 1117-1174.

Wang, M., Lin, Y., Tian, Q. & Si, G. (2021). Transfer learning promotes 6G wireless communications: Recent advances and future challenges. *IEEE Trans. Rel.*, 70(2), 790–807.

Xu, L., Wang, J., Li, X., Cai, F., Tao, Y. & Gulliver, T. A. (2021). Performance Analysis and Prediction for Mobile Internet-of-Things (IoT) Networks: A CNN Approach. *IEEE Internet Things J.*, 8(17), 13355-13366.

Xu, L., Zhou, X., Tao, Y., Yu, X., Yu, M. & Khan, F. (2022). AF Relaying Secrecy Performance Prediction for 6G Mobile Communication Networks in Industry 5.0. *IEEE Trans. Ind. Inform.*, 18(8), 5485-5493.

Yuan, L., Yang, N., Fang, F. & Ding, Z. (2022). Performance Analysis of UAV-Assisted Short-Packet Cooperative Communications. *IEEE Trans. Veh. Technol.*, 71(4), 4471-4476.

Zhang, J., Chen, M., Wu, F. & Xu, T. (2024). Wireless Hierarchical Federated Learning with Ultra-Reliable and Low-Latency Constraints in 6G Networks. *IEEE Internet of Things Journal*, 11(8), 12530–12545. doi: 10.1109/JIOT.2024.3378921.

Zhang, S., Chen, X., Li, Y., Jiang, D., Zhang, Q. & Guizani, M. (2022). Advancements in short packet communication for 6G networks. *Wirel. Commun.*, 29(1), 136–143. doi: 10.1109/MWC.2022.3149321.

Zhang, Z., Wang, W. & Lyu, Y. (2023). Finite Blocklength Communications in Cell-Free Massive MIMO for Delay-Constrained Networks. *IEEE Transactions on Communications*, 71(11), 6159-6173.

Zheng, S., Chen, S., Qi, P., Zhou, H. & Yang, X. (2020). Spectrum sensing based on deep learning classification for cognitive radios. *China Communications*, 17(2), 138–148.

Zhou, K., Liu, Z., Qiao, Y., Xiang, T. & Loy, C. C. (2023). Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4), 4396-4415.