



An International ICT Journal
Featuring Industry-University-Institute Cooperation

ISSN 1673-5188
CN 34-1294/TN

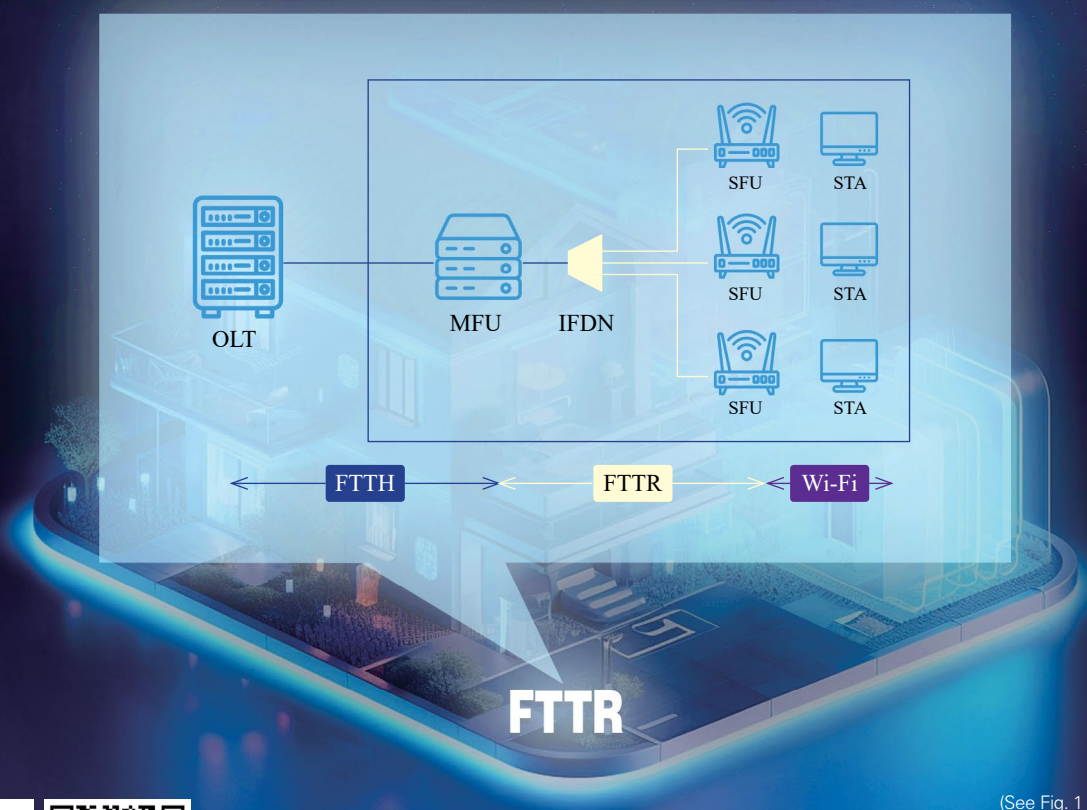
ZTE COMMUNICATIONS

中兴通讯技术(英文版)

<http://zte.magtechjournal.com>

December 2025, Vol. 23 No. 4

Special Topic: New Generation FTTR Communication and Networking Technology



ISSN 1673-5188



(See Fig. 1 on P. 50)

The 10th Editorial Board of ZTE Communications

Chairman

GAO Wen, Peking University (China)

Vice Chairmen

XU Ziyang, ZTE Corporation (China) | **XU Chengzhong**, University of Macau (China)

Members (Surname in Alphabetical Order)

AI Bo	Beijing Jiaotong University (China)
CAO Jiannong	The Hong Kong Polytechnic University (China)
CHEN Chang Wen	The Hong Kong Polytechnic University (China)
CHEN Yan	Northwestern University (USA)
CHI Nan	Fudan University (China)
CUI Shuguang	UC Davis (USA) and The Chinese University of Hong Kong, Shenzhen (China)
GAO Wen	Peking University (China)
GAO Yang	Nanjing University (China)
GAO Yue	Fudan University (China)
FANG Rong	ZTE Corporation (China)
GE Xiaohu	Huazhong University of Science and Technology (China)
HE Yejun	Shenzhen University (China)
Victor C. M. LEUNG	The University of British Columbia (Canada)
LI Xiangyang	University of Science and Technology of China (China)
LIAO Yong	Chongqing University (China)
LIN Xiaodong	ZTE Corporation (China)
LIU Chi	Beijing Institute of Technology (China)
LIU Jian	ZTE Corporation (China)
LIU Yue	Beijing Institute of Technology (China)
MA Jianhua	Hosei University (Japan)
MA Zheng	Southwest Jiaotong University (China)
PAN Yi	Shenzhen University of Advanced Technology, Chinese Academy of Sciences (China)
PENG Mugen	Beijing University of Posts and Telecommunications (China)
REN Fuji	Tokushima University (Japan)
REN Kui	Zhejiang University (China)
SHENG Min	Xidian University (China)
SU Zhou	Xi'an Jiaotong University (China)
SUN Huifang	Pengcheng Laboratory (China)
SUN Zhili	University of Surrey (UK)
TAO Meixia	Shanghai Jiao Tong University (China)
WANG Chengxiang	Southeast University (China)
WANG Haiming	Southeast University (China)
WANG Xiang	ZTE Corporation (China)
WANG Xiyu	ZTE Corporation (China)
WANG Yongjin	Nanjing University of Posts and Telecommunications (China)
XU Chengzhong	University of Macau (China)
XU Ziyang	ZTE Corporation (China)
YANG Kun	University of Essex (UK)
YU Hongfang	University of Electronic Science and Technology of China (China)
YUAN Jinhong	University of New South Wales (Australia)
ZENG Wenjun	Eastern Institute of Technology, Ningbo (China)
ZHANG Honggang	Macau University of Science and Technology (China)
ZHANG Jianhua	Beijing University of Posts and Telecommunications (China)
ZHANG Rui	The Chinese University of Hong Kong, Shenzhen (China)
ZHANG Wenqiang	Fudan University (China)
ZHANG Yueping	Nanyang Technological University (Singapore)
ZHOU Wanlei	City University of Macau (China)
ZHUANG Weihua	University of Waterloo (Canada)

CONTENTS

ZTE COMMUNICATIONS
December 2025 Vol. 23 No. 4 (Issue 93, End of Volume)

Special Topic ►	New Generation FTTR Communication and Networking Technology
01	Editorial GE Xiaohu, ZHONG Yi
03	PON Monitoring Scheme Based on TGD-OFDR with High Spatial Resolution and Dynamic Range ZHU Yidai, FAN Xinyu, ZHU Songlin, DONG Jiaxing, LI Guoqiang, HE Zuyuan
10	Insights on Next Generation WLAN: High Experiences (HEX) YANG Mao, LI Bo, YAN Zhongjiang
16	FTTR-MmWave Architecture for Next-Generation Indoor High-Speed Communications CHEN Zhe, ZHOU Peigen, WANG Long, HOU Debin, HU Yun, CHEN Jixin, HONG Wei
27	A Transformer-Based End-to-End Receiver Design for Wi-Fi 7 Physical Layer LIU Yichen, GAO Ruixin, ZENG Chen, LIU Yingzhuang
37	Root Cause Analysis of Poor FTTR Quality Based on Transformer Mechanisms YU Weichao, LIU Yang, ZHANG Junxiong, YE Junliang, GE Xiaohu
48	QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer for FTTR Networks CAI Jinhan, ZAN Mingyuan, SHEN Gangxiang
65	C-WAN for FTTR: Enabling Low-Overhead Joint Transmission with Deep Learning ZHANG Yang, CEN Zihan, ZHAN Wen, CHEN Xiang
Research Papers ►	77 Empowering Grounding DINO with MoE: An End-to-End Framework for Cross-Domain Few-Shot Object Detection DONG Xiugang, ZHANG Kaijin, NONG Qingpeng, JU Minhan, TU Yaofeng
86	Shortened PAC Codes and List Decoding LIU Aolin, FENG Bowen, LIANG Chulong, XU Jin, ZHANG Qinyu
97	Full-Duplex Massive MIMO Self-Interference Suppression Based on Beamforming ZHANG Boyu, ZHANG Ling, LI Zijing, SHEN Ying
110	A Root Cause Analysis Framework for Microservice Systems with Multimodal Data LI Yingke, HAN Jing, SUN Yongqian, SHI Binpeng, GONG Zican
Roundup ►	I Table of Contents for Volume 23, 2025

Serial parameters: CN 34-1294/TN*2003*q*16*122*en*P*¥30.00*2200*12* 2025-12

Submission of a manuscript implies that the submitted work has not been published before (except as part of a thesis or lecture note or report or in the form of an abstract); that it is not under consideration for publication elsewhere; that its publication has been approved by all co-authors as well as by the authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors hand over the transferable copyrights of the accepted manuscript to *ZTE Communications*; and that the manuscript or parts thereof will not be published elsewhere in any language without the consent of the copyright holder. Copyrights include, without spatial or timely limitation, the mechanical, electronic and visual reproduction and distribution; electronic storage and retrieval; and all other forms of electronic publication or any other types of publication including all subsidiary rights.

Responsibility for content rests on authors of signed articles and not on the editorial board of *ZTE Communications* or its sponsors.

Statement This magazine is a free publication for you. If you do not want to receive it in the future, you can send the “TD unsubscribe” mail to magazine@zte.com.cn. We will not send you this magazine again after receiving your email. Thank you for your support.

ZTE Communications Guidelines for Authors

Remit of Journal

ZTE Communications publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

Manuscript Preparation

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 3 000 to 8 000, and no more than 8 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

Abstract and Keywords

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Three to eight carefully chosen keywords must be provided with the abstract.

References

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to *ZTE Communications* Editorial Style. A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

Copyright and Declaration

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of *ZTE Communications*. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors; b) the manuscript has not been published elsewhere in its submitted form; c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

Content and Structure

ZTE Communications seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

Peer Review and Editing

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

Biographical Information

All authors are requested to provide a brief biography (approx. 100 words) that includes email address, educational background, career experience, research interests, awards, and publications.

Acknowledgments and Funding

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

Address for Submission

<http://mc03.manuscriptcentral.com/ztecom>



Special Topic on New Generation FTTR Communication and Networking Technology

Guest Editors



 GE Xiaohu



 ZHONG Yi

The convergence of optical and wireless technologies is driving the evolution of intelligent indoor networks, with Fiber-to-the-Room (FTTR) emerging as a key architecture for delivering gigabit connectivity in both home and enterprise environments. By deploying optical fiber directly to rooms and integrating it with advanced wireless solutions such as millimeter-wave and Wi-Fi 7, FTTR enables next-generation applications, including immersive Virtual Reality (VR)/Augmented Reality (AR) and industrial Internet of Things (IoT). Nevertheless, its large-scale deployment presents challenges in network management, energy efficiency, interference mitigation, and intelligent root cause analysis.

Following a rigorous review process, seven high-quality papers have been selected for this special issue. These works advance FTTR technology through contributions spanning physical layer monitoring, receiver design, energy-saving mechanisms, and artificial intelligence (AI)-driven optimization.

The first paper, “PON Monitoring Scheme Based on TGD-OFDR with High Spatial Resolution and Dynamic Range,” introduces a time-gated digital optical frequency-domain reflectometry (TGD-OFDR) system that decouples spatial resolution from pulse width. This breakthrough achieves both 0.3 m resolution and 30 dB dynamic range, substantially enhancing passive optical network (PON) monitoring capabilities.

The second paper, “Insights on Next Generation WLAN: High Experiences (HEX),” identifies poor quality of experience (QoE) as the primary challenge in current wireless local area network (WLAN) systems. The authors propose making high experiences (HEX) the core objective for next-generation networks through systematic architectural improvements.

The third paper, “FTTR-MmWave Architecture for Next Generation Indoor High-Speed Communications,” proposes a hybrid FTTR-millimeter-wave architecture, leveraging mmWave’s natural isolation to reduce interference while supporting multi-gigabit rates for ultra-high-speed indoor communications.

The fourth paper, “A Transformer-Based End-to-End Receiver Design for Wi-Fi 7 Physical Layer,” proposes a Transformer-based receiver for Wi-Fi 7 that directly decodes Orthogonal Frequency Division Multiplexing (OFDM) signals without channel estimation, achieving significant performance gains over conventional designs.

The fifth paper, “Root Cause Analysis of Poor FTTR Quality Based on Transformer Mechanisms,” addresses network maintenance challenges in FTTR deployments. The authors develop a multi-task Transformer model that achieves an accuracy of 96.75% in identifying degraded access points and classifying root causes, outperforming traditional methods significantly.

The sixth paper, “QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer for FTTR Networks,” tackles energy efficiency in dense FTTR deployments. The proposed Multi-threshold Buffer Energy Saving (MBES) scheme introduces dynamic buffer thresholds to enable intelligent sleep cycle management while maintaining strict QoS requirements, achieving energy savings of up to 17.75%.

DOI:10.12142/ZTECOM.202504001

Citation (Format 1): GE X H and ZHONG Y. Editorial: new generation FTTR communication and networking technology [J]. ZTE Communications, 2025, 23(4), pp. 1–2. doi: 10.12142/ZTECOM.202504001

Citation (Format 2): X. H. Ge and Y. Zhong, “Editorial: new generation FTTR communication and networking technology,” *ZTE Communications*, vol. 23, no. 4, pp. 1–2, Dec. 2025. doi: 10.12142/ZTECOM.202504001.

The seventh paper, “C-WAN for FTTR: Enabling Low-Overhead Joint Transmission with Deep Learning,” presents a centralized wireless access network solution that addresses synchronization and channel state information overhead in multi-access point (AP) coordination. The proposed deep learning model reduces sounding overhead while maintaining transmission performance.

In conclusion, this special issue provides comprehensive coverage of recent advances in FTTR and integrated fiber-wireless systems. We believe that these contributions will serve as valuable references for researchers, engineers, and industry professionals working toward next-generation communication networks. We extend our sincere gratitude to the authors, reviewers, and editorial team members for their dedicated efforts in making this special issue possible.

Biographies

GE Xiaohu is currently a full professor with the School of Electronic Information and Communications at Huazhong University of Science and Technology (HUST), China. He is also affiliated with the China International Joint Research Center of Green Communications and Networking at HUST. He serves as an Ad-

junct Professor with the Faculty of Engineering and Information Technology at University of Technology Sydney (UTS), Australia. He received his PhD degree in communication and information engineering from HUST in 2003. He has worked at HUST since Nov. 2005. Prior to that, he was a researcher at Ajou University, Republic of Korea, and Politecnico Di Torino, Italy from January 2004 to October 2005. His research interests include mobile communications, traffic modeling in wireless networks, green communications, and interference modeling in wireless communications. He has published more than 200 papers in refereed journals and conference proceedings and has been granted about 25 patents in China. He received the Best Paper Awards at IEEE GlobeCom 2010. He served as the General Chair for the 2015 IEEE International Conference on Green Computing and Communications (IEEE GreenCom 2015). He also serves as an associate editor for *IEEE Wireless Communications* and *IEEE Transactions on Vehicular Technology*. He is a Senior Member of the IEEE.

ZHONG Yi received his BS and PhD degrees in electronic engineering from the University of Science and Technology of China (USTC) in 2010 and 2015, respectively. From 2015 to 2016, he was a Postdoctoral Research Fellow at the Singapore University of Technology and Design (SUTD), where he worked with the Wireless Networks and Decision Systems (WNDS) Group. He is currently an associate professor with the School of Electronic Information and Communications at Huazhong University of Science and Technology (HUST), China. He served as an associate editor for *IEEE Wireless Communications Letters* and other academic journals. His research interests lie in advanced wireless network theory, with a focus on spatio-temporal modeling, stochastic geometry, spatial network calculus, and energy-efficient communications in next-generation networks. He is a Senior Member of the IEEE.



PON Monitoring Scheme Based on TGD-OFDR with High Spatial Resolution and Dynamic Range

ZHU Yidai¹, FAN Xinyu², ZHU Songlin¹, DONG Jiaxing²,
LI Guoqiang¹, HE Zuyuan²

(1. Wireline Product Planning Department, ZTE Corporation, Shanghai 201203, China;

2. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

DOI: 10.12142/ZTECOM.202504002

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251127.1332.002.html>,
published online November 27, 2025

Manuscript received: 2025-08-15

Abstract: Conventional optical time-domain reflectometry (OTDR) schemes for passive optical network (PON) link monitoring are limited by insufficient dynamic range and spatial resolution. The expansion of PONs, with increasing optical network units (ONUs) and cascaded splitters, imposes even more stringent demands on the dynamic range of monitoring systems. To address these challenges, we propose a time-gated digital optical frequency-domain reflectometry (TGD-OFDR) system for PON monitoring that effectively decouples the inherent coupling between spatial resolution and pulse width. The proposed system achieves both high spatial resolution (~0.3 m) and high dynamic range (~30 dB) simultaneously, marking a significant advancement in optical link monitoring.

Keywords: passive optical network (PON); time-gated digital optical frequency-domain reflectometry (TGD-OFDR); dynamic range; spatial resolution

Citation (Format 1): ZHU Y D, FAN X Y, ZHU S L, et al. PON monitoring scheme based on TGD-OFDR with high spatial resolution and dynamic range [J]. *ZTE Communications*, 2025, 23(4): 3 – 9. DOI: 10.12142/ZTECOM.202504002

Citation (Format 2): Y. D. Zhu, X. Y. Fan, S. L. Zhu, et al., “PON monitoring scheme based on TGD-OFDR with high spatial resolution and dynamic range,” *ZTE Communications*, vol. 23, no. 4, pp. 3 – 9, Dec. 2025. doi: 10.12142/ZTECOM.202504002.

1 Introduction

Optical fiber, serving as a key transmission medium in communication links, is widely deployed in data communication networks owing to its advantages, including large bandwidth, low loss, strong anti-interference capability, low cost, material abundance, and longevity. Its deployment in backbone and metropolitan area networks has significantly increased data bandwidth while reducing operational costs.

Leveraging the abundant broadband resources of optical fiber, passive optical network (PON) technology has gained prominence since its inception. As a point-to-multipoint (P2MP) optical access technology, PON comprises an optical line terminal (OLT) at the central office, optical network units (ONUs) at user premises, and an optical distribution network (ODN). To ensure reliable operation of diverse PON services and enhance network stability, online monitoring of each PON link status is essential. This enables the prediction of potential optical link issues and facilitates timely corrective actions to

guarantee service quality.

Traditional link monitoring primarily relies on optical time-domain reflectometry (OTDR)^[1]. By launching optical pulses and analyzing the time-domain Rayleigh backscattered signal intensity, OTDR assesses link conditions, including attenuation, loss, and breaks. While achieving detection ranges of tens of kilometers, OTDR typically provides meter-scale spatial resolution.

However, a fundamental trade-off exists in OTDR between dynamic range and spatial resolution, constrained by the optical pulse width. The optical splitters inherent in PON architecture introduce additional losses, demanding a higher system dynamic range. Furthermore, the meter-scale spatial resolution of OTDR poses significant limitations in PONs with numerous terminals. These limitations impede the development of advanced PON monitoring solutions.

Several approaches have been proposed to address these challenges. For instance, tunable-wavelength OTDR^[2] assigns distinct detection wavelengths to different ONUs. Alternatively, pulse coding techniques^[3] assign unique codes to branch links and increase the effective probe power, thereby boosting the system dynamic range. Continuous-wave optical frequency-

This work was supported by ZTE and Shanghai Jiao Tong University Collaborative Laboratory under Grant No. IA20241205014.

domain reflectometry (OFDR) enables high-spatial-resolution PON monitoring by utilizing the unique Rayleigh backscatter spectral “fingerprint” of individual fibers beyond the splitter^[4]. This approach offers high precision and can discern unique Rayleigh characteristics amidst overlapping signals from multiple fibers, serving as a powerful tool for real-time monitoring and individual event detection in PON. However, these studies^[2–4] still fail to achieve simultaneous long-distance reach, high dynamic range, and high spatial resolution^[5]. Furthermore, the number of ONUs that can be effectively monitored by these methods is limited. As the splitting ratio and number of cascaded splitters increase, the resulting higher insertion loss imposes stricter demands on the system dynamic range.

In a typical PON, the optical link needs to span up to 30 km. As the network gradually expands and the number of users increases, the use of two-stage 1:8 optical splitters to connect 64 ONUs has become a standard configuration. Considering such a PON structure featuring an optical fiber link exceeding 30 km and incorporating two stages of 1:8 optical splitters, as depicted in Fig. 1, a dynamic range of at least 26 dB is essential. This budget comprises 20 dB for the two-stage splitters and 6 dB for fiber transmission loss (0.2 dB/km @1 550 nm). Additionally, a spatial resolution finer than 0.3 m is required for detailed event monitoring. The high-spatial-resolution requirement renders OTDR inadequate. Conversely, long-distance OFDR detection is plagued by phase noise, limiting its achievable dynamic range. To meet these demanding requirements, we introduce a time-gated digital optical frequency-domain reflectometry (TGD-OFDR) system into PON. TGD-OFDR enables a long-distance link monitoring scheme achieving simultaneous high dynamic range and high spatial resolution. Since TGD-OFDR was first proposed by LIU et al. in 2015^[6], many studies have been undertaken with the aim of improving the performance of the system. XIAO et al. employed a commercial integrated tunable laser assembly (ITLA) with a 100 kHz linewidth and introduced a frequency-domain reciprocal distortion phase noise compensation (FDRD-PNC) algorithm to improve the signal-to-noise ratio^[7]. Consequently, a dynamic range of 21.8 dB and a spatial resolution of 3.48 m were achieved over a 50 km measurement fiber. Moreover, LUO et al. proposed a novel frequency-sweeping method by incorporating a frequency-shifting loop into the system to mitigate fading noise^[8]. This scheme effectively suppressed fading noise without compromising spatial resolution, thereby enabling high-performance acoustic wave detection with high-frequency response. A spatial resolution of 1.43 m was achieved. For the inspection of PON structures, FONTAINE et al. employed a polarization-diversity coherent OFDR system to identify fibers and detect events in a 1:32 split PON^[9]. They characterized an ODN spanning 21 km

with a 1:32 split ratio, achieving a spatial resolution of less than 0.5 m and a dynamic range exceeding 35 dB.

In this study, we first provide a brief introduction to the principles of the TGD-OFDR system and the methodology for constructing a noise-incorporated simulation model. Subsequently, we establish the simulation model and analyze the impact of key parameters, including laser linewidth, pulse width, and splitting ratio, on the system’s dynamic range.

Through simulations conducted under a 100 kHz laser linewidth condition, we demonstrate that the TGD-OFDR system meets the target specifications of 30 dB dynamic range and 0.3 m spatial resolution, validating its feasibility for high-precision distributed sensing applications.

2 Theory

TGD-OFDR launches a long-pulse frequency-swept probe into the fiber. By demodulating the returned Rayleigh backscatter signal via digital-domain heterodyne detection, it decouples spatial resolution from pulse width and mitigates the limiting impact of phase noise on the dynamic range. This section briefly introduces the basic principle of TGD-OFDR and establishes a preliminary simulation model incorporating phase noise.

Fig. 2 illustrates the basic structure of a TGD-OFDR system. An arbitrary waveform generator (AWG) generates a linear frequency-swept radio frequency (RF) pulse to drive an optical

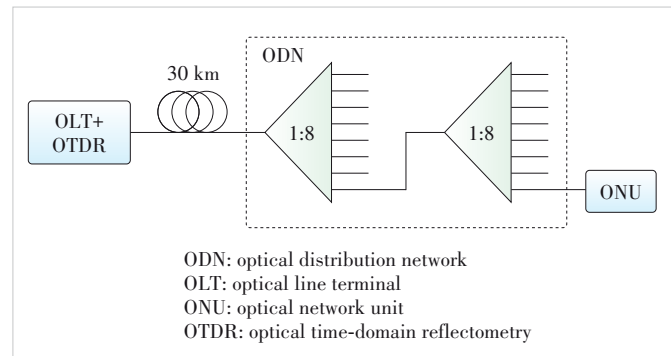


Figure 1. Structure of a typical passive optical network (PON)

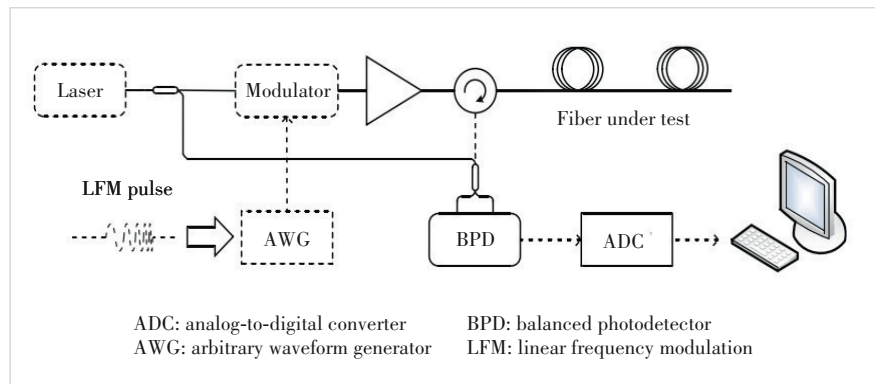


Figure 2. Basic structure of a time-gated digital optical frequency-domain reflectometry (TGD-OFDR) system

modulator, producing a linear frequency-swept optical pulse. The AWG-generated electrical signal is:

$$s(t) = W\left(\frac{t}{T_p}\right) e^{j[2\pi f_0 t + \pi \kappa t^2]} \quad (1),$$

where $W(\cdot)$ is the window function controlling the pulse shape over the pulse duration $[0, T_p]$, f_0 is the RF starting frequency, and κ is the frequency sweep rate. The resulting modulated optical field is:

$$E_p(t) = \sqrt{P_p} W\left(\frac{t}{T_p}\right) e^{j[\omega_c t + 2\pi f_0 t + \pi \kappa t^2]} \quad (2),$$

where P_p and ω_c are the power and frequency of the probe light (continuous-wave carrier), respectively. After Rayleigh backscattering occurs along the fiber, the total backscattered optical field returning to the receiver is:

$$E_R(t) = \sum_{i=1}^N E_i W\left(\frac{t - \tau_i}{T_p}\right) e^{j[\omega_c(t - \tau_i) + 2\pi f_0(t - \tau_i) + \pi \kappa(t - \tau_i)^2]} \quad (3),$$

where N is the total number of scattering points, E_i is the amplitude of the field from the i -th point (determined by the local Rayleigh scattering coefficient and accumulated fiber transmission loss up to that point), and τ_i is the round-trip time delay to the i -th scattering point. The returned backscattered light is combined with a portion of the original probe light (local oscillator, LO) and detected by a balanced photodetector (BPD). The BPD converts the optical interference into an electrical signal:

$$i(t) = \sum_{i=1}^N A_i W\left(\frac{t - \tau_i}{T_p}\right) \sin[2\pi f_0(t - \tau_i) + \pi \kappa(t - \tau_i)^2 - \omega_c \tau_i] \quad (4),$$

where A_i is the amplitude converted by the BPD.

A continuous digital replica of the frequency-swept signal is generated. This digital signal is mixed (multiplied) with the acquired electrical signal as shown in Eq. (4). This mixing process effectively down-converts the linear frequency-swept components into distance-dependent single-frequency signals. The sweep rate of this digital replica matches that of the optical probe signal, as illustrated in Fig. 3. The resulting beat frequency for each scattering point is directly proportional to its distance along the fiber. The amplitude of the signal component at frequency f_i reflects the local Rayleigh scattering coefficient at point i and the accumulated transmission loss from the beginning of the fiber to that point. Finally, applying a Fourier transform to the down-converted

digital signal yields a frequency spectrum. The amplitude and frequency of each peak in this spectrum correspond to the scattering strength and location, respectively, of each point along the optical fiber, thus providing the distributed Rayleigh backscatter profile.

The preceding derivation neglects phase noise. However, for practical simulation, time discretization is required. Consequently, we establish the simulation model as follows. The sensing fiber is divided into N discrete scattering units, each of length L_w . Computational constraints prevent setting L_w to be extremely small (i.e., much smaller than the incident optical wavelength λ). However, provided L_w is much smaller than the spatial extent (coverage length) of the detection pulse, the statistical properties of the Rayleigh backscatter remain valid. Owing to inherent non-uniformities in the fiber drawing process and potential variations in temperature and strain along the fiber, the refractive index n_i within each scattering unit i is not constant. The average core refractive index is typically 1.45. Each scattering unit is represented by a single effective scattering point, from which all scattered light originating within that unit is assumed to emanate. The position of the scattering point within its unit is defined by an offset δL_{wi} , measured from the unit's input end. These offsets are independent and identically distributed (i.i.d.) random variables, each uniformly distributed over $[0, L_w]$. The Rayleigh scattering coefficient for each scattering point is also modeled as an i.i.d. random variable with a uniform distribution. The round-trip time delay τ_i for light propagating from the fiber input end to the i -th scattering point and back is given by:

$$\tau_i = \frac{L_w \sum_{x=1}^{i-1} n_{1x} + \delta L_{wi} n_{1i}}{c} \quad (5).$$

Therefore, the backscattered field in Eq. (3) can be expressed as:

$$E_R(t) = \sum_{i=0}^{N-1} \sqrt{\gamma_i} \sqrt{\alpha_i} E_p(t - \tau_i) \quad (6),$$

where α_i is the loss coefficient of the scattered light from the

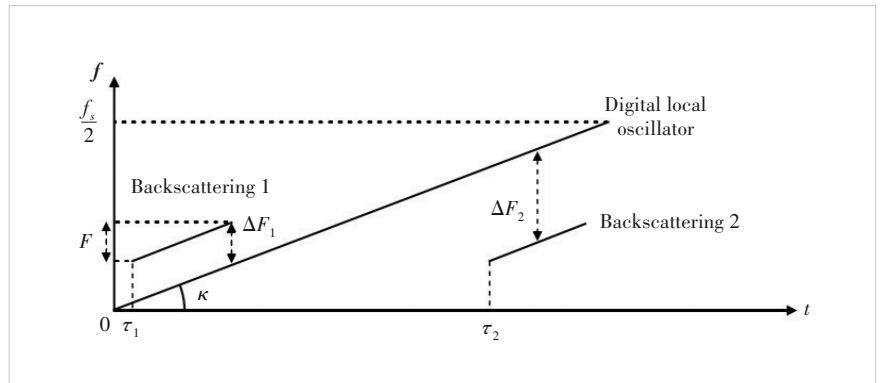


Figure 3. Schematic diagram of digital domain frequency-swept signal

i -th point along the entire optical path:

$$\alpha_i = \left(10^{\frac{L_w \alpha}{10}}\right)^{2i} \quad (7)$$

where α is the unit loss of optical fiber, normally taken as 0.2 dB/km@1 550 nm.

The phase noise $\phi(k)$, arising from the laser's finite linewidth, is modeled as a Wiener process affecting the instantaneous electric field of the probe light. Its increment $\phi(k - k_2) - \phi(k - k_1)$ follows a normal distribution with a mean value of 0 and a variance of $2\pi\Delta\nu(k_2 - k_1)T_s$, where $\Delta\nu$ is the linewidth. By generating k independent and identically distributed random variables Φ_n following $N(0, 2\pi\Delta\nu T_s)$ and accumulating them, the phase noise can be obtained:

$$\phi(k) = \sum_{n=1}^k \Phi_n \quad (8)$$

By substituting the phase noise into Eqs. (2) and (3), and revising Eq. (4) according to the distances of scattering points in the discrete case, we obtain:

$$i(k) = \eta(k) + R \cdot \sum_{i=0}^{N-1} \sqrt{\gamma_i} \sqrt{\alpha_i} \cdot s(k - \tau_i) e^{j[-\omega_s \tau_i + \phi(k - \tau_i) - \phi(k)]} \quad (9)$$

where $R = 2G \cdot R \cdot \sqrt{P_L} \sqrt{P_P}$, in which G is the gain of the BPD. Since the BPD gain amplifies both the signal and noise equally, it does not affect the signal-to-noise ratio (SNR). Therefore, in the simulation, we set the BPD gain to unity (i.e., 1) for simplified analysis without loss of generality. R is the responsivity of the BPD, which is set to 0.9 A/W in the simulation. The system noise, denoted by η , incorporates the effects of thermal noise, shot noise and amplified spontaneous emission (ASE) noise. Eq. (9) can be rewritten in the form of convolution^[10]:

$$i(k) = \eta(k) + R \cdot [s'(k) \otimes h(k)] \cdot e^{-j\phi(k)} \quad (10)$$

where $s'(k) = s(k) \cdot e^{j\phi(k)}$, which implies that the phase noise of the laser is imposed on the phase of the electrical signal; $h(k)$ represents the impulse response of the optical fiber:

$$h(k) = \sqrt{\gamma_k} \sqrt{\alpha_k} e^{j[-\omega_s \cdot \frac{L_w \sum_{i=1}^{k-1} n_{1i} + \delta L_{at} n_{1k}}{c}]} \quad (11)$$

First, the system bandwidth B is set according to the total sampling rate of the simulation model and its corresponding sampling time interval T_s . The length of the optical fiber scattering unit L_w is defined as $cT_s/2n$. The total fiber length is set, and the total number of scattering points N is calculated. Finally, the position distribution, scattering rate distribution and attenuation distribution

of the scattering points are defined according to Eqs. (5) and (7). These distributions are substituted into Eq. (11) to obtain the optical fiber impulse response $h(k)$.

Subsequently, we generate the frequency-swept pulsed light. First, a single-frequency continuous light is generated according to the incident light wavelength and the speed of light. According to Eq. (8), the phase noise corresponding to the total duration of the optical signal is generated and imposed on the phase of the optical signal. Then we generate the corresponding window signal mask $W(\cdot)$ based on the rise time, fall time, and acousto-optic modulator (AOM) extinction ratio. The above information is substituted into Eq. (1) to generate the frequency-swept electrical signal $s(k)$.

Finally, we convolve the frequency-swept electrical signal $s(k)$ with the optical fiber impulse response $h(k)$, and calculate the amplitude coefficient R resulting from the optical power, PD gain, and responsivity. Then, according to Eq. (10), we superimpose the phase noise term and other noise terms $\eta(k)$, and finally obtain the photocurrent from coherent reception.

3 Simulation

In conventional TGD-OFDR systems, the frequency-swept pulse is directly applied to the AOM. However, due to the limitations in the operating frequency and bandwidth of the AOM, the frequency sweep range of the system cannot exceed hundreds of MHz. To achieve a wide-range frequency sweep, we have modified the system architecture as illustrated in Fig. 4. A continuous frequency-swept optical signal is modulated via an in-phase/quadrature (IQ) modulator, followed by an injection-locked scheme for amplification and filtering. Subsequently, the AOM is utilized to extract the pulsed signal. After being amplified by the Erbium-doped fiber amplifier (EDFA), the probe signal is injected into the fiber under test for sensing, and is received with the local signal through a polarization diversity receiver (PDR). In this scheme, the AOM functions solely as a high-speed optical switch. The overall frequency sweep range of the system is determined by the bandwidth of the modulator and the performance of the AWG, thereby substantially elevating the upper limit of the sweep range.

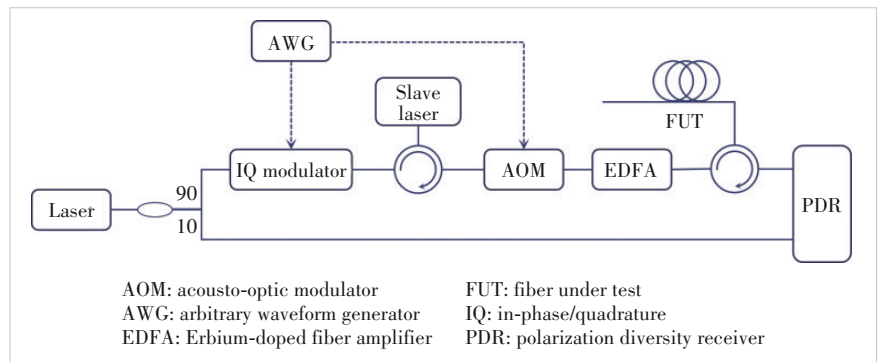


Figure 4. Structure of a simulated time-gated digital optical frequency-domain reflectometry (TGD-OFDR) system

The simulation platform was developed within the MATLAB environment. The key simulation parameters are shown in Table 1.

For computational efficiency in the simulation, the fiber link was first modeled as three concatenated 1 km fiber segments with 10 dB insertion loss at each splicing point to emulate the power attenuation caused by optical splitters in practical deployments; 100 independent measurements were averaged to enhance system performance. Using these parameters, the dynamic range performance was simulated for different laser linewidths (100 Hz to 1 MHz), pulse widths (10 μ s to 100 μ s), and splitting ratios (1:99, 10:90, and 50:50), respectively.

1) Linewidth

The incident optical power was fixed at 0 dBm and the pulse width at 100 μ s. The laser linewidth was varied sequentially from 100 Hz to 1 MHz, increasing by a factor of ten at each step. Representative Rayleigh spectra are shown in Fig. 5. The corresponding dynamic ranges are compiled in Table 2. It is evident that a tenfold increase in laser linewidth results in an approximate 5 dB decrease in the dynamic range.

2) Pulse width

The laser linewidth was fixed at 1 kHz, the source optical power at 0 dBm, and the splitting ratio between the LO path and the sensing path at 10:90. The pulse width was varied from 10 μ s to 100 μ s. Rayleigh spectra obtained for different pulse widths are shown in Fig. 6. The resulting dy-

Table 1. Key parameters for the proposed TGD-OFDR simulation model

Parameter	Value
Sampling rate	8 GS/s
Wavelength	1 550 nm
Speed of light	299 792 458 m/s
Extinction ratio (AOM)	40 dB
Operating frequency (AOM)	80 MHz
Gain (EDFA)	30 dB
Detection optical power	0 dBm
Fiber attenuation	0.2 dB/km

AOM: acousto-optic modulator
EDFA: Erbium-doped fiber amplifier
TGD-OFDR: time-gated digital optical frequency-domain reflectometry

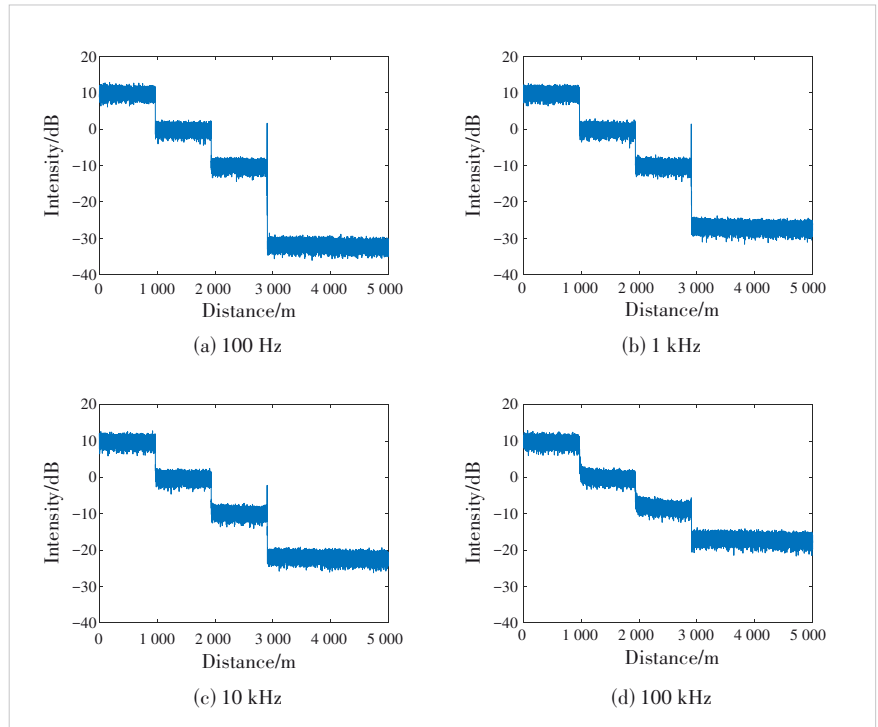


Figure 5. Simulated Rayleigh spectra with different laser linewidths

Table 2. Statistics of dynamic ranges with different linewidths

Linewidth	100 Hz	1 kHz	10 kHz	100 kHz	1 MHz
Dynamic range/dB	44.527 2	39.921 2	34.619 2	29.396 1	23.841 6

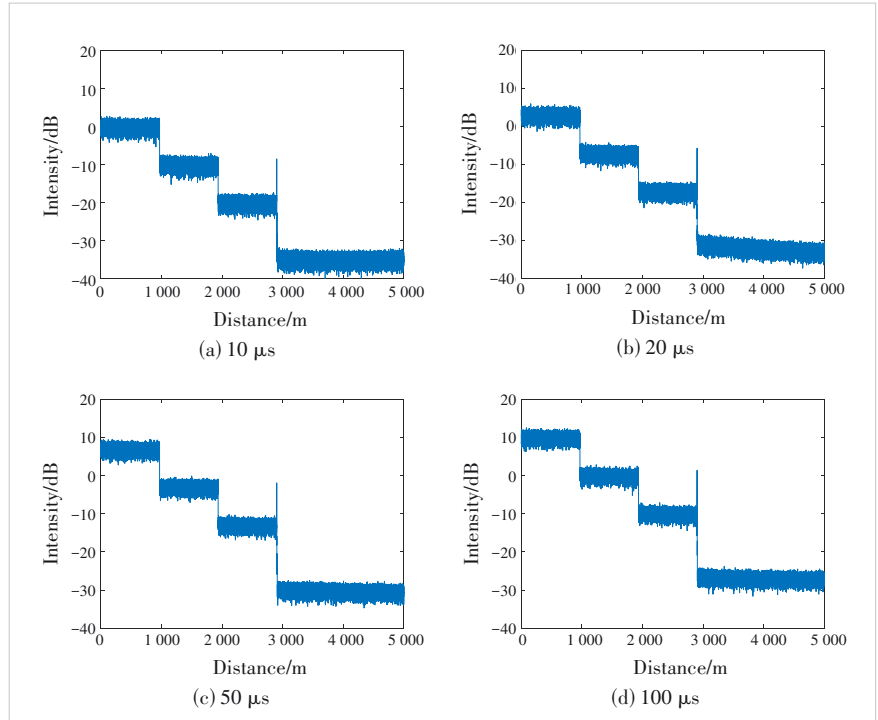


Figure 6. Simulated Rayleigh spectra with different pulse widths

Table 3. Statistics of dynamic ranges with different pulse widths

Pulse width/ μ s	10	20	50	100
Dynamic range/dB	33.014 6	35.234 8	37.152 3	39.921 2

dynamic ranges are compiled in Table 3. As the pulse width increases, the system's dynamic range improves. This is because TGD-OFDR shares similarities with OTDR, which utilizes pulsed light for detection; thus, increasing the pulse power enhances the dynamic range.

3) Splitting ratio

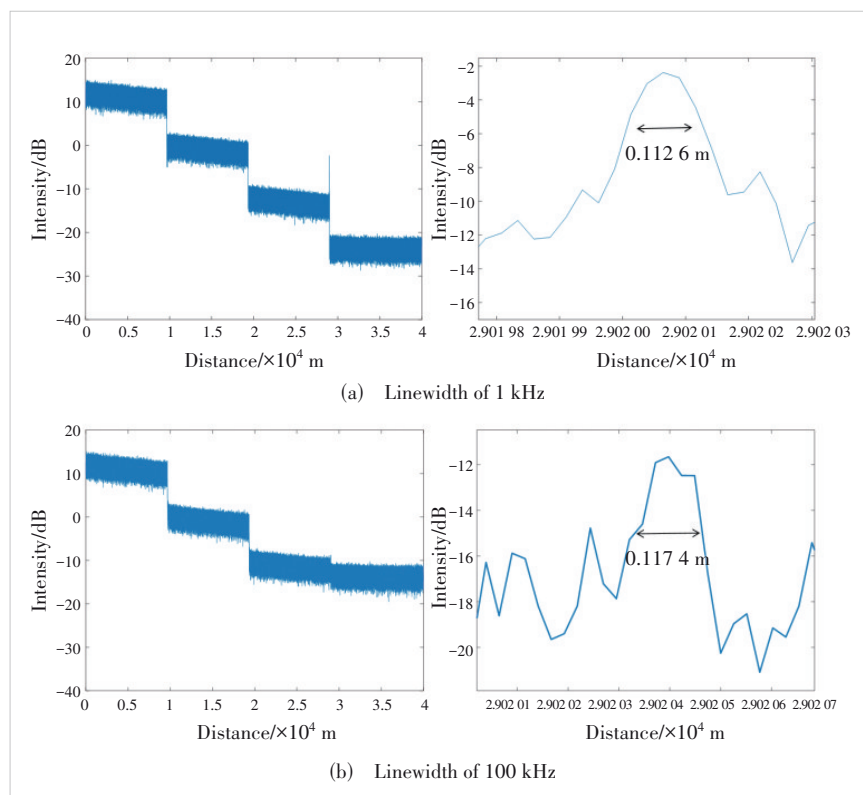
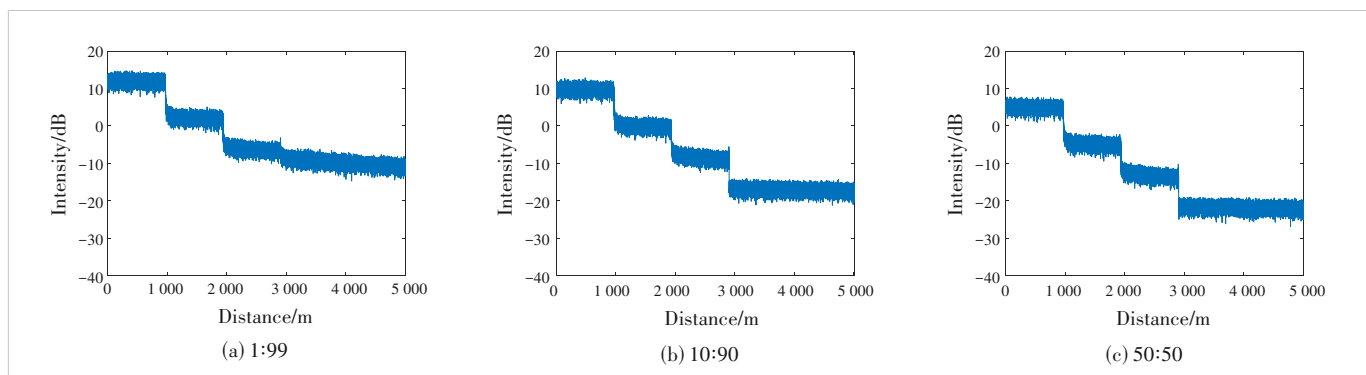
Adjusting the splitting ratio can effectively increase the LO power, thereby improving the system SNR. With the incident optical power fixed at 0 dBm, the dynamic range was simulated for different splitting ratios. The results are shown in Fig. 7. Increasing the splitting ratio (directing more power to the LO) significantly enhances the dynamic range by increasing the LO power. Common splitting ratios achievable with standard couplers are listed in Table 4. However, the improvement in this dynamic range is limited. When the LO power becomes excessive, the shot noise it generates begins to dominate over the thermal noise, and the system SNR becomes limited solely by the sensing path power. Conversely, increasing the splitting ratio reduces the sensing path power, establishing a trade-off between the LO power and sensing path power. Therefore, the simulations indicate that a splitting ratio of 10:90 yields the optimal dynamic range.

Finally, simulations were conducted with the following optimized parameters: an incident optical power of 10 dBm, a splitting ratio of 1:9, a pulse width of 100 μ s, and 100 averages. We extended the fiber under test to three concatenated 10 km fiber segments with parameters closer to those of a real-world

PON, while maintaining the two-stage splitter configuration. The simulation results are shown in Fig. 8. The total fiber length of 30 km corresponds to an attenuation loss of 6 dB. Combined with the 20 dB transmission loss of the two-stage splitters, the overall link loss reaches 26 dB. The pulse width is set to 100 μ s, and the sweep range is set from 100 MHz to 1 100 MHz. Under the condition of linewidths of 1 kHz and 100 kHz, the total dynamic

Table 4. Statistics of dynamic ranges with different splitting ratios

Splitting ratio	1:99	10:90	50:50
Dynamic range/dB	26.004 2	30.182 8	31.477 6

**Figure 8. Simulated Rayleigh spectra of a 30 km fiber under test with different laser linewidths and calculated spatial resolution****Figure 7. Simulated Rayleigh spectra with different splitting ratios**

range reaches 34.013 5 dB and 28.157 0 dB, and the spatial resolution (using the 3 dB width of the end reflection peak as the benchmark) is 0.112 6 m and 0.117 4 m. These simulations demonstrate that the system achieves a sensing distance of 30 km and a spatial resolution of less than 0.3 m, thereby meeting the requirements for high-spatial-resolution and high-dynamic-range PON monitoring.

4 Conclusions

Simulation results demonstrate that the TGD-OFDR system overcomes key limitations of conventional OTDR, specifically insufficient incident optical pulse power and the inherent constraint linking spatial resolution to pulse width. While maintaining high spatial resolution, the system achieves a significantly enhanced dynamic range. Compared to conventional OTDR systems, which require hundreds of thousands of averages to achieve a dynamic range of 15 dB with a spatial resolution on the order of meters, the TGD-OFDR technique achieves superior spatial resolution (<0.3 m) and dynamic range (>25 dB) with significantly fewer averages (<500 times) and a much faster measurement rate. When applied to PONs, the system enables the monitoring of multi-stage optical splitters, supporting a larger number of ONUs over an extended reach while delivering a superior SNR. For a representative scenario involving two cascaded 1:8 splitters (imposing a minimum attenuation of ~ 20 dB), the TGD-OFDR-based PON monitoring system achieves a sensing distance of 30 km, a spatial resolution below 0.3 meters, and a dynamic range approaching 30 dB. This advancement holds significant potential for both fundamental research and the practical realization of intelligent digital PONs.

References

- [1] CHAMPAVÈRE A. New OTDR measurement and monitoring techniques [C]//Proc. Optical Fiber Communication Conference. OSA, 2014. DOI: 10.1364/ofc.2014.w3d.1
- [2] OZAWA K. Field trial of in-service individual line monitoring of PONs using a tunable OTDR [C]//Proc. Fourteenth International Conference on Optical Fiber Sensors. SPIE, 2000: 101 – 104. DOI: 10.1117/12.2302161
- [3] FATHALLAH H, RAD M M, RUSCH L A. PON monitoring: periodic encoders with low capital and operational cost [J]. IEEE photonics technology letters, 2008, 20(24): 2039 – 2041. DOI: 10.1109/LPT.2008.2006060
- [4] FONTAINE N K, MAZUR M, PUTTNAM B J, et al. Ultra-high resolution and long-range OFDRs for characterizing and monitoring Hollow-core DNANFs [C]//Proc. Optical Fiber Communication Conference (OFC). Optica Publishing Group, 2025: 101 – 104. DOI: 10.1364/ofc.2025.th4d.6
- [5] ZHOU X, ZHANG F D, SUN M M, et al. A modified optical coding monitoring scheme in PON with electronic decoding processing [J]. IEEE communications letters, 2013, 17(9): 1849 – 1851. DOI: 10.1109/LCOMM.2013.090213.131501
- [6] LIU Q W, FAN X Y, HE Z Y. Time-gated digital optical frequency domain reflectometry with 16-m spatial resolution over entire 110-km range [J]. Optics express, 2015, 23(20): 25988. DOI: 10.1364/oe.23.025988
- [7] XIAO Z Y, PAN Y C, CHEN J G, et al. High-performance and low-cost distributed acoustic sensor with phase noise compensation [C]//Proc. Asia Communications and Photonics Conference (ACP) and International Conference on Information Photonics and Optical Communications (IPOC). IEEE, 2024. DOI: 10.1109/ACP/IPOC63121.2024.10810028
- [8] LUO Y M, LIU Q W, LIU C Z, et al. A fading noise suppression method based on frequency-shift loop in TGD-OFDR system [C]//Proc. Asia Communications and Photonics Conference (ACP) and International Conference on Information Photonics and Optical Communications (IPOC). IEEE, 2024. DOI: 10.1109/ACP/IPOC63121.2024.10810004
- [9] DALLACHIESA L, MAZUR M, IANNONE P, et al. Coherent OFDR-based individual fiber identification and event detection over a 21-km passive optical distribution network with a 1:32 split [C]//2025 Optical Fiber Communications Conference and Exhibition (OFC). OPTICA, 2025: 1 – 3
- [10] CHEN D, LIU Q W, HE Z Y. Phase-detection distributed fiber-optic vibration sensor without fading-noise based on time-gated digital OFDR [J]. Optics express, 2017, 25(7): 8315 – 8325

Biographies

ZHU Yidai received his BS degree in optoelectronic science and engineering from the University of Electronic Science and Technology of China in 2020, and PhD degree in information and communication engineering from Shanghai Jiao Tong University, China in 2025. He joined the Wireline Product Planning Department, ZTE Corporation in 2025. His research interests include optical communications and optical sensing applications.

FAN Xinyu (fan.xinyu@sjtu.edu.cn) received his BS and MS degrees in applied physics from Shanghai Jiao Tong University, China in 2000 and 2003, respectively, and PhD degree in electrical engineering from the University of Tokyo, Japan in 2006. In the same year, he joined the NTT Access Network Service Systems Laboratories, Japan, where his research interests included optical reflectometry and optical measurement. Since 2012, he has been with the Department of Electrical Engineering, Shanghai Jiao Tong University, where he is currently a Professor. His research interests include optical fiber sensing technology, optical fiber measurement, and fiber-related applications.

ZHU Songlin received his MS degree in theoretical physics from Hangzhou University, China in 1998 and PhD degree in electronic science and technology from Zhejiang University, China in 2001. He joined the Wireline Product Planning Department, ZTE Corporation in 2001. His research interests include FT-Tx technology for optical communications and OTDR technology for optical sensing applications.

DONG Jiaxing received his BS degree in electronic information and electrical engineering from Shanghai Jiao Tong University, China in 2024. He is currently working toward the master's degree at the School of Integrated Circuits, Shanghai Jiao Tong University. His research focuses on optical fiber sensors.

LI Guoqiang received his BS degree in electronic information engineering from Wuhan University, China in 2019 and PhD degree in electromagnetic field and microwave technology from Fudan University, China in 2024. He joined the Wireline Product Planning Department, ZTE Corporation in 2024. His research interests include digital signal processing for optical communications and OTDR technology for optical sensing applications.

HE Zuyuan received his BS and MS degrees in electronic engineering from Shanghai Jiao Tong University, China in 1984 and 1987, respectively, and PhD degree in photonics from the University of Tokyo, Japan in 1999. He joined CI-ENA Corporation, USA in 2001, as a Lead Engineer, where he headed the optical testing and optical process development group. He returned to the University of Tokyo as a Lecturer in 2003, and then became an Associate Professor in 2005 and a Full Professor in 2010. He is now a Chair Professor and Head of the Department of Electronic Engineering at Shanghai Jiao Tong University. His current research interests include optical fiber sensors, specialty optical fibers, and optical interconnects.

Insights on Next Generation WLAN: High Experiences (HEX)



YANG Mao, LI Bo, YAN Zhongjiang

(Northwestern Polytechnical University, Xi'an 710072, China)

DOI: 10.12142/ZTECOM.202504003

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251107.1125.002.html>,
published online November 7, 2025

Manuscript received: 2025-09-11

Abstract: Wireless local area networks (WLANs) have witnessed rapid growth in the past 20 years, with maximum throughput as the key technical objective. However, quality of experience (QoE) remains the primary concern for wireless network users. We point out that poor QoE is the most challenging issue in current WLANs and further analyze the key technical problems that cause poor QoE in WLANs, including fully distributed networking architectures, chaotic random access, awkward “high capability” issues, coarse-grained quality of service (QoS) architectures, ubiquitous and complicated interference, “no place” for AI issues, and heavy burden of standard evolution. To the best of our knowledge, this is the first work to point out that poor QoE is the most challenging problem in current WLANs, and the first to systematically analyze the technical problems that cause poor QoE in WLANs. We strongly suggest that achieving high experience (HEX) be the key objective of the next-generation WLANs.

Keywords: wireless local area network; IEEE 802.11; quality of experiences; Wi-Fi; 802.11bn; ultra-high reliability

Citation (Format 1): YANG M, LI B, YAN Z J. Insights on next generation WLAN: high experiences (HEX) [J]. ZTE Communications, 2025, 23 (4): 10 – 15. DOI: 10.12142/ZTECOM.202504003

Citation (Format 2): M. Yang, B. Li, Z. J. Yan, “Insights on next generation WLAN: high experiences (HEX),” *ZTE Communications*, vol. 23, no. 4, pp. 10 – 15, Dec. 2025. doi: 10.12142/ZTECOM.202504003.

1 Introduction

Wireless networks are an indispensable and fast-growing technology in human lives. After more than 20 years of development, wireless local area networks (WLANs) and cellular networks have become dominant types of wireless networks. WLANs are standardized by IEEE 802.11, and IEEE 802.11be^[1-3] is the latest commercially available WLAN standard. IEEE 802.11be, also known as Extremely High Throughput (EHT) or Wi-Fi 7, was officially released in 2025^[4]. Now, both industry and academia are focusing on the key technology research and standardization of IEEE 802.11bn, which is the next generation of IEEE 802.11be. IEEE 802.11bn, also called Ultra-High Reliability (UHR) or Wi-Fi 8, is expected to be officially published in 2028 or 2029^[5-11]. In 2025, Draft 1.0 of IEEE 802.11bn was released. It introduces several notable features, including seamless roaming, non-primary channel access, coordination of multiple access points (APs), in-device coexistence, dynamic power save, etc. It can be seen that WLAN technology and its standardization process are developing very rapidly.

IEEE 802.11 has undergone eight major versions, including IEEE 802.11a, b, g, n, ac, ax, be, and bn. Among these, only

IEEE 802.11ax and bn take high efficiency as the technical objective, while other WLAN standards take maximum throughput as their key technical objective. Throughput is quite important, because higher throughput means WLANs can provide more capacity for the wireless traffic. However, with the increasing diversity of wireless services, such as virtual reality (VR), meta universe, ultra-high resolution online video, real-time games, remote medical services, and industry applications, it is increasingly difficult for WLANs to guarantee the quality of service (QoS) for these diverse services. More importantly, quality of experience (QoE) is the most important concern of wireless network users. QoE is not only related to the QoS, but also to other factors like human subjective perceptions. This makes QoE an important and urgent consideration for wireless networks.

Poor QoE is the most challenging problem in current WLANs. For instance, in many parts of China, as long as the traffic data fee for cellular networks is within budget, people would rather choose 4G/5G than free Wi-Fi. The problem lies not in peak performance, but in performance stability: Wi-Fi performance changes dramatically over time. In practice, the claimed “very/extremely high throughput” can hardly be experienced. Therefore, we believe high QoE should be the key objective of the next generation WLAN standard. The problem

becomes serious in the fiber-to-the-room (FTTR) scenario, since potential interference sources are complex, e.g., neighboring networks, multiple intra-home cells and different kinds of devices. Table 1 shows some comparisons between WLANs and cellular networks. In this paper, we analyze the key technical problems that lead to poor QoE of WLANs. This is the first work to identify poor QoE as the most challenging problem in current WLANs, and the first to systematically analyze the technical factors behind it.

2 Overview of Key Technical Problems

Based on our analysis, there are seven key technical problems that lead to poor QoE of WLANs. They are summarized as follows:

- **Network architecture:** Different from cellular networks, WLANs adopt a fully distributed network architecture. This architecture has its advantages, but it usually leads to a disordered network status, resulting in low QoE.
- **Channel access:** Stations (STAs) and APs randomly contend for the channel, resulting in chaotic channel access among STAs.
- **Transmission:** IEEE 802.11be includes new features,

such as multiple link operation (MLO), larger bandwidth (e.g., 320 MHz), higher-order modulation (e.g., 4096-QAM), 1024-frame aggregation, more spatial streams (16 SS), etc. These potential features sound attractive. However, the actual available performance is far from its stated high capability.

- **QoS guarantee:** The current QoS architecture is traffic class based. Eight traffic identifiers (TIDs) and four access categories (ACs) cannot keep pace with the increasingly diverse services.
- **Interference management:** Unlike that in cellular networks, random interference from the overlapping basic service set (OBSS), intra-BSS, and non-Wi-Fi systems is ubiquitous and complex.
- **Network intelligence:** Making networks more intelligent is a natural and promising target, but there appears to be no place for AI in the standardization process.
- **Legacy dilemma:** Essentially, with each evolution of the IEEE 802.11 standard, the newly revised version has no choice but to coexist with many versions of legacy STAs.

As shown in Fig. 1, network architecture problem is the framework-related challenge derived from topology, deployment and composition. Channel access and transmission prob-

Table 1. Comparisons between WLANs and cellular networks

Feature	WLAN	Cellular Network
Network deployment	Advantages: easy, open, and flexible to deploy Disadvantages: interference is difficult to control	Advantages: interference is controllable Disadvantages: in need of careful design and test; inflexible that only allows the mobile operators to deploy
Channel access and transmission	Advantages: easy to access the channel for each device with less control and management signaling Disadvantages: strong interference and collisions and low resource utilization ratio	Advantages: collision-free; little interference and high resource utilization ratio Disadvantages: all uplink and downlink channel access and transmission need to be scheduled by the base station with heavy control and management signaling
QoS and QoE	Advantages: in light traffic load scenarios, the access latency is low because of random access strategy Disadvantages: in medium or heavy traffic load scenarios, the QoS and QoE are quite poor and uncertain	Advantages: high certainty Disadvantages: high access latency because of heavy signaling in light traffic load scenario

QoS: quality of service QoE: quality of experience WLAN: wireless local area network

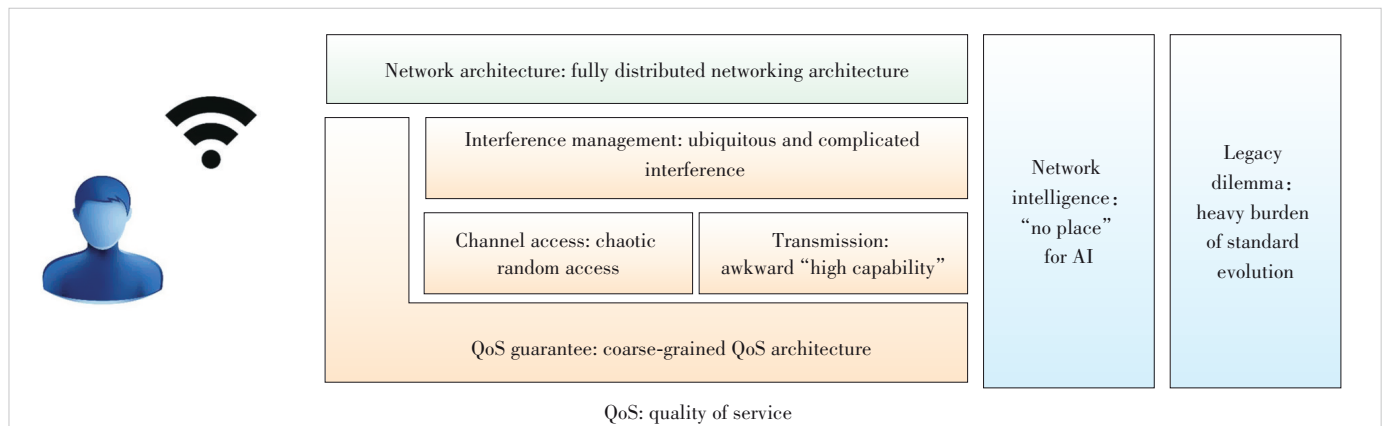


Figure 1. Architecture of the key technical problems that cause poor QoE of WLANs

lems are summarized from the perspective of a single BSS or any BSS, while the interference management problem is analyzed from the perspective of multiple BSSs or the OBSS. Furthermore, the QoS guarantee problem is obtained by analyzing the problems existing in the QoS guarantee methods at all levels, including channel access, transmission, and interference management. Moreover, network intelligence and legacy dilemma problems are two relatively independent aspects that have a significant impact on WLAN performance and standardization. In the following section, we analyze these technical problems in detail.

3 Analysis of Factors Influencing Key Technical Problems

3.1 Fully Distributed Networking Architecture

Different from cellular networks, WLANs adopt a fully distributed network architecture, in which there is no central controller to regulate or manage the behaviors of multiple BSSs. Even in a single BSS, the control and management functions of the AP are limited. Typically, APs and STAs independently make their own decisions without coordination. A distributed architecture has its advantages, such as flexibility and easy deployment, but it usually leads to a network disorder. Consequently, the fully distributed architecture makes the network inefficient, resulting in low QoE.

For example, within a single BSS, the network is distributed even in the presence of an AP. Both the AP and STAs randomly contend for channel access. The STAs can determine their own configurations and parameters for channel access, transmission, etc. Furthermore, coordination within an extended service set is still very limited. We highlight that network architecture is of greater significance for FTTR scenarios, as FTTR involves more flexible yet complex in-home network topologies. Even in deployed enterprise networks, distributed features also result in many problems, such as the starvation problem (flow-in-the-middle), which makes it quite difficult for some BSSs or STAs to access the channel.

The influencing factors of the fully distributed networking architecture are summarized as follows.

- It lacks overall and top-level architecture design. Since the inception of IEEE 802.11, WLANs have always followed the distributed networking architecture to facilitate deployment. With the standard versions evolving from generation to generation, this distributed architecture is more deeply rooted and difficult to change.
- The distributed architecture is the basis of network management, control, and data transmission. This means that all aspects of network functions are designed and developed based on it, making the whole system affected by this architecture.
- The current industrial, scientific, and medical (ISM) bands are limited and the environment is complex. WLANs operate on the ISM bands, especially the 2.4 GHz, 5 GHz, and

6 GHz bands. The ISM band's frequency bandwidth is limited, which leads to insufficient resources for WLAN architecture changes. Moreover, many wireless network types work on the ISM bands, such as Wi-Fi, Bluetooth, Zigbee, and microwave ovens. Thus, the channel environment is quite complicated, which increases the technical and policy difficulties for WLANs to change the distributed architecture.

3.2 Chaotic Random Access

In the media access control (MAC) layer, STAs and APs randomly contend for the channel according to the enhanced distribution channel access (EDCA) scheme based on the carrier sense multiple access with collision avoidance (CSMA/CA) mechanism, resulting in chaotic channel access among STAs. Chaotic random access significantly exacerbates collisions, ineffective access, and transmission issues, which lead to low QoE.

For example, STAs randomly access channels by consistently using the maximum TX power, setting a transmission opportunity (TXOP) limit exceeding the actual demand. For home and enterprise scenarios including FTTR, multiple BSSs or multiple STAs within a single BSS often act as "enemies" to one another. Moreover, IEEE 802.11be introduces restricted target wakeup time (r-TWT) to improve latency performance, but the rule "non-AP EHT STAs may behave as if overlapping quiet intervals do not exist" exacerbates random collisions.

The factors influencing chaotic random access are summarized as follows.

- Deep-rooted random channel access: In the first version of IEEE 802.11, CSMA/CA mechanism was adopted. Although the channel access mechanism has been improved during the WLAN standard evolution, such as the transition from distributed channel function (DCF) to EDCA, the CSMA/CA mechanism based on the idea of random access is still deep-rooted.
- Rigid clear channel assessment (CCA) and energy detection (ED) thresholds incompatible with diverse scenarios: In order to support CSMA/CA, IEEE 802.11 adopts the ED threshold and the CCA threshold to determine channel status. However, the rigid CCA and ED thresholds do not dynamically change as the environmental status varies.
- Lack of a well-designed channel access structure: Well-designed channel access mechanisms, such as scheduled channel access and reservation-based channel access, have several performance advantages. These mechanisms can be important supplements to random channel access and further form a more complete channel access structure.
- Inappropriate channel access rules for new bands: Originally, a clean 6 GHz band was available for Wi-Fi, but we adopted the CSMA/CA mechanism directly and contaminated this frequency band. If there are new bands for WLANs in the future, we can fully seize the opportunity to carry out new designs.

3.3 Awkward “High Capability”

Several new features for IEEE 802.11be have been mentioned in Section 2. While these potential features appear promising in theory, their actual performance falls far short of the claimed capabilities. A series of valuable technologies cannot be implemented in practice, and thus, high-quality user experiences are also unattainable.

For example, in high-density deployment scenarios, complex interference among multiple BSSs and STAs makes it nearly impossible to achieve 320 MHz bandwidth; instead, 20 MHz or 40 MHz is normal. Furthermore, 4 096-QAM or even 1 024-QAM is difficult to use because of the demanding signal-to-interference and noise ratio (SINR) threshold. Notably, 5G only uses 256-QAM as the highest modulation order. Increasing the number of spatial streams is also challenging because of inter-stream interference and complex user grouping.

The influencing factors behind the awkward “high capability” are summarized as follows.

- **Complicated channel environment and interference:** The WLAN operates at the ISM band, resulting in more cross-system interference. The fully distributed architecture and the chaos random access significantly increase the inter-BSS and inter-STA interference. In this case, the channel environment and interference become quite complicated, making the conditions of the potential high performance enabling technologies hard to meet.
- **Lack of mechanisms to realize these valuable technologies in practice.** There are few mechanisms to guarantee that the conditions required by the potential high-performance enabling technologies can be met. This makes these technologies remain high-performance only in theory rather than in real-world applications.
- **Declining efficiency:** The higher the peak throughput, the lower the efficiency. The primary reason is that higher peak throughput leads to relatively larger protocol overhead. The transmission duration of the control frames, channel access time, and the inter-frame space do not decrease with the data transmission rate. Moreover, some rules restrict each other; for example, the maximum physical layer protocol data unit (PPDU) length restricts the aggregation size.

3.4 Coarse-Grained QoS Architecture

The current QoS architecture is traffic class-based. Traffic received from upper layers is mapped to eight TIDs and four ACs. But traffic (or service) types are getting increasingly diverse. Thus, the current class-based QoS mechanism cannot keep pace with the increasingly diverse services. QoS is the premise of QoE, and poor QoS will inevitably lead to poor QoE.

For example, increasingly diverse services (such as VR, real-time cloud games, the metaverse, digital twins, and remote healthcare) require highly diverse and challenging QoS. Furthermore, it seems we have to achieve the “unattainable” low-latency objective. Although low latency is a feature of

IEEE 802.11be, there are few effective solutions in IEEE 802.11be except r-TWT. Several standard proposals discuss a latency guarantee of less than 1 ms, which seems “unsolvable” for Wi-Fi. Moreover, complicated environments lead to unstable resource acquisition, further resulting in inconsistent performance.

Some new scenarios like FTTR and Internet of Things (IoT) are considered important application scenarios. Various types of applications coexist in FTTR scenarios, such as videos, screen casting, and VR. The QoS and QoE requirements of these applications are quite different. However, WLANs only pursue increasingly larger bandwidth and other enablers, making them “unfriendly” to FTTR and IoT.

Mobility support is increasingly important for WLANs. However, seamless Wi-Fi roaming is challenging for the current standards.

The influencing factors of the coarse-grained QoS architecture are summarized as follows.

- **Need for QoS architecture with appropriate granularity:** The current class-based QoS mechanism (especially its eight TIDs and four ACs) cannot meet the requirements of the increasingly diverse services. Thus, other QoS mechanisms (e.g., packet-level or finer granularity) should be fully studied. It is worth noting that we propose a particle access method and theory for wireless networks by treating each packet with several attributes as an “information particle”^[12]. After optimally combining some information particles into an information particle group, a low-complexity scheduling strategy is used to deploy packet-level resource allocation and provide the packet-level QoS guarantee, ensuring the latency and the throughput requirements of the “particle” are met.
- **Fine scalability:** The QoS mechanism should keep pace with the continuous evolution of wireless services. Thus, better scalability is quite important.
- **Lack of truly effective solutions for low latency/jitter:** WLANs face great challenges in ensuring low latency. If low latency/jitter cannot be guaranteed, the future of WLANs will be bleak.
- **Stable resource acquisition:** Only with stable access to resources can QoS and QoE be improved. However, there is a lack of mechanisms for resource acquisition.
- **Integrated wide-band systems and narrow-band IoT:** IoT usually requires a small bandwidth (narrow band) in many scenarios because of energy consumption and cost. But large bandwidth has always been the technical goal of WLANs. For the future, if wide band and narrow band systems cannot be integrated in the WLAN standard, the WLAN will probably not keep up with the development tide of the IoT.
- **Seamless Wi-Fi roaming:** As the operating frequency becomes higher, the BSS becomes smaller and denser. Thus the mobility of nodes has to be considered. If QoS cannot be guaranteed during the node movement, QoE will be seriously affected. The challenge lies in the fact that, unlike cellular net-

works, WLAN did not take supporting mobility as an important goal at the beginning.

3.5 Ubiquitous and Complicated Interference

Quite different from cellular networks, random interferences from OBSS, intra-BSS, and non-Wi-Fi systems are ubiquitous and complicated. Interference leads to poor network performance, which in turn results in poor QoE.

For example, high-density deployment scenarios will be typical in future wireless networks. WLAN devices made by different vendors adopt different implementations to select channels, bandwidth, access parameters, etc., which makes interference among BSSs extremely complicated. IEEE 802.11ax introduces spatial reuse (SR), which further increases interference among BSSs. Moreover, cross-system interferences may come from various non-Wi-Fi systems such as licensed assisted access (LAA), LTE-u, Sparklink (Nearlink), etc.

The influencing factors of the ubiquitous and complicated interference are summarized as follows.

- Lack of inter-BSS coordination: As deployment scenarios become increasingly dense, ubiquitous and complicated interference emerges. However, the inter-BSS coordination is very limited in current WLANs.
- Complicated ISM band interference. WLANs operate on the ISM band. Many heterogeneous systems, such as Bluetooth, LAA, LTE-u, Zigbee, work on the same band. More importantly, devices cannot understand frames sent from different systems, making cross-system interference ubiquitous.
- Lack of explicit technical rules for WLAN coexistence with other systems. For the coexistence of the heterogeneous systems, there is a lack of explicit technical rules, such as channel access rules among different systems, which aggravates interference complexity.
- Shortage of new bands for WLANs. New bands can mitigate interference, but spectrum resources are limited.

3.6 “No Place” for AI

User experience is subjective and complex. Making networks more intelligent is a natural and promising target for WLANs, but it seems there is “no place” for AI in standardization. Network intelligence can help us evaluate complex user experience and choose the optimal method to guarantee QoE. Many studies on supervised learning and reinforcement learning without deep learning have been proposed in recent years, but they lack support from standards, making AI solutions difficult to extend.

For example, several standard proposals discuss the importance of AI for WLANs. However, machine learning is usually considered an internal tool for each individual module in a single device. Such a kind of AI obtains limited performance gains and poor scalability. Moreover, there are lots of AI models, training methods, and algorithms. Simply standardizing a specific one is not scalable.

To better embrace AI, we need to answer the following questions: What can AI do for WLANs? What can standardization do for AI? What is the AI standardization architecture in IEEE 802.11? At least, network AI does not simply mean implementing specific AI models or algorithms.

3.7 Heavy Burden of Standard Evolution

Different from cellular networks, every time the IEEE 802.11 standard evolves, it has no choice but to coexist with many versions of legacy STAs, which is a heavy burden for standard evolution. Dropping the heavy burden of legacy is beneficial to the technical innovation, as it allows us to focus on and further guarantee the user experience enhancement of the new-generation standard.

For cellular networks, standard evolution is unfettered, featuring new bands, new designs, new frame formats, etc. However, for WLANs, standard evolution has to keep good backward compatibility, which is a “double-edged sword”. For example, a clean 6 GHz band was available for Wi-Fi, yet we simply incorporated all the old technologies. Moreover, the frame format is continuously patched. As the standard evolves, these patches have become fragmented.

The influencing factors of the heavy burden of standard evolution are summarized as follows.

- Old and new versions share the same bands. This makes it difficult for new-generation standards to break free from coexistence constraints and carry out new designs.
- Devices, especially APs installed with new standards, have to serve all old versions. This makes it impossible for the new standards to break away from compatibility limitations and carry out new designs.
- Frame formats are full of patches, but little emphasis is placed on scalable designs or clean-sheet designs.

4 Conclusions

In this article, we point out that poor QoE is the trickiest problem affecting the evolution and user growth of WLANs. More importantly, we analyze in detail seven technical problems that cause poor QoE of WLANs. To the best of our knowledge, this is the first work to point out that poor QoE is the most challenging problem for current WLANs, and also the first to systematically analyze the technical problems that lead to poor QoE of WLANs. Table 2 summarizes the key technical problems and further analyzes the threat types and the importance of each problem.

The vision of wireless networks is, in our opinion, to enable people to enjoy wireless connectivity and to enhance ubiquitous interconnection. We are glad that WLANs have achieved and continue to achieve great success. It is time for us to carefully think about things from the users’ perspective. Therefore, to achieve high quality experiences is highly suggested as the key objective of the next-generation WLAN: Wi-Fi 9.

Table 2. Summary of the key technical problems that lead to poor quality of experience of WLANs

Feature	Problem	Threat	Importance
Network architecture	Fully-distributed networking architecture	Disorder	★★★★★
Channel access	Chaotic random access	Disorder	★★★★★
Transmission	Awkward “high capability”	Low efficiency	★★★★
QoS guarantee	Coarse-grained QoS architecture	Weak adaptability	★★★★
Interference management	Ubiquitous and complicated interference	Disorder	★★★★
Network intelligence	“No place” for AI	Weak adaptability	★★★
Legacy dilemma	Heavy burden of standard evolution	Hindering evolution	★★★

QoS: quality of service

References

- [1] YANG M, LI B. Survey and perspective on extremely high throughput (EHT) WLAN: IEEE 802.11be [J]. Mobile networks and applications, 2020, 25(5): 1765 – 1780. DOI: 10.1007/s11036-020-01567-7
- [2] Cisco. Visual networking index: white paper [R]. San Jose, USA: Cisco, 2020
- [3] DENG C L, FANG X M, HAN X, et al. IEEE 802.11be Wi-Fi 7: new challenges and opportunities [J]. IEEE communications surveys & tutorials, 2020, 22(4): 2136 – 2166. DOI: 10.1109/COMST.2020.3012715
- [4] LAN/MAN Standards Committee of IEEE Computer Society. IEEE standard for information technology-telecommunications and information exchange between systems local and metropolitan area networks specific requirements. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 2: enhancements for extremely high throughput (EHT): IEEE Std 802.11be-2024 [S]. 2025
- [5] GALATI-GIORDANO L, GERACI G, CARRASCOSA M, et al. What will Wi-Fi 8 be? A primer on IEEE 802.11bn ultra high reliability [J]. IEEE communications magazine, 2024, 62(8): 126 – 132. DOI: 10.1109/MCOM.001.2300728
- [6] WEI D Y, CAO L, ZHANG L, et al. Optimized non-primary channel access design in IEEE 802.11bn [C]//IEEE Global Communications Conference. IEEE, 2024: 4588 – 4593. DOI: 10.1109/GLOBECOM52923.2024.10901367
- [7] WILHELM F, GALATI-GIORDANO L, GERACI G, et al. Throughput analysis of IEEE 802.11bn coordinated spatial reuse [C]//IEEE Conference on Standards for Communications and Networking (CSCN). IEEE, 2023: 401 – 407. DOI: 10.1109/CSCN60443.2023.10453190
- [8] MECKLENBURG B, ANSARI A H, RICHERZHAGEN B, et al. Seamless roaming based on distributed multi-link operation over IEEE 802.11bn [C]// International Conference on Factory Communication Systems (WFCS). IEEE, 2025: 1 – 8. DOI: 10.1109/WFCS63373.2025.11077616
- [9] NUNEZ D, WILHELM F, GALATI-GIORDANO L, et al. Spatial reuse in IEEE 802.11bn coordinated multi-AP WLANs: a throughput analysis [C]// IEEE Conference on Standards for Communications and Networking (CSCN). IEEE, 2024: 265 – 270. DOI: 10.1109/CSCN63874.2024.10849731
- [10] YAN R, GUO Z Y, LIU P, et al. Multi-agent reinforcement learning-based channel access optimization for IEEE 802.11bn [J]. IEEE transactions on green communications and networking, 2025, 9(3): 1429 – 1441. DOI: 10.1109/TGCN.2024.3495236
- [11] WOJNAR M, CIEZOBKA W, KOSEK-SZOTT K, et al. IEEE 802.11bn multi-AP coordinated spatial reuse with hierarchical multi-armed bandits [J]. IEEE communications letters, 2025, 29(3): 428 – 432. DOI: 10.1109/LCOMM.2024.3521079
- [12] LI B, SUN K, YAN Z J, et al. Idea and theory of particle access [EB/OL]. (2023-05-15) [2025-09-16]. <https://arxiv.org/abs/2203.15191>

Biographies

YANG Mao (yangmao@nwpu.edu.cn) received his BE and MS degrees in information and telecommunication engineering from Xidian University, China in 2006 and 2009, respectively, and his PhD degree in electronic engineering from Tsinghua University, China in 2014. He is currently an associate professor in the School of Electronics and Information, Northwestern Polytechnical University, China. His research interests focus on wireless networking and communications, including next-generation cellular networks (6G) and WLAN (e.g. IEEE 802.11bn), the MAC and higher-layer technologies, non-orthogonal multiple access for 6G, software-defined wireless networking, and wireless network virtualization. His research projects are funded by the National Natural Science Foundation of China. He has published over 100 research papers and has over 100 granted and pending international and Chinese patents. He is a member of the IEEE and a voting member of IEEE 802.11 Working Group.

LI Bo received his BE, MS, and PhD degrees in telecommunication engineering from Xidian University in 1994, 1997, and 2002, respectively. From 1997 to 1998, he was a visiting scholar in the Department of Engineering, Shizuoka University, Japan. From 2002 to 2004, he was a postdoctoral fellow at the University of Trento, Italy. From July to December 2007, he was a visiting professor at the Institut National des Sciences Appliquées (INSA), France. He is currently a professor at the School of Electronics and Information, Northwestern Polytechnical University, China. His research interests are in the area of wireless networking and communications, including next-generation cellular networks (5G) and WLAN (e.g. IEEE 802.11ax and 11ay), MAC and higher-layer technologies, and non-orthogonal multiple access for 5G. He has published over 80 research papers and has over 40 granted and pending international and Chinese patents. He is a member of the IEEE.

YAN Zhongjiang received his BE and PhD degrees in telecommunication engineering from Xidian University, China in 2006 and 2011, respectively. From September 2010 to December 2011, he was a visiting PhD student in the Department of Electrical and Computer Engineering, University of Alberta, Canada. He is currently an associate professor with the School of Electronics and Information, Northwestern Polytechnical University, China. His research interests focus on wireless networking and communications, including protocol and algorithm design, modeling, performance analysis, and implementation of the media access control layer, radio resource management, and traffic scheduling strategies for wireless networks. He is a member of the IEEE and a voting member of IEEE 802.11 Working Group.

FTTR-MmWave Architecture for Next-Generation Indoor High-Speed Communications



CHEN Zhe¹, ZHOU Peigen¹, WANG Long²,
HOU Debin², HU Yun³, CHEN Jixin¹, HONG Wei¹

(1. Southeast University, Nanjing 210096, China;
2. MiSic Microelectronics Co., Ltd., Nanjing 211111, China;
3. Nanjing Ziwei Technology Co., Ltd., Nanjing 211111, China)

DOI: 10.12142/ZTECOM.202504004

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251216.1124.002.html>,
published online December 16, 2025

Manuscript received: 2025-09-15

Abstract: Millimeter-wave (mmWave) technology has been extensively studied for indoor short-range communications. In such fixed network applications, the emerging FTTR architecture allows mmWave technology to be well cascaded with in-room optical network terminals, supporting high-speed communication at rates over tens of Gbit/s. In this Fiber-to-the-Room (FTTR)-mmWave system, the severe signal attenuation over distance and high penetration loss through room walls are no longer bottlenecks for practical mmWave deployment. Instead, these properties create high spatial isolation, which prevents mutual interference between data streams and ensures information security. This paper surveys the promising integration of FTTR and mmWave access for next-generation indoor high-speed communications, with a particular focus on the Ultra-Converged Access Network (U-CAN) architecture. It is structured in two main parts: it first traces this new FTTR-mmWave architecture from the perspective of Wi-Fi and mmWave communication evolution, and then focuses specifically on the development of key mmWave chipsets for FTTR-mmWave Wi-Fi applications. This work aims to provide a comprehensive reference for researchers working toward immersive, untethered indoor wireless experiences for users.

Keywords: Fiber-to-the-Room; millimeter wave; Wi-Fi; cloud virtual reality (cloud VR); beamforming

Citation (Format 1): CHEN Z, ZHOU P G, WANG L, et al. FTTR-mmWave architecture for next-generation indoor high-speed communications [J]. *ZTE Communications*, 2025, 23(4): 16 – 26. DOI: 10.12142/ZTECOM.202504004

Citation (Format 2): Z. Chen, P. G. Zhou, L. Wang, et al., “FTTR-mmWave architecture for next-generation indoor high-speed communications,” *ZTE Communications*, vol. 23, no. 4, pp. 16 – 26, Dec. 2025. doi: 10.12142/ZTECOM.202504004.

1 Introduction

Over the past two decades, fixed broadband access technologies have continuously evolved to provide high-speed internet connectivity to end-users worldwide. The rise of new audio-visual media services such as augmented reality (AR), virtual reality (VR), 8K high-definition video, and cloud gaming has created higher demands for network throughput, coverage, and latency from content producers to end-users. The quality of fixed wireless broadband access significantly impacts the end-user experience, affecting factors like maximum service throughput, network stability, ease of installation, and easy management capabilities^[1–5].

Currently, mainstream commercial wireless local area networks (WLANs) and cellular networks primarily operate in the sub-6 GHz frequency band, which has limited data transmission bandwidth and struggles to meet the demands

of these new services. For example, for highly interactive, extreme cloud VR video transmission, a home network must guarantee a stable rate of at least 4.4 Gbit/s, with round-trip latency and jitter not exceeding 10 ms. High-quality VR headsets are typically connected via High-Definition Multimedia Interface (HDMI) cables, which can transmit up to 40 Gbit/s and require next-generation wireless ultra-high-speed interconnection technologies to provide users with an immersive, untethered experience^[6]. The millimeter-wave (mmWave) band offers vast, continuous spectrum, which can provide greater transmission bandwidth. Furthermore, due to the short wavelength of mmWave, the transmitting antennas can be very small, which is conducive to making devices lightweight. To achieve higher wireless communication speeds, researchers across academia and industry have conducted extensive research on mmWave channel characterization, system architectures, software algorithms, hardware circuits, and multiple standards^[7–22] over the past two

decades. Based on home network scenarios and use cases, the challenges of 60 GHz mmWave transmission have been investigated^[17]. Efforts have also been dedicated to channel measurements and modeling at 45 GHz and 60 GHz^[19, 21 - 22]. Similarly, the use of beamforming in short-range communication has been widely studied^[14, 18, 20].

The emergence of Fifth-Generation Fixed Networks (F5G) has seen the proposal of Fiber-to-the-Room (FTTR) technology. FTTR extends fiber optic cables to each room, providing gigabit broadband coverage for entire indoor spaces^[23 - 25]. By leveraging the fiber infrastructure laid into each room, FTTR creates a large-bandwidth connection channel for high-speed mmWave communications, enabling efficient mmWave coverage in every room. At the same time, the high wall-penetration loss of the mmWave band creates a favorable condition for eliminating interference between wireless communication systems in different rooms. This allows FTTR to achieve same-frequency networking and ensure information security. To address the demands of new scenarios and services, China Unicom's Ultra-Converged Access Network (U-CAN) hyper-converged access network architecture integrates advanced technologies like 50 Gigabit-capable Passive Optical Network (50G-PON), FTTR, and next-generation Wi-Fi 8 to support the continuous development of indoor tens-gigabit services^[26]. New services, such as video production, cloud PCs, and XR computing buses, have stringent requirements including deterministic latency, mobile roaming, and multi-domain networking. Consequently, the architecture proposes a new frequency division duplex access protocol for the next-generation Wi-Fi 8 mmWave air interface, efficient beam management and scanning mechanisms, and a high-low-frequency collaboration mechanism to address mmWave obstruction issues.

Furthermore, a multi-band hybrid networking solution is used to solve challenges like continuous indoor mmWave coverage. The Centralized/Cloud Wireless-Optical Access Network (C-WAN) is a centralized Wi-Fi architecture. It aims to provide full-house network coverage and efficient resource management through the synergy of optical networks and Wi-Fi. Its core features include 1) centralized control, in which a master device collects information and makes decisions, enabling unified and coordinated configuration of optical links and air interface links; 2) seamless roaming, which ensures fast handover of terminals between different access points, thereby enhancing the user experience; 3) interference optimization through Wi-Fi power balancing and interference mitigation. By innovating both the FTTR C-WAN architecture and air interface technology, the system effectively addresses issues of lagging and roaming in home and enterprise Wi-Fi applications, further improving the experience of short-range indoor wireless networks and enhancing efficiency^[27 - 29].

This paper reviews the U-CAN architecture enabled by

FTTR-mmWave Wi-Fi technology, with a primary focus on recent advances in Chinese mmWave chips and modules for next-generation FTTR-mmWave Wi-Fi applications. It is organized as follows. Section 2 briefly reviews the evolution of Wi-Fi and mmWave standards and the FTTR-mmWave system architecture. Section 3 focuses on the design of typical millimeter-wave chips, as well as recent Chinese commercial mmWave chip and module developments. Finally, Section 4 concludes the paper.

2 Wi-Fi and MmWave Standard Evolution and FTTR-MmWave Solution

2.1 Wi-Fi and MmWave Standard Evolution

In 1990, the IEEE 802.11 committee was established to promote research and development of standards for WLANs. In 1997, the first version, 802.11-1997, was officially released^[30]. The Wi-Fi Alliance was founded in 1999 to apply the 802.11 standard to the industry and facilitate compatibility among different manufacturers. With the growth of WLAN technology and the proliferation of mobile wireless devices, it has become a crucial technology for household network coverage. The release of the 802.11n protocol in 2009 was a major milestone, significantly increasing Wi-Fi network speeds with features like multiple antennas and beamforming that are still used today. Currently, 802.11ax technologies and network devices are widely used, and Wi-Fi 7 (802.11be) is poised to offer a better wireless network experience.

Millimeter-wave refers to electromagnetic waves with a wavelength between 1 mm and 10 mm, corresponding to a frequency range of 30 GHz to 300 GHz. The mmWave band offers vast spectrum resources. According to Shannon's theorem, the communication capacity of the system is proportional to the channel bandwidth, giving mmWave systems a natural advantage in channel capacity. By 2018, several mmWave-related standards had been introduced globally. The ECMA-387 standard primarily defines the characteristics of the physical layer (PHY), protocol adapter layer (PAL) for HD media interfaces, and media access control (MAC) layer in the 60 GHz band, mainly providing high-speed audio and HD video transmission. The IEEE 802.15.3c standard, operating in the 57 - 66 GHz band, primarily specifies the PHY and MAC layers for Wireless Personal Area Networks (WPANs). WPANs have an effective transmission distance of 10 m and are mainly used for high-speed network downloading. The 802.11ad physical layer supports various modes, including Single Carrier (SC) and Orthogonal Frequency Division Multiplexing (OFDM). As an evolution of 802.11ad, IEEE 802.11ay enhances theoretical speeds (theoretically up to 20 Gbit/s) and expands application scenarios. The IEEE 802.11aj standard is intended for indoor wireless communications, with frequency ranges of 42.3 - 48.4 GHz and 59 - 64 GHz. The MAC layer is designed similarly to

other IEEE 802.11 protocol standards, while the physical layer introduces analog beamforming and antenna array technologies to improve mmWave coverage. The next-generation Wi-Fi 8 (802.11bq) standard working group has included both the low microwave band (sub-7.25 GHz) and the mmWave Q/V bands (42 – 71 GHz) as candidate working frequency bands.

2.2 FTTR-MmWave System Architecture

The combination of FTTR and mmWave WLAN technology is a recent research topic that is expected to become an optimal solution for service and data transmission in F5G. One such architecture is the U-CAN, which defines a unified access architecture from the cloud to the terminal side and flexibly supports the allocation of transmission resources for industry applications. The U-CAN integrates 50G-PON, FTTR, C-WAN, Wi-Fi 8 (mmWave band), and other technologies to achieve a comprehensive upgrade of the next-generation hyper-converged access network. In this new architecture, fiber optic cables are laid into each room to provide a high-capacity mmWave backhaul channel. With fiber penetrating into each room, using mmWave has two distinct advantages: extremely high speed enabled by more available spectrum and less interference between adjacent networks due to high wall-penetration loss. In this setup, each FTTR terminal can be cascaded with an mmWave access point (AP) on the downlink while connecting to the same Optical Network Terminal (ONT) on the uplink, as shown in Fig. 1.

Compared to traditional Wi-Fi, the FTTR-mmWave system can eliminate interference and conflicts between different rooms, effectively improving data rates and transmission reliability while reducing latency. This performance is superior to that of traditional Wi-Fi Mesh network architectures. Currently, a Chinese company has designed an FTTR mmWave prototype using 4×4 multi-input multi-output (MIMO), which has achieved communication rates higher than 10 Gbit/s, validating the performance of the FTTR mmWave system for indoor communications.

3 MmWave Chips for FTTR-MmWave Wi-Fi Applications

FTTR technology is now in large-scale commercial use to address the high-rate transmission requirements of enterprises and homes. However, the bottleneck for even higher transmission rates lies in the fiber-to-wireless transition link. Given its high bandwidth and low latency, the mmWave band is an excellent solution to high-speed wireless transmission. This section focuses on the development of mmWave chips operating in 26 GHz, 40 GHz, and 45 GHz for next-generation indoor high-speed communication applications. While ensuring low cost and high integration, these chips are designed to optimize key metrics such as system bandwidth, power, noise, and linearity to meet the requirements of 16-QAM to 64-QAM modulation schemes to achieve higher transmission rates.

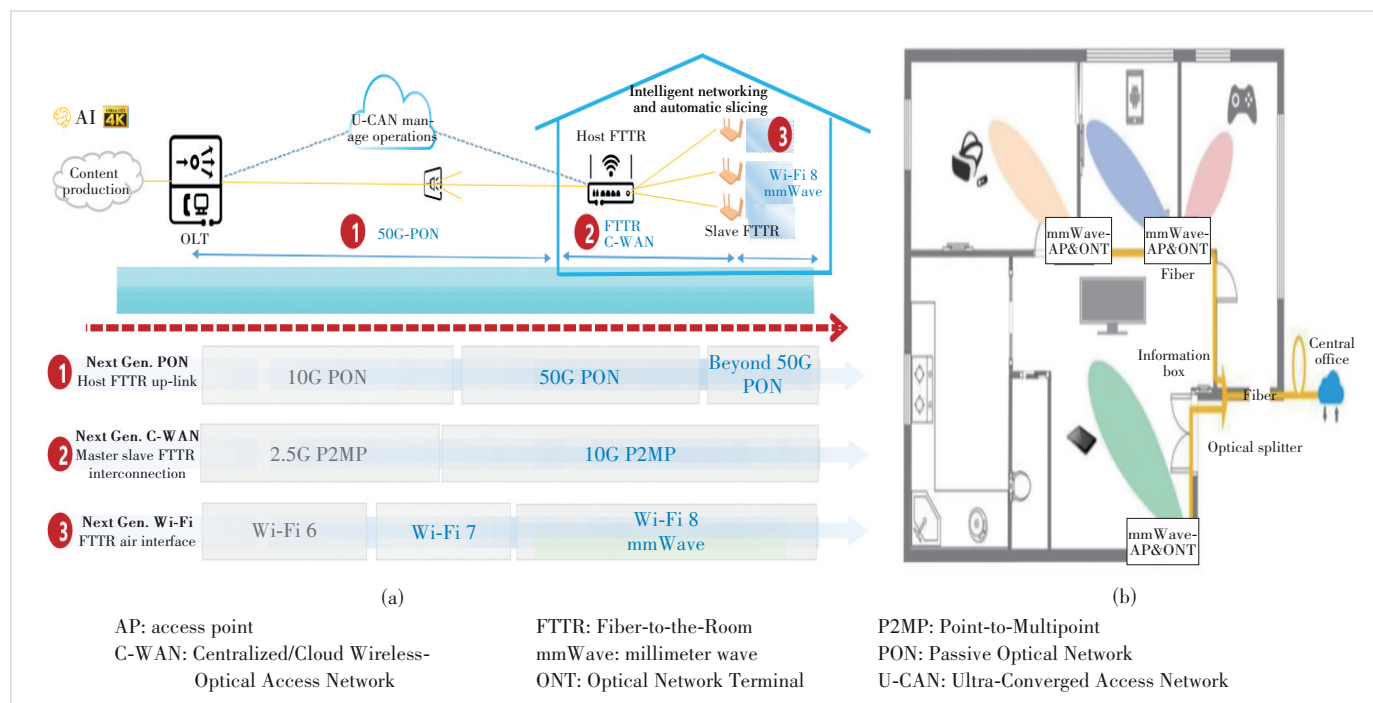
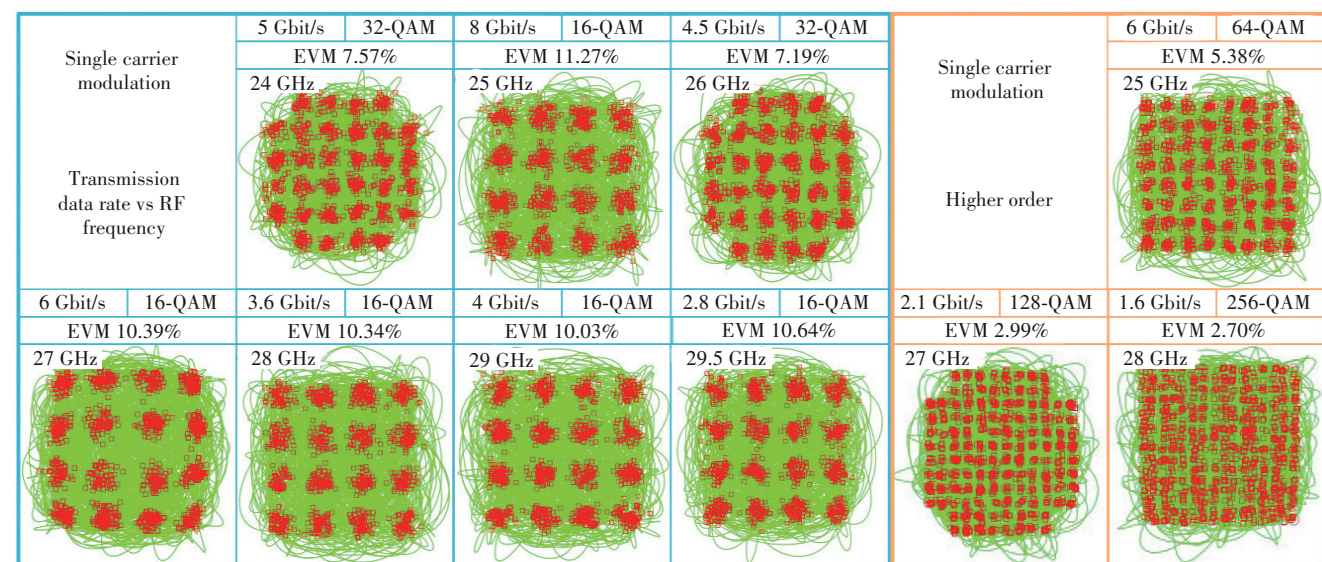


Figure 1. Schematic diagram of FTTR-mmWave cascading in the U-CAN architecture: (a) overall U-CAN architecture and (b) FTTR-mmWave system



(d)			
Att. : attenuator	I-Q: In-phase and Quadrature	NR: New Radio	SCPI: Standard Commands for
BPF: Band-Pass Filter	IF: intermediate frequency	PA: Power Amplifier	Programmable Instruments
DA: Driver Amplifier	LNA: Low-Noise Amplifier	QAM: Quadrature Amplitude	SPI: Serial Peripheral Interface
EVM: error vector magnitude	LO: local oscillator	Modulation	
FSPL: Free Space Path Loss	MCS: Modulation and Coding Scheme	RF: radio frequency	

Figure 3. Chip micrograms and measurement results: (a) micrograph of the 24 – 29.5 GHz Rx chip; (b) micrograph of the 24 – 29.5 GHz Tx chip; (c) over-the-air (OTA) measurement setup; (d) measured EVM with single carrier modulation signal

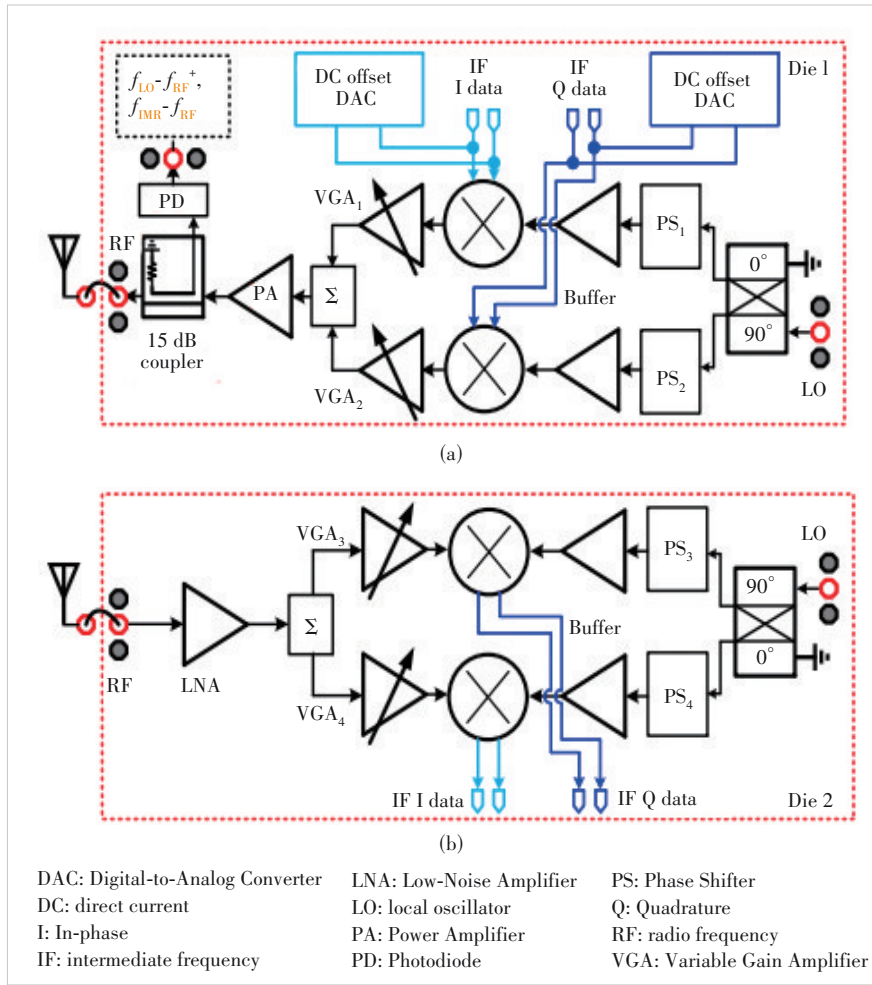


Figure 4. Block diagram of the 45-GHz direct-conversion chipset: (a) Tx chip and (b) Rx chip

lent 89600 vector signal analyzer was used to evaluate constellation and EVM performance. During the wireless data transmission measurement, the distance between the transmitter and the receiver was fixed at 1 m. At 42 GHz, 45 GHz, and 48.5 GHz, the measured EVMs of the SC 64-QAM signal (with a baud rate of 100 Mbit/s) were -29.11 dB, -29.44 dB, and -28.56 dB, respectively. As demonstrated in Fig. 5d, the measured 64-QAM constellations over the IEEE 802.11aj frequency band show high quality. Within the IEEE 802.11aj working band of 42.3 – 48.4 GHz, the transmitter achieved an output 1 dB compression point (OP1 dB) of 14.6 – 16.4 dBm, with a conversion gain (CG) exceeding 24 dB. Within the IEEE 802.11aj band, the receiver exhibited a noise figure of 3.8 – 4.1 dB, an input 1 dB compression point ranging from -22 dBm to -18.7 dBm, and a conversion gain greater than 43.8 dB. Overall, the measured transceiver fully meets the IEEE 802.11aj EVM requirements for 64-QAM modulation. In the OTA measurement with a T/R distance of 1 meter, the measured EVM from the transmitter to the receiver is better than -28.5 dB.

3.3 40 GHz Integer-N Phase-Locked Loop in CMOS Process

This subsection presents a 40 GHz integer-N phase-locked loop (PLL) featuring a linearized CMOS LC voltage-controlled oscillator (VCO), implemented in a standard 90-nm CMOS process. Its block diagram and micrograph are shown in Figs. 6 and 7, respectively^[33]. The VCO uses a tri-coupled inductor to couple the varactor diode pairs, achieving gain linearization through mutual compensation. The tri-coupled inductor eliminates the need for a tuning voltage offset circuit or DC-blocking capacitors, making it suitable for high-performance mmWave design.

Measurement results show the linear VCO has a tuning bandwidth of 15.8% and a phase noise of -100.7 dBc/Hz at 1 MHz offset, yielding a figure of merit (FOMT) of -181.8 dBc/Hz. The designed PLL, using this linear VCO, can lock stably from 38.61 GHz to 44.55 GHz. At 40 GHz, it exhibits an in-band phase noise of -81 dBc/Hz at 100 kHz offset and an out-of-band phase noise of -114.5 dBc/Hz at 10 MHz offset, with a total power consumption of 76 mW. Compared with other reported research, the proposed linearization method results in very small variations in VCO gain. The developed PLL demonstrates good stability and minimal loop bandwidth variation across its entire operating frequency range, suitable for wideband mmWave low phase noise LO signal generation.

3.4 Recent Advances in Chinese Commercial Millimeter-Wave Chipsets and Modules

After years of research and development, Chinese RF chip design companies have developed competitive silicon-based millimeter-wave multi-channel beamforming chips and up/down conversion chipsets. For instance, the MSTR111 and MSTR205 chips from MiSiC Microelectronics Co., Ltd. integrate eight RF transceiver channels in a single chip at 26/28 GHz (Fig. 8). The P1dB transmit power of a single channel reaches 20.5 dBm. Meanwhile, in Q-band (45 GHz), the MSTR201 (Fig. 9) supports the IEEE 802.11aj (45 GHz) standard. It integrates dual-channel transceivers, frequency conversion units, and a low-phase-noise PLL, with the P1dB of the transmit channel reaching 18 dBm and the noise figure (NF) of the receive channel below 6 dB, making it suitable for high-speed indoor millimeter-wave Wi-Fi application scenarios^[34]. The mmWave AP and terminal prototype have also been developed based

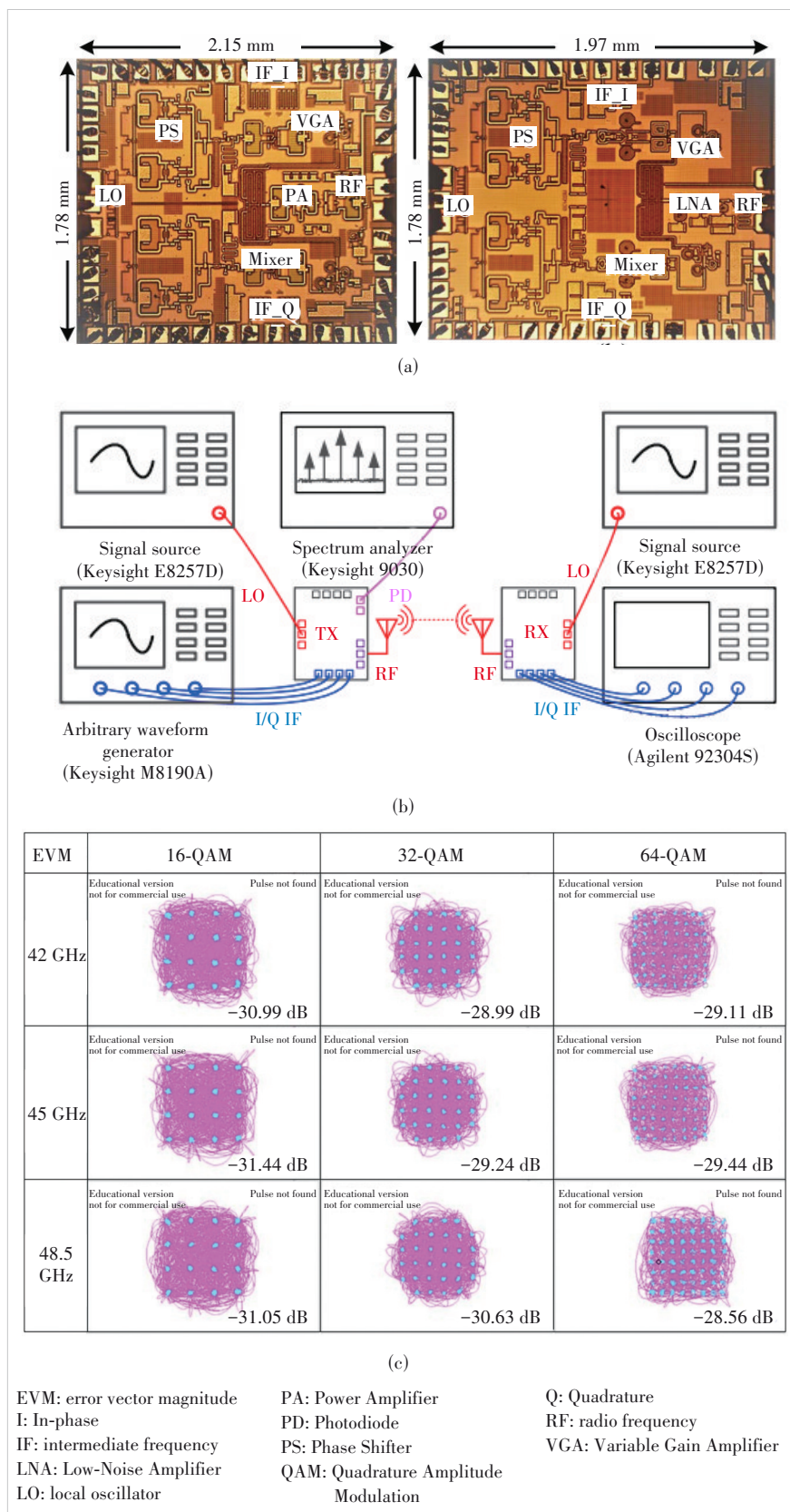


Figure 5. Chip micrographs and measurement results for the 45 GHz transceiver: (a) Tx chip; (b) Rx chip; (c) OTA measurement setup; (d) measured 64-QAM constellation

on the millimeter-wave Wi-Fi architecture by Nanjing Ziwei Technology Co., Ltd., as shown in Fig. 10. The system operates in the 24 – 26 GHz frequency band, adopting the millimeter-wave module and the commercial 802.11ax baseband. The system uses active phased array antennas, each containing 2×2 antenna units, which can cover a distance of over 100 meters, with a maximum data rate of 1.6 Gbit/s.

3.5 Challenges in Future MmW-Wi-Fi Applications

Compared to classical sub-6GHz Wi-Fi technology, mmWave Wi-Fi presents distinct challenges that require further research. Beamforming (BF) technologies improve mmWave communication capability by focusing the transmitted signal towards the receiver to provide additional link gain. To enhance coverage in the Q/V-band, the conventional approach is to use a phased array front-end with a moderate 16-element array size, which is a good balance between complexity and performance. In some application cases, the terminals rarely move after deployment, where the real-time BF can be omitted. In other cases, it needs multiple beam coverage for multiple high-throughput terminal users. The analog BF needs multiple sub-arrays for each beam. However, considering the cost and extra power consumption for multi-beam tracking, it is also a good choice to use several well-planned wider beams from the AP side to the mobile terminals coverage with MIMO techniques. Meanwhile, efficiently coordinating mmWave Wi-Fi with the classical Wi-Fi remains a significant challenge. This requires an advanced digital baseband chipset with integrated analog-to-digital and digital-to-analog converters (ADCs/DACs).

4 Conclusions

The integration of FTTR and mmWave technologies is a crucial means to achieve highly stable connections and low-latency transmission. The study of mmWave chip technology and system applications under the new FTTR architecture will be a significant research topic for the development of next-generation high-speed wire-

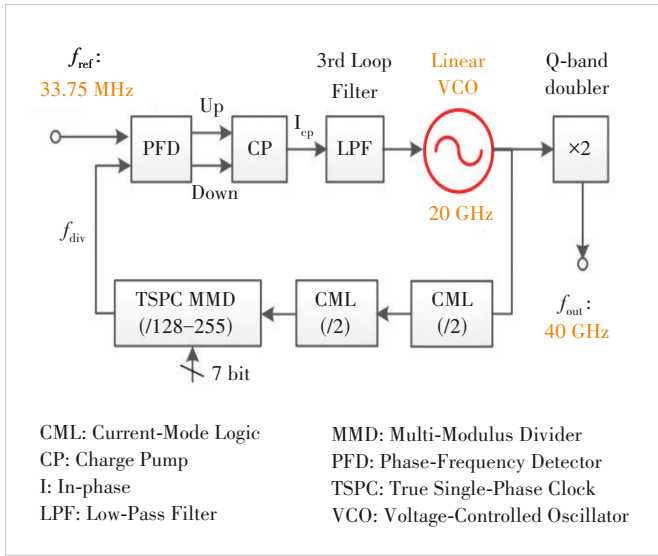


Figure 6. Block diagram of the 40 GHz Integer-N PLL

less access technologies in fixed access scenarios. Based on the advantages of the FTTR-mmWave architecture, it is of great research significance to design mmWave chips and systems with high transmission rates, low latency, high reliability, and low cost, while fully utilizing both sub-7.25 GHz and mmWave Wi-Fi frequency resources to achieve multi-band collaborative transmission. This work briefly reviews the U-

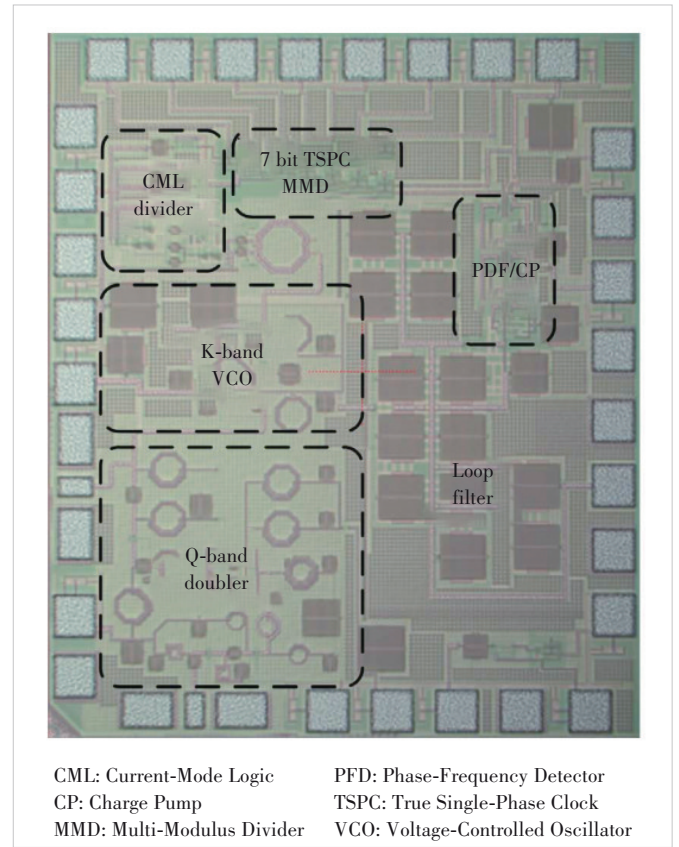


Figure 7. Micrograph of the 40 GHz Integer-N PLL

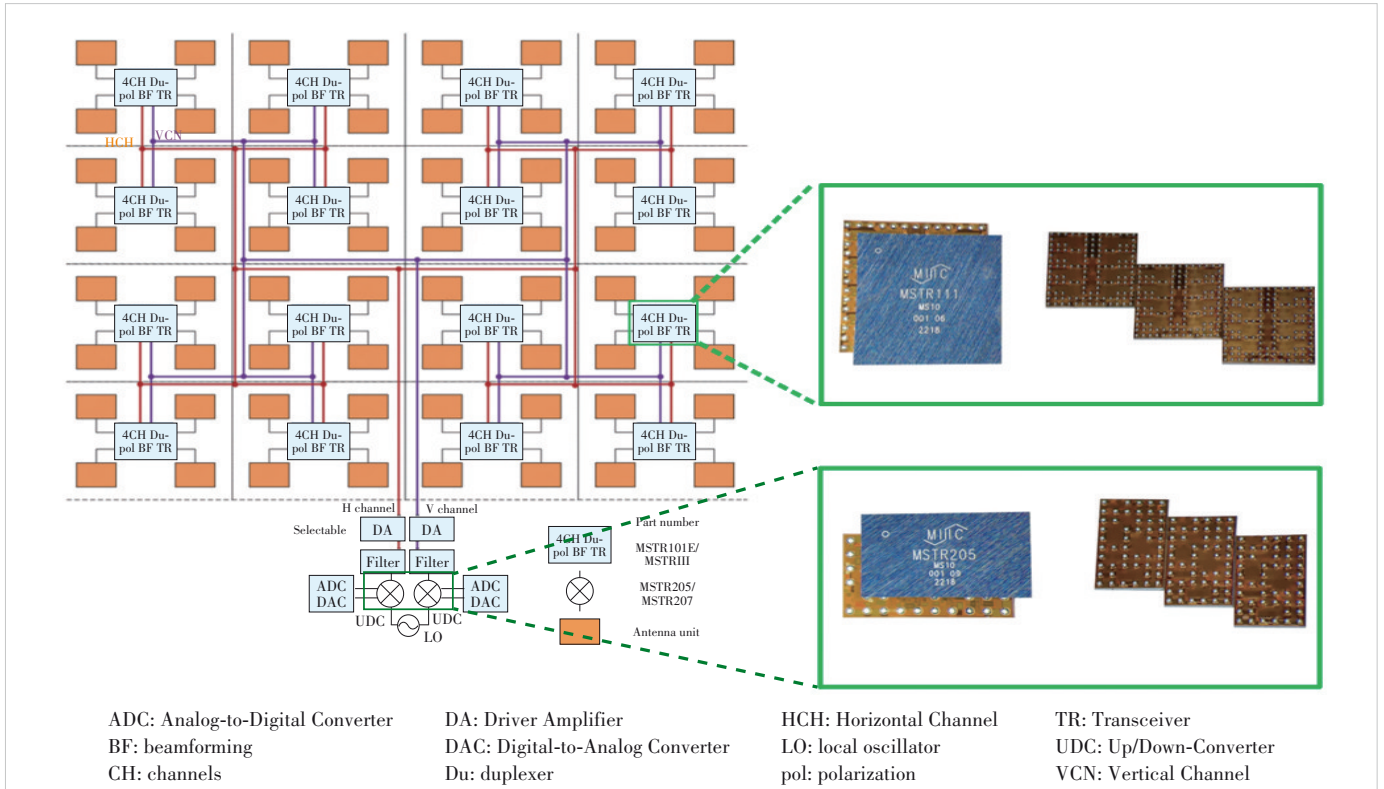


Figure 8. The 26/28-GHz 8-channel TR and Up/Down-conversion chipsets (Image courtesy of MiSic Microelectronics Co., Ltd.)

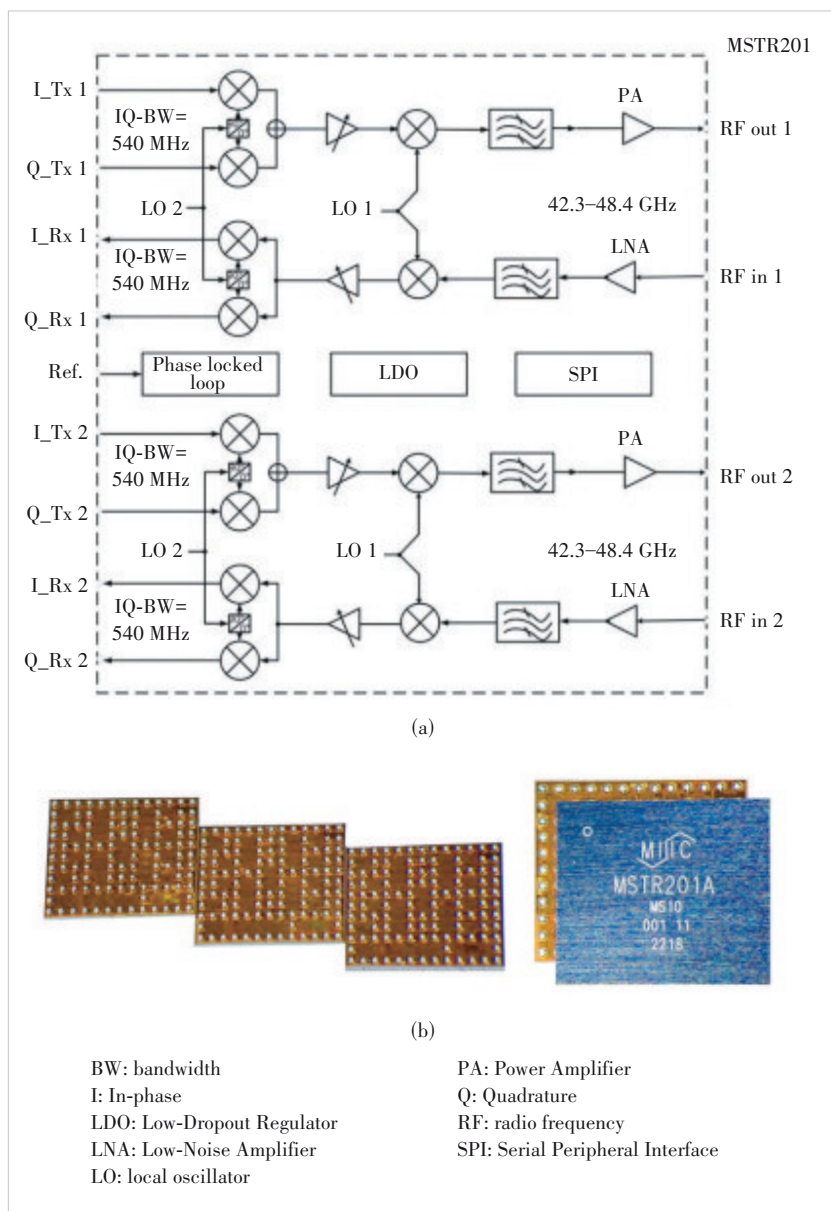


Figure 9. 45-GHz 2-channel transceiver chip for mmWave Wi-Fi (IEEE802.11aj) applications (Image courtesy of MiSic Microelectronics Co., Ltd.)



Figure 10. MmWave Wi-Fi modules for: (a) access point and (b) user terminal (Image courtesy of MiSic and Ziwei Tech., etc.)

CAN architecture with FTTR-mmWave Wi-Fi technology, as well as the recent Chinese mmWave chipset and RF module developments. The combination of fiber optics and mmWave will effectively solve problems related to indoor mmWave signal coverage, high-throughput traffic, and interference between wireless access points. As the key element in this system, the recent mmWave chip and RF module developments have made rapid progress, providing key support for indoor 10-giga-bit and higher-speed traffic and millisecond-level transmission services.

References

- [1] SCHNEIDER T, WIATREK A, PREÜBLER S, et al. Maximum transmittable data rates for millimeter-wave fixed wireless links [C]//Proc. 2012 ELEKTRO. IEEE, 2012: 94 - 98. DOI: 10.1109/ELEKTRO.2012.6225579
- [2] GHASEMPOUR Y, DA SILVA C R C M, CORDEIRO C, et al. IEEE 802.11ay: next-generation 60 GHz communication for 100 Gb/s Wi-Fi [J]. IEEE communications magazine, 2017, 55(12): 186 - 192. DOI: 10.1109/MCOM.2017.1700393
- [3] KIM M, ROPITAULT T, LEE S, et al. Efficient MU-MIMO beamforming protocol for IEEE 802.11ay WLANs [J]. IEEE communications letters, 2019, 23(1): 144 - 147. DOI: 10.1109/LCOMM.2018.2879476
- [4] KIM M S. Deep learning based multiple-beam transmission for IEEE 802.11ay beamforming training [C]//Proc. 27th Asia Pacific Conference on Communications (APCC). IEEE, 2022: 494 - 495. DOI: 10.1109/APCC55198.2022.9943719
- [5] ALDUBAIKHY K, WU W, ZHANG N, et al. MmWave IEEE 802.11ay for 5G fixed wireless access [J]. IEEE wireless communications, 2020, 27(2): 88 - 95. DOI: 10.1109/MWC.001.1900174
- [6] HE C, REN Z X, WANG X, et al. Millimeter-wave wireless communications for home network in fiber-to-the-room scenario [J]. Frontiers of information technology & electronic engineering, 2021, 22(4): 441 - 456. DOI: 10.1631/FITEE.2000440
- [7] IEEE. IEEE standard for information technology—local and metropolitan area networks—specific requirements part 15.3: amendment 2: millimeter-wave-based alternative physical layer extension: IEEE Std 802.15.3c-2009 [S]. 2009
- [8] SEYEDI A. On the physical layer performance of ecma-387: a standard for 60GHz WPANs [C]//Proc. IEEE International Conference on Ultra-Wideband. IEEE, 2009: 28 - 32. DOI: 10.1109/ICUWB.2009.5288820
- [9] IEEE. IEEE standard for information technology—telecommunications and information exchange between systems—local and metropolitan area networks—specific requirements part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: enhancements for very high throughput in the 60 GHz band: IEEE Std 802.11ad-2012 [S]. 2012

- [10] IEEE. IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: enhancements for very high throughput to support chinese millimeter wave frequency bands (60 GHz and 45 GHz): IEEE Std 802.11aj-2018 [S]. 2018
- [11] IEEE. IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements part 11: wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 2: enhanced throughput for operation in license-exempt bands above 45 GHz: IEEE Std 802.11ay-2021 [S]. 2021
- [12] AI B, GUAN K, HE R S, et al. On indoor millimeter wave massive MIMO channels: measurement and simulation [J]. IEEE journal on selected areas in communications, 2017, 35(7): 1678 – 1690. DOI: 10.1109/JSAC.2017.2698780
- [13] AKDENIZ M R, LIU Y P, SAMIMI M K, et al. Millimeter wave channel modeling and cellular capacity evaluation [J]. IEEE journal on selected areas in communications, 2014, 32(6): 1164 – 1179. DOI: 10.1109/JSAC.2014.2328154
- [14] ALKHATEEB A, EL AYACH O, LEUS G, et al. Channel estimation and hybrid precoding for millimeter wave cellular systems [J]. IEEE journal of selected topics in signal processing, 2014, 8(5): 831 – 846. DOI: 10.1109/JSTSP.2014.2334278
- [15] CHEN J N, LI S, TAO J Y, et al. Wireless beam modulation: an energy- and spectrum-efficient communication technology for future massive IoT systems [J]. IEEE wireless communications, 2020, 27(5): 60 – 66. DOI: 10.1109/MWC.001.2000021
- [16] CUI P F, ZHANG J A, LU W J, et al. Statistical sparse channel modeling for measured and simulated wireless temporal channels [J]. IEEE transactions on wireless communications, 2019, 18(12): 5868 – 5881. DOI: 10.1109/TWC.2019.2940017
- [17] GENC Z, DANG B L, WANG J, et al. Home networking at 60 GHz: challenges and research issues [C]//The International Federation for Information Processing (IFIP). Springer, 2008. DOI: 10.1007/978-0-387-77216-5_5
- [18] HUR S, KIM T, LOVE D J, et al. Millimeter wave beamforming for wireless backhaul and access in small cell networks [J]. IEEE transactions on communications, 2013, 61(10): 4391 – 4403. DOI: 10.1109/TCOMM.2013.090513.120848
- [19] MORAITIS N, CONSTANTINOU P. Indoor channel measurements and characterization at 60 GHz for wireless local area network applications [J]. IEEE transactions on antennas and propagation, 2004, 52(12): 3180 – 3189. DOI: 10.1109/TAP.2004.836422
- [20] RANGAN S, RAPPAPORT T S, ERKIP E. Millimeter-wave cellular wireless networks: potentials and challenges [J]. Proceedings of the IEEE, 2014, 102(3): 366-385. DOI: 10.1109/JPROC.2014.2299397
- [21] SATO K, MANABE T, IHARA T, et al. Measurements of reflection and transmission characteristics of interior structures of office building in the 60-GHz band [J]. IEEE transactions on antennas and propagation, 1997, 45(12): 1783 – 1792. DOI: 10.1109/8.650196
- [22] WU X Y, WANG C X, SUN J, et al. 60-GHz millimeter-wave channel measurements and modeling for indoor office environments [J]. IEEE transactions on antennas and propagation, 2017, 65(4): 1912 – 1924. DOI: 10.1109/TAP.2017.2669721
- [23] LIU X, LI J W, WU X M, et al. Fiber-to-the-room (FTTR) technologies for the 5th generation fixed network (F5G) and beyond [C]//Proc. IEEE Future Networks World Forum (FNWF). IEEE, 2022: 351 – 354. DOI: 10.1109/FNWF55208.2022.00068
- [24] CAI Y. Discussion on the construction of FTTR gigabit full optical network for smart home [J]. Telecom power technologies, 2022, 39(2): 92 – 94. DOI: 10.19399/j.cnki.tpt.2022.02.029
- [25] CHEN P N, ZHAO Z J, WANG X X. Research on the development and application of FTTR technology in the context of gigabit network development [J]. Telecom power technologies, 2021, 38(17): 85 – 87. DOI: 10.19399/j.cnki.tpt.2021.17.024
- [26] SUN L. U-CAN ultra-converged access network architecture and key technologies [C]//2024 National Conference on Microwave and Millimeter Wave. CIE, 2024. DOI: 10.26914/c.cnkihy.2024.019769
- [27] NOR A M. Access point selection in beyond 5G hybrid MmWave/Wi-Fi/Li-Fi network [J]. Physical communication, 2021, 46(6): 1 – 11. DOI: 10.1016/j.phycom.2021.101299
- [28] SUN X, YAO W B, MAO J H, et al. Research on 5G multi frequency coordination technology [J]. Telecom engineering technics and standardization, 2021, 34(3): 75 – 81. DOI: 10.13992/j.cnki.tetas.2021.03.017
- [29] HU T Y, LI L X, CHEN Z. New opportunities for multi-band collaborative communication: Integrated communications and sensing at terahertz band [J]. ZTE technology journal, 2022, 28(4): 14 – 18. DOI: 10.12142/ZTETJ.202204004
- [30] IEEE. IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: IEEE Std 802.11-1997 [S]. 1997
- [31] WANG L, CHEN J X, CHEN Z, et al. A 24 – 29.5-GHz scalable 2 × 2 I-Q TX/RX chipset with streamlined IF interfaces for DBF systems [J]. IEEE transactions on circuits and systems I: regular papers, 2025, 72 (1): 155 – 168. DOI: 10.1109/TCSI.2024.3438555
- [32] ZHOU P G, CHEN J X, YAN P P, et al. A -28.5-dB EVM 64-QAM 45-GHz transceiver for IEEE 802.11aj [J]. IEEE journal of solid-state circuits, 2021, 56(10): 3077 – 3093. DOI: 10.1109/JSSC.2021.3057077
- [33] CHEN Z, WANG M, CHEN J X, et al. Linear CMOS LC-VCO based on triple-coupled inductors and its application to 40-GHz phase-locked loop [J]. IEEE transactions on microwave theory and techniques, 2017, 65(8): 2977 – 2989. DOI: 10.1109/TMTT.2017.2663401
- [34] MiSiC Microelectronics Co., Ltd. Product center [DB/OL]. [2025-08-10]. <https://www.misic.com.cn>

Biographies

CHEN Zhe (zhechen@seu.edu.cn) received his BS degree in microwave engineering from University of Electronic Science and Technology of China in 2006, and his PhD degree in electronic engineering from Southeast University, China in 2014. From 2016 to 2018, he was with The University of Texas, USA. He is currently an associate professor with the State Key Laboratory of Millimeter Waves, School of Information Science and Technology, Southeast University. His research interests include terahertz signal sources and high-speed communications and millimeter wave integrated circuits in nanometer silicon-based technology. He served as the Session Chair for the IEEE International Symposium on Radio-Frequency Integration Technology (RFIT), the IEEE International Wireless Symposium (IWS), and the IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT).

ZHOU Peigen received his BS degree in radio engineering and PhD degree in electromagnetic field and microwave technique from Southeast University (SEU), China in 2015 and 2020, respectively. From 2019 to 2020, he was a visiting student researcher with Eindhoven University of Technology, The Netherlands. Since 2021, he has been with the State Key Laboratory of Millimeter Waves, SEU, where he is currently an associate professor with the School of Information Science and Engineering. His current research interests include silicon-based millimeter wave/terahertz on chip wireless communication/radar phased-array transceivers.

WANG Long received his BE degree from Hunan Institute of Science and Technology, China in 2015, MS degree from Hangzhou Dianzi University, China in 2018, and PhD degree from Southeast University, China in 2024. From 2018 to 2020, he was with MiSiC Microelectronics Technology Company Ltd., Nanjing, as a research engineer. His current research interests include silicon-based millimeter-wave integrated circuits and systems. He was a recipient of the IEEE International Wireless Symposium (IWS) Best Student Paper Award in 2021.

HOU Debin received his BS degree from the School of Physical Electronics, University of Electronic Science and Technology of China in 2007, and his PhD degree from the School of Information Science and Technology, Southeast University, China in 2013. He was with the Blekinge Institute of Technology (BTH), Sweden in 2009, and the Institute of Microelectronics (IME), Agency for Science, Technology and Research (A*STAR), Singapore, as an exchange student, from 2010 to 2012. From 2013 to 2022, he was with the State Key Laboratory of Millimeter Waves, Southeast University, China. He has authored over 20 technical publications. He is now with the MiSiC Microelectronics Co., Ltd., China. His current research interests include silicon based/GaAs millimeter wave/terahertz on-chip components, antennas, and integrated circuits. Dr. HOU received the Jiangsu Excellent 100 Doctoral Dissertation Prize in 2014.

HU Yun received his BE degree in electronics and information engineering and ME degree in circuits and systems from University of Electronic Science and Technology of China in 2008 and 2011, respectively, and his PhD degree in electromagnetic and microwave engineering from Southeast University, China in 2019. From July 2011 to February 2015, he was with the Second Academy of China Aerospace Science and Industry Corporation, as a research engineer. He is now with Nanjing Ziwei Tech. Co., Ltd., China. His current research interests include multibeam antennas, antenna arrays, beamforming networks, and RF front-end design. Dr. HU was a recipient of the Student Paper Award of the 2018 IEEE 7th Asia-Pacific Conference on Antennas and Propagation (APCAP 2018).

CHEN Jixin received his BS degree in radio engineering and MS and PhD degrees in electromagnetic field and microwave technique from Southeast University, China in 1998, 2002, and 2006, respectively. Since 1998, he has been with the State Key Laboratory of Millimeter-Waves, Southeast University, where he is currently a professor with the School of Information Science and Engineering and the director of the Department of Electromagnetic Field and Microwave Engineering. His current research interests include the integration technology of millimeter-wave circuits and systems. Dr. CHEN was a recipient of the 2016 Keysight Early Career Professor Award and the 2016 National Natural Science Prize Second Prize of China. He served as the TPC Chair for NCMMW2021 and RFIT2019 and the TPC Co-Chair for HSIC2012 and UCMMT2012.

HONG Wei received his BS degree from University of Information Engineering, China in 1982, and MS and PhD degrees from Southeast University, China in 1985 and 1988, respectively, all in radio engineering. Since 1988, he has been with the State Key Laboratory of Millimeter Waves, has served as the director of the lab since 2003, and is currently a professor at the School of Information Science and Engineering, Southeast University. In 1993, 1995, 1996, 1997 and 1998, he was a short-term Visiting Scholar with the University of California at Berkeley and at Santa Cruz, USA. He has been engaged in numerical methods for electromagnetic problems, millimeter-wave theory and technology, antennas, and RF technology for wireless communications. He has authored and co-authored over 400 technical publications and two books. He has received numerous awards, including the National Natural Science Award of China (twice), the first-class Science and Technology Progress Award from the Chinese Ministry of Education and the Jiangsu Provincial Government (four times), the 2023 CIE Science and Technology Innovation Team Award, and the 2021 IEEE MTT-S Microwave Prize. Dr. HONG is a Fellow of IEEE, a Fellow of CIE, the Vice President of both the CIE Microwave Society and the Antenna Society, and the Chair of the IEEE MTT-S/AP-S/EMC-S Joint Nanjing Chapter. He was an elected member of the IEEE MTT-S Administrative Committee (AdCom) from 2014 to 2016. He served as Associate Editor of *IEEE Transactions on Microwave Theory and Techniques* from 2007 to 2010 and as one of the guest editors for the 5G special issue of *IEEE Transactions on Antennas and Propagation* in 2017.



A Transformer-Based End-to-End Receiver Design for Wi-Fi 7 Physical Layer

LIU Yichen, GAO Ruixin, ZENG Chen, LIU Yingzhuang

(School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China)

DOI: 10.12142/ZTECOM.202504005

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251201.0957.002.html>,
published online December 1, 2025

Manuscript received: 2025-09-16

Abstract: The increasing demand for high throughput and low latency in Wi-Fi 7 necessitates a robust receiver design. Traditional receiver architectures, which rely on a cascade of complex, independent signal processing modules, often face performance bottlenecks. Rather than focusing on semantic-level tasks or simplified Additive White Gaussian Noise (AWGN) channels, this paper investigates a bit-level end-to-end receiver for a practical Wi-Fi 7 Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) physical layer. A lightweight Transformer-based encoder-only architecture is proposed to directly map synchronized OFDM signals to decoded bit-streams, replacing the conventional channel estimation, equalization, and data detection. By leveraging the multi-head self-attention mechanism of the Transformer encoder, our model effectively captures long-range spatial-temporal dependencies across antennas and subcarriers, thus learning to compensate for channel distortions without explicit channel state information. This mechanism eliminates the need for explicit channel estimation, enabling the direct extraction of crucial channel and signal features. Experimental results validate the efficacy of the proposed design, demonstrating the significant potential of deep learning for future wireless receiver architectures.

Keywords: Transformer; receiver design; Wi-Fi 7; deep learning

Citation (Format 1): LIU Y C, GAO R X, ZENG C, et al. A transformer-based end-to-end receiver design for Wi-Fi 7 physical layer [J]. *ZTE Communications*, 2025, 23(4): 27 – 36. DOI: 10.12142/ZTECOM.202504005

Citation (Format 2): Y. C. Liu, R. X. Gao, C. Zeng, et al., “A transformer-based end-to-end receiver design for Wi-Fi 7 physical layer,” *ZTE Communications*, vol. 23, no. 4, pp. 27 – 36, Dec. 2025. doi: 10.12142/ZTECOM.202504005.

1 Introduction

The proliferation of data-intensive applications, ranging from immersive virtual reality and high-definition streaming to industrial automation and the Internet of Things (IoT), has continuously driven the evolution of wireless communication standards. As the next generation of Wi-Fi technology, the IEEE 802.11be standard, commercially known as Wi-Fi 7^[1–2], is poised to meet these demands by delivering higher throughput, lower latency, and enhanced reliability. Key technological advancements such as 320 MHz channel bandwidth, Multi-Link Operation (MLO)^[3], and higher-order 4096-Quadrature Amplitude Modulation (QAM) are at the core of Wi-Fi 7’s performance gains. While these innovations push the theoretical limits of data transmission, they simultaneously introduce significant complexity to the physical layer (PHY) receiver design.

The traditional Wi-Fi PHY receiver architecture operates

as a cascaded chain of independent signal processing blocks, including time-frequency synchronization, channel estimation, equalization, and decoding. Each module is meticulously designed and optimized based on expert knowledge and mathematical models of the communication channel. However, this modular, block-by-block approach suffers from two fundamental limitations. First, the performance of each module is highly sensitive to the imperfections of its preceding stages, leading to a “domino effect” where errors propagate and accumulate. Second, the explicit channel estimation module, while crucial, can be computationally intensive and may not always accurately capture the complex, time-varying nature of wireless channels, especially in multi-path and multi-antenna environments. As Wi-Fi 7 leverages multi-antenna technologies, the intricate spatial and temporal correlations across the received signals present a formidable challenge that conventional methods struggle to address holistically.

In recent years, the paradigm of applying artificial intelligence (AI)/deep learning (DL) to wireless communication systems has attracted significant attention as a promising alternative to traditional model-based designs. Early works have dem-

This work was supported by the Huawei Technologies Co., Ltd. under Grant No. TP20250612004.

onstrated the potential of convolutional neural networks (CNNs)^[4], deep neural networks (DNNs) and recurrent neural networks (RNNs)^[5-7] for signal processing blocks such as channel modeling^[8], coding/decoding^[9], channel estimation^[10] and equalization^[11]. DONG et al. applied CNNs to millimeter-wave large-scale multiple-input multiple-output (MIMO) channel estimation^[4], effectively utilizing the local correlations of the channel in the space-frequency-time domain. However, CNNs rely on stacked convolution layers to perceive global information, and this inherent local processing mechanism makes it insufficient when directly modeling long-distance, non-local dependencies within the channel matrix. Ref. [5] used RNNs with bidirectional Long Short-Term Memory (LSTM) to recover data directly from the received signal, effectively simplifying the receiver design. However, the model's understanding of the entire Orthogonal Frequency Division Multiplexing (OFDM) symbol is based on sequential information propagation rather than a one-time global perception. Ref. [6] used RNNs to predict MIMO channels and performed offline training using complex hybrid evolutionary algorithms. However, the inherent sequential processing paradigm of RNNs limits their ability to capture long-range correlations in channel time series effectively, and their performance is heavily dependent on specific offline training processes, making it difficult to adapt to dynamic changes in the channel environment. Ref. [7] proposed a receiver framework based on DNNs that predicts future channel coefficients by learning the time-domain correlation of the channel, thereby effectively reducing pilot overhead. However, this method relies on a recurrent structure for sequence prediction, which makes it difficult for the model to capture long-term dependencies in channel changes; as a result, prediction errors accumulate over time.

As mentioned above, the cascaded architecture suffers from the performance limitations, and emphasis has shifted to the DL-based end-to-end (E2E) PHY design by interpreting a communication system as an autoencoder that jointly optimizes the transmitter and receiver over simplified channels^[12-14]. Ref. [12] provides the theoretical justification for introducing DL to PHY and incorporates Radio Transformer Networks to embed physical priors, thereby enhancing the model's interpretability and generalization ability. AIT AOUDIA and HOYDIS^[13] systematically evaluated E2E learning for OFDM systems under time-frequency selective fading and channel aging scenarios, and demonstrated that learning only the receiver can maintain bit error rate (BER) performance with sparse pilot tones. SONG et al.^[14] conducted a systematic benchmark evaluation of E2E autoencoders, re-evaluating the actual gains of autoencoders under more standardized training assumptions and improved baselines. However, they lacked a systematic assessment of complexity, scalability, and training costs. Moreover, these studies mainly consider single-antenna or small-scale systems with Additive White Gaussian Noise (AWGN) or simple fading

channels and predominantly rely on fully connected networks or CNNs rather than Transformer architectures.

The Transformer, originally a cornerstone of natural language processing, excels at processing sequential data by employing a multi-head self-attention mechanism^[15]. More recently, Transformer architectures have been extensively explored in the context of semantic communications and joint source-channel coding (JSCC). HUANG et al. proposed a JSCC framework for semantic communications of images^[16], which combines deep source coding with a hyper-prior model and conventional digital channel block coding, and derived a two-step rate control algorithm that adapts the source-channel rate split to the channel signal-to-noise ratio (SNR). For image transmission, BOURTSOULATZE et al. introduced a deep JSCC scheme^[9] that directly maps image pixels to complex channel symbols using a convolutional autoencoder, with the noisy channel implemented as a non-trainable layer in the network. In parallel, Transformer-based decoders have been proposed for algebraic block codes and Low-Density Parity-Check (LDPC) codes^[17], showing that self-attention can effectively capture the code structure and improve soft-decoding performance.

In contrast to the above studies, this paper targets the bit-level demodulation in a practical Wi-Fi MIMO-OFDM system and proposes a novel DL-based E2E receiver for the Wi-Fi 7 PHY layer, leveraging the power of the Transformer architecture, which can inherently capture long-range spatial and temporal dependencies across subcarriers and antennas of the synchronized OFDM symbols. By learning these intricate relationships, our proposed E2E model completely bypasses the need for an explicit channel estimation block, directly extracting crucial channel and signal features from the input data. This unified approach not only simplifies the receiver architecture but also offers a more robust and adaptive solution to the complexities of Wi-Fi 7 channels. Our experimental results validate the efficacy of this design. The key contributions of this work are summarized as follows:

- 1) We propose a receiver-only, encoder-only Transformer architecture for the Wi-Fi 7 MIMO-OFDM system, which operates directly on the synchronized multi-antenna OFDM frequency grid and outputs the coded bitstream, thereby replacing conventional channel estimation and equalization, and bit-detection chain while keeping a standard Wi-Fi 7 transmitter.
- 2) We design a complexity-aware lightweight head, consisting of a Convolutional Feature Enhancement Module (CFEM) and a compact bitstream recovery layer, which jointly exploit global (Transformer) and local (1D-CNN) features, making the model suitable for resource-constrained receivers.
- 3) We present a comprehensive experimental validation showcasing the superior performance of the proposed design in terms of BER compared to conventional methods.

The rest of this paper is organized as follows. Section 2 describes the system model and architecture. Section 3 presents

the proposed DL-based E2E receiver design. Following that, the algorithm verification and application experiment are described in Section 4. Finally, Section 5 concludes the paper.

2 System Model

We consider a single-user MIMO OFDM communication system, which serves as the equivalent model for our Wi-Fi 7 physical layer simulation. The number of transmit and receive antennas is configured with $N_t=4$, $N_r=4$. The traditional modular design of PHY transmission, from the transmitter to the receiver, is depicted in Fig. 1. The following subsections provide detailed descriptions of the key settings for the transmitter, the TGax Non-Line-of-Sight (NLOS) office channel model, and receiver.

2.1 Transmitter Design

At the transmitter, a PHY Service Data Unit (PSDU) stream is generated as the source message, followed by forward error correction coding (FEC), QAM, OFDM waveform generation, and other processes to generate the waveform to be transmitted. Specifically, the original data length of the PSDU is 2 792 bit. A 16 bit service field is then added to the data header, and after concatenation, a 2 808 bit payload is formed, which serves as the basic data unit for PHY transmission. Following this, the data is scrambled using a bitwise Exclusive OR operation with a pseudo-random sequence, and then forward error correction coding is applied using a low-density parity-check code, expanding the original data block into a 3 744 bit codeword. Finally, the data stream is modulated into data symbols using 16-QAM modulation, and OFDM symbol generation is completed through operations such as spatial stream partitioning and subcarrier mapping.

2.2 Channel Model

To simulate complex indoor environments and effectively characterize multipath effects and time-varying fading characteristics, we adopt the TGax NLOS office channel model (Model-D), as defined in the IEEE 802.11ax standard. Based on the Model-D channel, a large-scale shadow fading model characterized by a log-normal distribution is superimposed to simulate the path loss and occlusion attenuation of signals during long-distance transmission. This combined channel model con-

forms to the signal attenuation characteristics of typical office environments. In addition, to simulate a real noisy environment, AWGN is added to the channel output, and the SNR parameter is adjusted to achieve noise intensity control within the dynamic range of 0 – 34 dB, reproducing the signal transmission characteristics in complex wireless communication scenarios.

For each OFDM subcarrier, the baseband input-output relation can be written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k \quad (1),$$

where $\mathbf{x}_k \in \mathbb{C}^{N_s \times 1}$ is the collection of the N_s spatial data streams ($1 \leq N_s \leq 4$), $\mathbf{y}_k \in \mathbb{C}^{N_r \times 1}$ is the received signal vector, $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_s}$ is the effective MIMO channel matrix on subcarrier k , and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$ denotes AWGN. In our main simulations, we set $N_t = N_r = N_s = 4$, so that four spatial streams are transmitted simultaneously.

The MIMO channel \mathbf{H}_k is estimated in the traditional baseline receiver using the Wi-Fi 7 preamble structure. Specifically, least-squares (LS) channel estimation is performed based on the orthogonal pilot symbols transmitted in the Extremely High Throughput Long Training Field (EHT-LTF) fields, which enables recovery of the full 4×4 MIMO channel matrix across all subcarriers. The estimated channels are then used to construct a singular value decomposition (SVD)-based minimum mean square error (MMSE) equalizer, as detailed in Section 2.3. The spatial multiplexing characteristics of this multi-stream channel lead to a linear superposition of multipath components at the receiver, resulting in spatial correlation that must be handled by the equalizer or, in our proposed approach, by the Transformer-based neural receiver.

2.3 Traditional Receiver Design

The details of the traditional receiver scheme are described in this subsection. The receiver employs a multi-stage signal processing framework that incorporates time-frequency synchronization, channel estimation, and equalization to achieve precise signal recovery. The detailed flowchart of the receiver is shown in Fig. 2.

The initial synchronization phase aims to correct time and frequency deviations. By leveraging the periodic characteristics embedded in the pilot symbols, the system deploys an autocorrelation-based detection mechanism to precisely locate the starting boundary of the OFDM signal, thereby enabling packet detection and coarse timing. Subsequently, the carrier frequency offset (CFO) f_{offset} is precisely estimated by analyzing the phase difference between known periodic structures as follows:

$$\Delta\phi = \frac{\varphi}{2\pi} \quad (2),$$

$$f_{\text{offset}} = \Delta\phi \frac{f_s}{T_s}$$

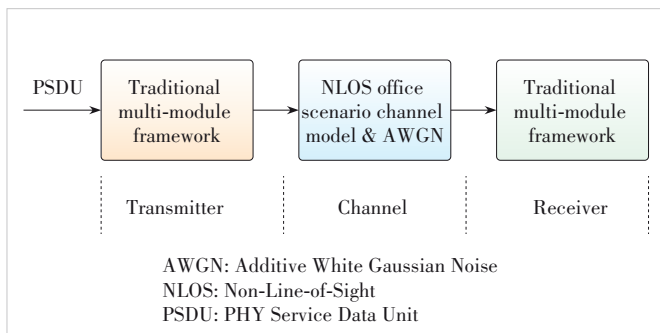


Figure 1. A traditional modular design of PHY transmission

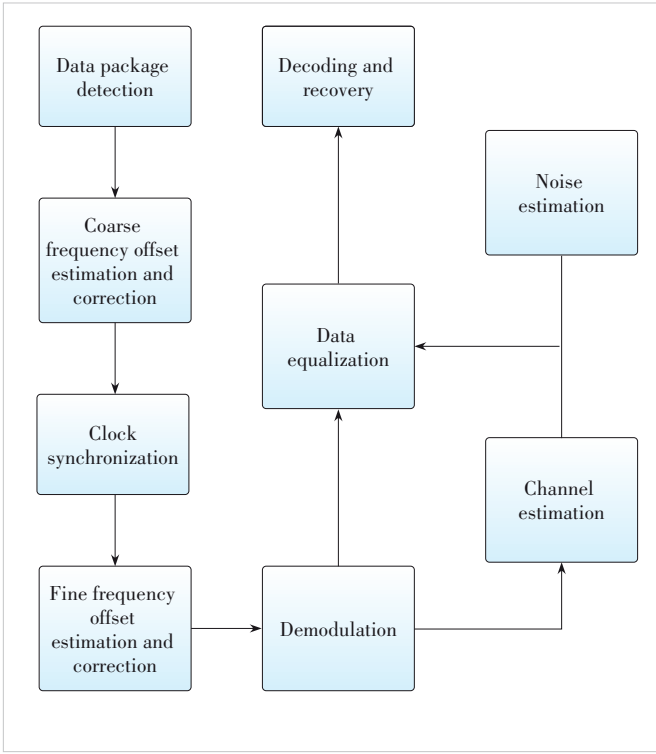


Figure 2. Operation flowchart of the receiver

where φ , $\Delta\phi$, f_s , and T_s are the phase angle, normalized phase difference, sampling rate, and symbol period, respectively. In the next stage, the channel frequency response (CFR) is estimated based on the Least Squares (LS) algorithm as follows:

$$\begin{aligned} Y &= HX + N \\ \hat{H} &= (X^H X)^{-1} X^H Y \end{aligned} \quad (3)$$

where X , H and N are the original transmitted signal, CFR, and AWGN, respectively. The average channel response \bar{H} is obtained by performing a time-domain average of the pilot estimate values. Finally, the noise variance σ_n^2 is calculated as follows:

$$\begin{aligned} \mu &= \hat{H} - \bar{H} \\ \sigma_n^2 &= I \cdot E[|\mu|^2] \end{aligned} \quad (4)$$

where I is the identity matrix. Using this estimate as a basis, SVD is first used to decompose \bar{H} into independent parallel subchannels, with each subchannel i corresponding to a singular value S_i :

$$\bar{H} = USV^H \quad (5)$$

where the columns of U span the received signal space, $S = \text{diag}(S_1, \dots, S_P)$ contains the singular values corresponding to the P spatial subchannels, and V^H is the orthogonal basis of

the transmitted signal space, mapping the equalized signal back to the transmitted signal space. We seek a linear equalizer W such that $\hat{x} = Wy$ minimizes the mean square error $\mathbb{E}\{\|x - Wy\|^2\}$. Under the standard assumption $\mathbb{E}\{xx^H\} = I$ and $\mathbb{E}\{nn^H\} = \sigma_n^2 I$, the MMSE solution is

$$W_{\text{MMSE}} = V(S^H S + \sigma_n^2 I)^{-1} S^H U^H \quad (6)$$

which assigns different weights to each singular mode according to its channel gain and the noise variance. Finally, the equalization matrix is applied to the received signal to obtain

$$\hat{x} = W_{\text{MMSE}} y = \sum_{i=1}^P \frac{S_i}{S_i^2 + \sigma_n^2} v_i (u_i^H y) \quad (7)$$

where S_i is the i -th singular value, and u_i and v_i are the i -th columns of U and V , respectively. This expression shows that the MIMO channel is decomposed into P parallel subchannels, each employing an MMSE scalar coefficient for equalization. In the traditional baseline, the subsequent step is to demodulate and decode \hat{x} to reconstruct the original transmitted bitstream. In contrast, our DL receiver replaces this entire LS+MMSE equalization and bit-detection chain with the proposed neural network.

3 Transformer-Based End-to-End Receiver Design

In a single-user 4×4 MIMO baseband transmission scenario, the spatial multiplexing characteristics of the channel cause the received signals to exhibit linear superposition of multipath propagation signals, resulting in spatial correlation. Traditional receiver architectures, such as those based on MMSE or zero-forcing equalizers, rely on the accuracy of channel state information (CSI) obtained through pilots, which introduces unavoidable pilot overhead; on the other hand, in typical Wi-Fi indoor environments, characterized by multipath fading, dynamic interference, shadow effects, and high-intensity noise, accurate CSI estimation is highly challenging, leading to a significant degradation in the performance of linear equalizers. Additionally, the complex scrambling and channel coding modules on the transmitter side, while enhancing link reliability, significantly increase the system's computational complexity and processing latency.

To address the limitations of traditional approaches, we explore the feasibility of modeling physical layer signal processing as an E2E-DL network. Specifically, given the Transformer model's strong global feature extraction capabilities in sequence data modeling, we propose a lightweight receiver based on a Transformer encoder, as shown in Fig. 3. At the transmitter (TX), a random TX PSDU bitstream is processed by a Wi-Fi 7 (IEEE 802.11be EHT) baseband chain and transmitted over a TGax NLOS MIMO-OFDM channel. At the re-

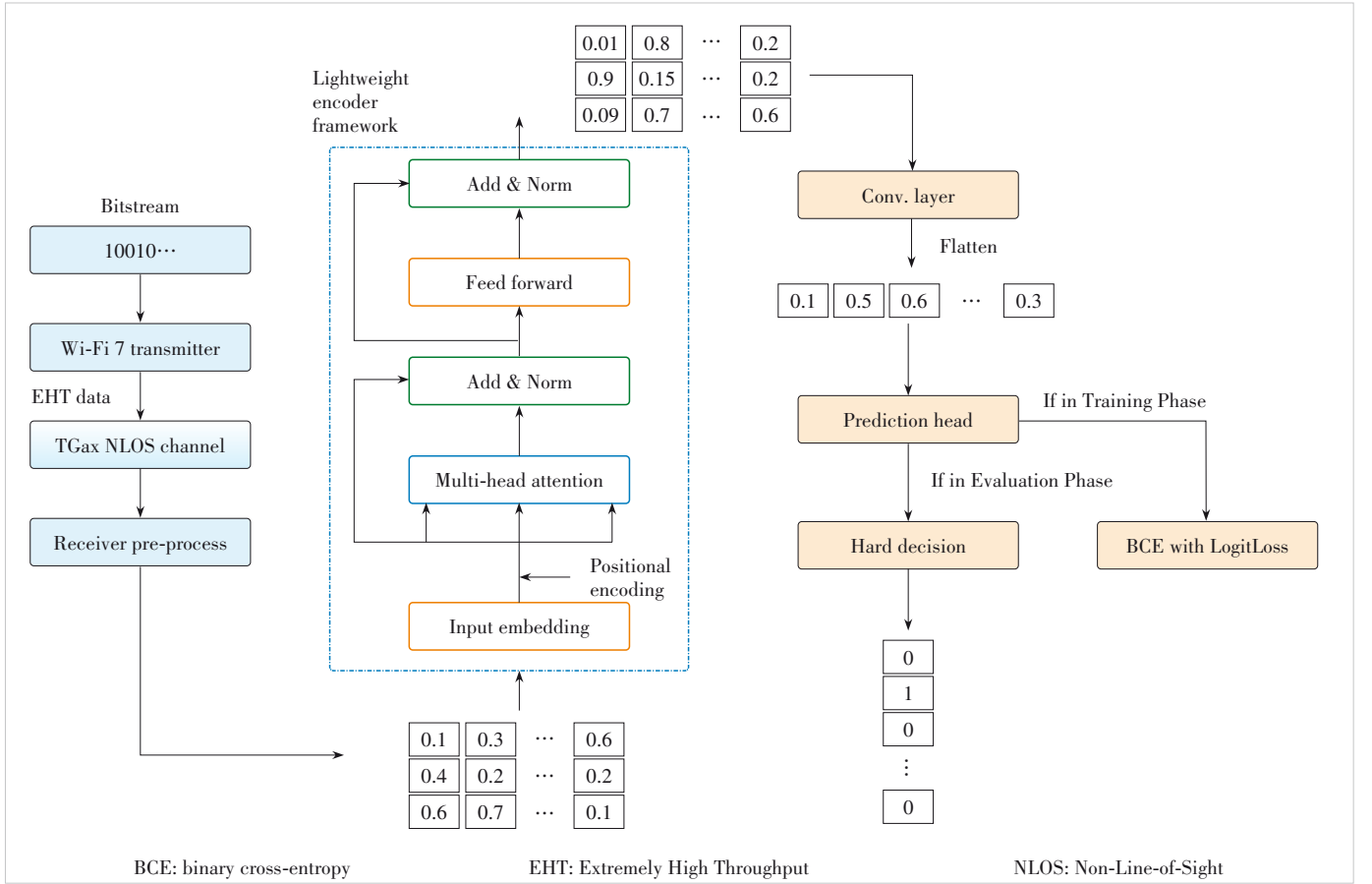


Figure 3. Lightweight encoder framework of the proposed Transformer-based receiver

ceiver side, the conventional front-end performs packet detection, time-frequency synchronization, Fast Fourier Transform (FFT), data demodulation, and phase tracking/correction. The resulting multi-antenna frequency-domain data symbols, after the above pre-processing stage, are used as the input to the proposed neural receiver, while the corresponding TX PSDU serves as the label. Thus, the network is trained to recover the original bitstream directly from the demodulated MIMO-OFDM data, effectively replacing the traditional chain of noise variance estimation, channel estimation and equalization, and bit-level detection.

The proposed neural receiver comprises three main components: an encoder-only Transformer, a CFEM, and a bitstream recovery layer. The Transformer encoder provides a robust representation of the MIMO-OFDM signal by exploiting multi-head self-attention over antenna and time/frequency dimensions. The CFEM further refines these features and reduces the effective sequence length. The final multi-layer perceptron (MLP)-based bitstream recovery layer maps the hierarchical features to bit-wise logits, which are trained with a binary cross-entropy with logits (BCE-with-Logits) loss and converted into hard decisions during evaluation. The following subsections will provide a detailed explanation of the model's net-

work structure and operational principles.

3.1 Lightweight Encoder Framework

The encoder consists of six layers of identical structures stacked together, with each layer containing four self-attention heads, residual connections, layer normalization, and a feed-forward network (FFN). This design focuses on modeling the spatial correlation of signals across antennas, processes MIMO-OFDM signals through multi-layer feature extraction and nonlinear transformations, and provides a highly robust feature representation for subsequent signal detection and demodulation.

3.1.1 Tokenization

The number of data subcarriers in the OFDM system is set to $N_s = 234$. Each antenna on each subcarrier carries one information symbol ($N_i = 1$). After synchronization and OFDM demodulation, the received signal yields a complex signal matrix $S_c \in \mathbb{R}^{N_s \times N_i}$. Since the model supports processing real-valued data, the complex numbers are decomposed into two independent dimensions based on their real and imaginary parts, resulting in the signal tensor $S_r \in \mathbb{R}^{N_s \times N_i \times 2}$. The signal received by each antenna can be regarded as a “token” for the

model input. By calculating the correlation weights between different antennas via self-attention, the model can dynamically learn which antenna signals are more important for restoring specific transmitted data.

The corresponding original randomly generated data bitstreams are directly used as supervision labels to construct the “signal feature-original bit” mapping pairs required for E2E training. The number of bits is $N_b = N_s \times N_i \times N_t = 936$. The data in the dataset is input into the model after being reshaped, with a signal dimension of inputs $\mathcal{I} \in \mathbb{R}^{N_d \times N_b \times 512}$, where N_d is the number of signals while also adapting to the original Transformer encoder input dimension [batchsize, sequencelength, d_{model}]. The supervised label dimension is $\mathcal{L} \in \mathbb{R}^{N_d \times N_b}$.

3.1.2 Input Embedding and Positional Encoding

In the embedding module, the encoder employs a linear projection layer to expand the input to 512 dimensions. This enhancement improves the model’s ability to represent features, allowing it to learn more complex representations. Based on empirical findings, the value of d_{model} is 512, demonstrating its feasibility and effectiveness in NLP and sequence modeling tasks. Following this, layer normalization is applied to each sample to stabilize the training process and help the model converge more effectively.

Finally, the Leaky Rectified Linear Unit (LeakyReLU) is used as the nonlinear activation function. Unlike the standard ReLU function that has a zero gradient in the negative value region, LeakyReLU maintains a slight positive slope in this interval^[18]. This design avoids the “dying ReLU” problem, where neurons fail to update their weights due to continuously outputting negative values and thus lead to permanent failure during training. The concrete formula is as follows:

$$X_{\text{embedding}} = \text{LeakyReLU}\left(\text{LayNorm}\left(W_e \cdot X_{\text{token}} + b_e\right)\right) \quad (8).$$

In the position encoding module, a cosine-based position encoding function is used to compensate for the Transformer’s lack of sensitivity to sequence order. This function generates position vectors corresponding to the antenna index to distinguish the spatial characteristics of different receiving antennas. The position vector, whose dimension is consistent with the embedding dimension, is added element-wise to the signal features. This explicitly embeds the position information into the model so that the model can better understand the spatial characteristics between antennas. The concrete formula is as follows:

$$\begin{aligned} \text{PE}_{(\text{pos}, 2i)} &= \sin(\text{pos}/10000^{2i/d_{\text{model}}}) \\ \text{PE}_{(\text{pos}, 2i+1)} &= \cos(\text{pos}/10000^{2i/d_{\text{model}}}) \end{aligned} \quad (9),$$

where $i \in [0, (d_{\text{model}} - 1)/2]$ represents the n -th element of the position vector.

3.1.3 Multi-Head Attention Mechanism

The multi-head attention mechanism can focus on feature correlations from different angles in the received signal, improving the model’s accuracy and robustness. This model uses a 4-head self-attention mechanism to process different feature subspaces in parallel and independently, which helps improve overall computational efficiency. Simultaneously, the model analyzes the input sequence from multiple perspectives to integrate a more comprehensive set of features. Finally, the outputs from each head are concatenated and subjected to linear projection as follows:

$$\begin{aligned} \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \\ \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \dots, \text{head}_4)W^O \end{aligned} \quad (10).$$

3.2 Convolutional Feature Enhancement Module

After the Transformer encoder completes the sequence modeling of the received signal, we design a CFEM further to extract the local spatio-temporal features of the signal. This hierarchical feature extraction strategy, which combines global and local information, significantly improves the accuracy of signal recovery.

The architecture of this module is as follows: first, a one-dimensional CNN (1D-CNN) processes the output features of the Transformer. The convolutional kernels in this layer map the input channel count from 4 to 16, with a stride of 4, thereby compressing the sequence length from 512 to 128. This operation not only expands the feature dimension but also serves as an effective downsampling mechanism to reduce computational complexity. Following the convolutional layer are two components: a BatchNorm layer to accelerate convergence and prevent overfitting, and a LeakyReLU activation function to introduce nonlinearity. Additionally, to ensure the integrity of the information flow and optimize the training process, we introduce a skip connection that adds the module’s input to the output after convolution, normalization, and activation on an element-wise basis. This residual structure preserves the original global features and provides a shortcut for gradient flow during backpropagation, thereby enhancing the network’s trainability.

3.3 Bitstream Recovery Layer

The bitstream prediction head serves as the terminal of the network, with its core task being to decode from hierarchical feature representations into the final bitstream. This module is implemented as a MLP, with the specific process as follows:

1) Feature integration and transformation

First, the feature tensor (16×128) output by the upstream convolutional module is flattened into a single 2048-dimensional vector, effectively integrating all the high-level features extracted by the network across antennas and time-frequency domains. Next, the first fully-connected layer (FC, also known as the feature transformation layer) linearly maps this vector to

3744 dimensions, aligning its dimension with the number of bits (N_b) in a single prediction.

2) Normalization and nonlinear activation

Before the nonlinear transformation, layer normalization is applied to normalize the distribution of activations, thereby accelerating convergence and improving training stability. We then use the LeakyReLU activation function. This choice is consistent with the selection for input embedding: LeakyReLU preserves a small gradient for negative inputs, enabling better handling and propagation of negative feature components related to signal phase. For phase-sensitive modulation schemes like 16-QAM, this property helps minimize information loss.

3) Regularization and output projection

The activated features pass through a Dropout layer with a dropout probability of 0.4, reducing the risk of overfitting in the FC layer. The second FC layer (output projection layer) then performs a final linear transformation of these features to refine the prediction results.

During the training phase, LogitsLoss is used to directly optimize the log odds output, avoiding the Sigmoid saturation problem. The loss expression is as follows:

$$L = \frac{1}{N} \left[-b_i \log \sigma(\hat{b}_i) - (1 - b_i) \log (1 - \sigma(\hat{b}_i)) \right] \quad (11).$$

During the testing phase, bitstreams $\hat{b} \in \{0, 1\}^{3744}$ are generated through hard decisions with a threshold of 0.5 to generate the final bits as follows:

$$\hat{b} = \begin{cases} 1, & f_{fc}(x)_i \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (12),$$

where $f_{fc}(x)_i$ is the FC layer function. Finally, the output results are compared with the label data bit by bit to count the number of errors, and the bit accuracy is calculated based on the total number of error bits as:

$$\text{BitAc} = 1 - \frac{\sum_{i=1}^{N_{\text{test}}} |\hat{b}_i - b_i|}{N_b \times N_{\text{test}}} \quad (13).$$

3.4 Complexity Analysis

The per-layer computational complexity of a standard multi-head self-attention block with sequence length L_{seq} , model dimension d_{model} , and H heads (each of size $d_{\text{head}} = d_{\text{model}}/H$) scales approximately as:

$$\mathcal{O}(L_{\text{seq}}^2 d_{\text{model}}) + \mathcal{O}(L_{\text{seq}} d_{\text{model}}^2) \quad (14),$$

where the first term corresponds to attention score computation (QK^T and softmax-weighted V), and the second term corresponds to the FFN. In our receiver, the effective sequence length is $L_{\text{seq}} = N_r \times N_{\text{sc}}$, e.g., four receive antennas and 234

active subcarriers give $L_{\text{seq}} = 936$. We note that the bit-length $N_{\text{bits}} = 3744$ is the output dimensionality of the final fully-connected head and does not directly enter the attention complexity. Moreover, to further reduce inference latency and memory footprint, we append CFEM that downsamples the sequence dimension by a stride s , and projects $d_{\text{model}} \rightarrow d_{\text{red}}$. We choose $d_{\text{red}} \in \left[\frac{1}{8} d_{\text{model}}, \frac{1}{2} d_{\text{model}} \right]$ and set $d_{\text{red}} = 128$ by default ($d_{\text{model}} = 512$), which is approximately one head width when $H = 4$ ($d_{\text{head}} = 128$). This choice preserves the information aggregated by the encoder and makes the final dense bit head operate on a compressed representation, rather than on the full $L_{\text{seq}} \times d_{\text{model}}$ tensor. In practice, this reduces the size of the final linear layer and lowers on-chip activation bandwidth.

4 Experiment

This section evaluates the performance of the proposed DL Transformer receiver against a traditional physical layer baseline via numerical simulation.

4.1 Experiment Settings

4.1.1 Baseline Scheme

The baseline was evaluated using the Monte Carlo method. We measured its BER performance over an SNR range of 0 to 34 dB, in 2 dB increments. To ensure statistical validity, each SNR point was assessed by transmitting a sufficient number of packets over independently realized random channels. In cases of packet detection failure, all bits within the packet were considered erroneous.

4.1.2 DL Transformer Receiver

1) Dataset generation: The proposed receiver was trained and evaluated on a large-scale simulated dataset. The data were generated using the same transmission link as the baseline, but with the transmitter's scrambling and channel coding modules disabled. To promote model generalization, each data sample corresponds to an independent channel realization. Both large-scale fading and small-scale fading were fully randomized to cover typical NLOS conditions.

2) Dataset configuration and preprocessing: The training and validation sets were generated at a fixed SNR of 30 dB, whereas dedicated test sets were generated for each evaluated SNR point (0 – 34 dB). Concretely, the training set consists of 1 200 mini-batches with 64 samples in each mini-batch, resulting in 76 800 training samples in total. The validation and test sets each contain 120 mini-batches with 64 samples per mini-batch, i.e., 7 680 samples per set, which corresponds to an approximate 10:1:1 ratio for training, validation, and test data. For each sample, we first generate a fresh random bitstream and then reset the channel model with new random seeds and parameters, such that the transmitter-receiver distance, path loss, shadowing, multipath delays, and fading co-

efficients (both large-scale and small-scale) are independently randomized. As a result, every sample is associated with an independent channel realization, covering a representative dynamic range of NLOS channel conditions and producing diverse fading characteristics, path loss levels, and noise realizations. For preprocessing, complex-valued signals were decomposed into their real and imaginary components. Furthermore, we applied data augmentation to the training set by adding zero-mean Gaussian noise ($\sigma = 0.05$) to enhance model robustness.

4.2 Training Strategy and Stabilization Techniques

To ensure stable and efficient training of the proposed neural receiver and to mitigate both underfitting and overfitting, we adopt the following training strategy.

1) Initialization and normalization

Network parameters are initialized according to the type of layer and activation. Layers with Gaussian Error Linear Unit (GELU) activations (e.g., in the Transformer encoder) use Xavier initialization, which balances the variance of forward and backward signals. Layers with LeakyReLU activations (e.g., in the CFEM) use Kaiming initialization, which is tailored to ReLU-type nonlinearities. For LayerNorm and BatchNorm, the scale parameters are initialized to 1 and the biases to 0 so that normalization does not distort the feature distribution at the beginning of training. The self-attention projection matrices in the Transformer are also initialized with Xavier to keep the variance of dot-product attention stable.

2) Optimizer and learning-rate scheduling

We use the AdamW optimizer to decouple weight decay from the gradient update, which provides more controlled regularization than classical Adam. Unless otherwise stated, the hyperparameters are set to weight decay 10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. A ReduceLROnPlateau scheduler monitors the validation loss and reduces the learning rate by a factor of 0.8 if no improvement is observed for five consecutive epochs. This strategy accelerates convergence in the early stage while allowing finer adjustments near convergence.

3) Gradient clipping

To avoid gradient explosion and overly large parameter updates, we apply gradient-norm clipping with a maximum l_2 norm of 2. This improves the stability of training, especially given the long input sequences and the depth of the encoder.

4) Regularization and stabilization

To enhance robustness and prevent overfitting, we combine several standard regularization techniques. L2 weight decay is applied through AdamW; Dropout and LayerNorm are used within the Transformer encoder and the fully connected layers; and BatchNorm is applied after convolutional layers in the CFEM to stabilize feature statistics across mini-batches.

With this training setup, the network learns to map the demodulated MIMO-OFDM symbols directly to the transmitted bitstream in an end-to-end fashion, enabling joint optimization

of signal detection and recovery under realistic Wi-Fi 7 channel conditions.

4.3 Experiment Results

The experiment used BER as the core evaluation indicator, defined as

$$\text{BER} = \frac{B_{\text{error}}}{B_{\text{total}}} \quad (15).$$

Fig. 4 shows the performance curve of the BER of the traditional scheme as a function of SNR. This curve exhibits three distinct regions. In the low SNR region (SNR < 10 dB), system performance is primarily constrained by intense channel noise, resulting in a sharp decline in BER from a high error level. Some errors originate from frequent packet detection failures during this phase. As the SNR enters the medium SNR range (10 dB to 26 dB), the slope of the BER curve slows significantly, indicating that system performance improvements are beginning to be constrained by non-noise factors. In the high SNR region (SNR > 26 dB), the BER decreases rapidly again, and when the SNR reaches 32 dB, the system approaches error-free transmission, achieving highly reliable communication.

Figs. 5a and 5b illustrate the learning dynamics of the proposed model over 1 300 epochs. The model exhibits a convergence pattern. In the initial training phase (approximately the first 200 epochs), the training and validation losses decrease sharply. Correspondingly, accuracy rises rapidly from the random guess baseline of 0.5, indicating that the model is effectively capturing the underlying data features. Following this initial phase, the learning process enters a stable convergence regime. A consistent generalization gap emerges between the training and validation curves, stabilizing in later epochs at approximately 0.05 for the loss and 0.045 for the accuracy.

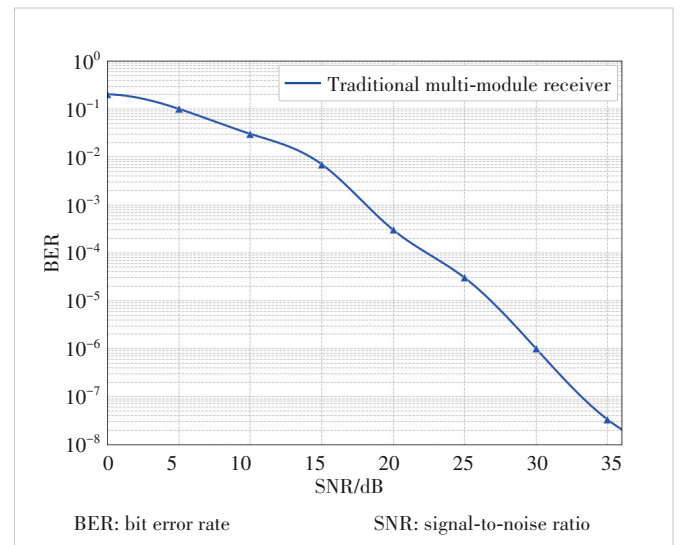


Figure 4. BER performance of the traditional scheme

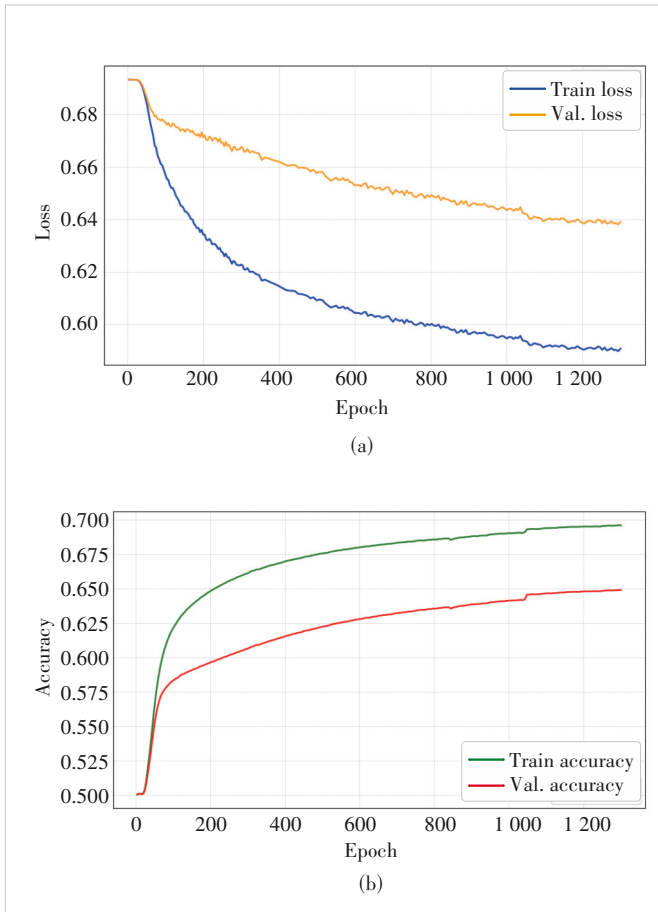


Figure 5. Training and validation curves: (a) loss versus epochs; (b) accuracy versus epochs

Crucially, the validation loss plateaus without a significant upward trend, which suggests that while minor overfitting is present, the model's generalization ability remains well-controlled. The minor fluctuations observed across the curves are attributable to the stochastic nature of minibatch gradient descent when applied to a complex and diverse dataset.

In summary, the leveling off of the validation metrics indicates that the model's performance has likely saturated, given its capacity and the complexity of the dataset. While this confirms the effectiveness of the proposed architecture, it also suggests that its performance ceiling is constrained by either its representational power or the intrinsic noise within the data.

Fig. 6 shows the BER performance of the proposed deep learning receiver, demonstrating that the model achieves excellent signal recovery capability. As shown in the figure, the BER begins to decline when the SNR reaches approximately 10 dB, demonstrating that the model has successfully learned effective features capable of robustly countering channel noise and fading. When the SNR exceeds 23 dB, the decline rate of the BER curve significantly slows down, exhibiting a conver-

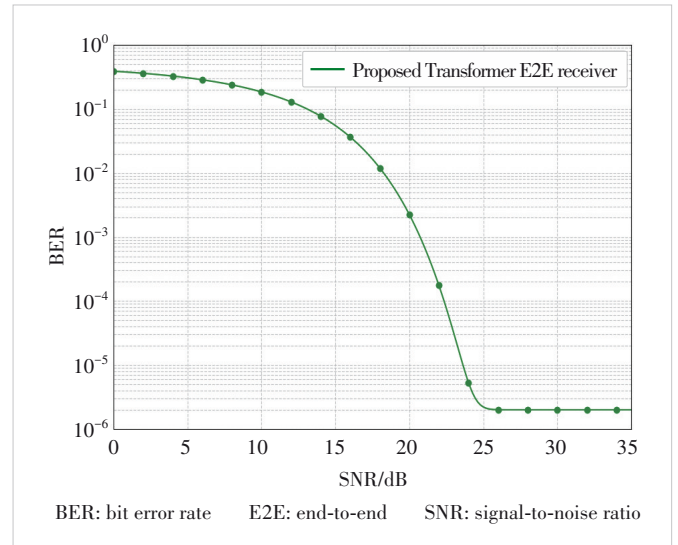


Figure 6. BER performance of the proposed scheme

gent trend, indicating that the model has reached its performance limit. Overall, these results demonstrate that the proposed deep learning model can successfully learn the E2E signal recovery task without explicit channel estimation and equalization modules. The model exhibits strong robustness across the entire evaluated SNR range, with its achieved low BER meeting the performance standards for high-reliability communication.

5 Conclusions

This paper proposes and validates a novel DL-E2E receiver based on the Transformer architecture, designed to overcome the limitations of traditional multi-module designs in Wi-Fi 7 environments. Our work demonstrates that the model's multi-head self-attention mechanism is effective at implicitly learning and jointly performing noise estimation, channel estimation, and equalization, thus bypassing the need for the explicit modular design. The experimental results underscore the superiority of this data-driven paradigm: The proposed receiver achieves an excellent BER performance across a wide range of SNRs, meeting the requirements for high-reliability communication and serving as a compelling proof-of-concept for AI-native wireless systems.

Building on these promising results, future efforts will be directed toward two critical areas. First, we will investigate the practical implementation and deployment of this architecture, focusing on complexity reduction and optimization for real-time processing on hardware platforms. Second, we will pursue research into the theoretical interpretability of the network's learned features, aiming to move beyond a "black-box" understanding and gain deeper insights into how the model makes its decoding decisions.

References

- [1] DENG C L, FANG X M, HAN X, et al. IEEE 802.11be Wi-Fi 7: new challenges and opportunities [J]. *IEEE communications surveys & tutorials*, 2020, 22(4): 2136 – 2166. DOI: 10.1109/COMST.2020.3012715
- [2] KHOROV E, LEVITSKY I, AKYILDIZ I F. Current status and directions of IEEE 802.11be, the future Wi-Fi 7 [J]. *IEEE access*, 2020, 8: 88664 – 88688. DOI:10.1109/ACCESS.2020.2993448
- [3] MENG J, ZHAO Q L, WU W M, et al. Enhancing IEEE 802.11ax network performance: an investigation and modeling into multi-user transmission [J]. *IEEE transactions on mobile computing*, 2025, 24(3): 2151 – 2165. DOI: 10.1109/TMC.2024.3493032
- [4] DONG P H, ZHANG H, LI G Y, et al. Deep CNN-based channel estimation for mmWave massive MIMO systems [J]. *IEEE journal of selected topics in signal processing*, 2019, 13(5): 989 – 1000. DOI: 10.1109/JSTSP.2019.2925975
- [5] WANG S Y, YAO R G, TSIFTIS T A, et al. Signal detection in uplink time-varying OFDM systems using RNN with bidirectional LSTM [J]. *IEEE wireless communications letters*, 2020, 9(11): 1947 – 1951. DOI: 10.1109/LWC.2020.3009170
- [6] POTTER C, VENAYAGAMOORTHY G K, KOSBAR K. RNN based MIMO channel prediction [J]. *Signal processing*, 2010, 90(2): 440 – 450. DOI: 10.1016/j.sigpro.2009.07.013
- [7] MATTU S R, THEAGARAJAN L N, CHOCKALINGAM A. Deep channel prediction: a DNN framework for receiver design in time-varying fading channels [J]. *IEEE transactions on vehicular technology*, 2022, 71(6): 6439 – 6453. DOI: 10.1109/TVT.2022.3162887
- [8] YANG Y, LI Y, ZHANG W X, et al. Generative-adversarial-network-based wireless channel modeling: challenges and opportunities [J]. *IEEE communications magazine*, 2019, 57(3): 22 – 27. DOI: 10.1109/MCOM.2019.1800635
- [9] BOURTSOULATZE E, BURTH KURKA D, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission [J]. *IEEE transactions on cognitive communications and networking*, 2019, 5(3): 567 – 579. DOI: 10.1109/TCCN.2019.2919300
- [10] HE H T, WEN C K, JIN S, et al. Deep learning-based channel estimation for beamspace mmWave massive MIMO systems [J]. *IEEE wireless communications letters*, 2018, 7(5): 852 – 855. DOI: 10.1109/LWC.2018.2832128
- [11] CHEN J, WANG X B. Learning-based intermittent CSI estimation with adaptive intervals in integrated sensing and communication systems [J]. *IEEE journal of selected topics in signal processing*, 2024, 18(5): 917 – 932. DOI: 10.1109/JSTSP.2024.3468037
- [12] O'SHEA T, HOYDIS J. An introduction to deep learning for the physical layer [J]. *IEEE transactions on cognitive communications and networking*, 2017, 3(4): 563 – 575. DOI: 10.1109/TCCN.2017.2758370
- [13] AIT AOUDIA F, HOYDIS J. End-to-end learning for OFDM: from neural receivers to pilotless communication [J]. *IEEE transactions on wireless communications*, 2022, 21(2): 1049 – 1063. DOI: 10.1109/TWC.2021.3101364
- [14] SONG J X, HÄGER C, SCHRÖDER J, et al. Benchmarking and interpreting end-to-end learning of MIMO and multi-user communication [J]. *IEEE transactions on wireless communications*, 2022, 21(9): 7287 – 7298. DOI: 10.1109/TWC.2022.3157467
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proc. 31st International Conference on Neural Information Processing Systems (NIPS'17). ACM, 2017: 6000 – 6010
- [16] HUANG J H, YUAN K, HUANG C, et al. D2-JSCC: digital deep joint source-channel coding for semantic communications [J]. *IEEE journal on selected areas in communications*, 2025, 43(4): 1246 – 1261. DOI: 10.1109/JSAC.2025.3531546.
- [17] CHOUKROUN Y, WOLF L. Error correction code transformer [C]//Proc. 36th International Conference on Neural Information Processing Systems (NIPS'22). ACM, 2022: 38695 – 38705
- [18] XU J, LI Z S, DU B W, et al. Reluplex made more practical: leaky ReLU [C]//Proc. IEEE Symposium on Computers and Communications (ISCC). IEEE, 2020: 1 – 7. DOI: 10.1109/ISCC50000.2020.9219587

Biographies

LIU Yichen is currently pursuing the PhD degree in the School of Electronic Information and Communications, Huazhong University of Science and Technology, China. His research interests include wireless communications, artificial intelligence, and WLAN systems.

GAO Ruixin is currently pursuing the MS degree in the School of Electronic Information and Communications, Huazhong University of Science and Technology, China. Her research interests include machine learning for wireless communications, artificial intelligence, and WLAN systems.

ZENG Chen is currently pursuing the PhD degree in the School of Electronic Information and Communications, Huazhong University of Science and Technology, China. His research interests include machine learning for wireless communications, artificial intelligence, and WLAN systems.

LIU Yingzhuang (liuyz@hust.edu.cn) is currently a professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology, China. Prior to that, he was a postdoctoral researcher with University of Paris XI, France from 2000 to 2001. Since 2003, he has led more than 10 national key projects, published more than 100 papers, and obtained more than 50 patents in broadband wireless communications. His main research interests are in broadband wireless communications, including LTE-Advanced, 5G/6G, and WLAN systems.

Root Cause Analysis of Poor FTTR Quality Based on Transformer Mechanisms



YU Weichao, LIU Yang, ZHANG Junxiong,

YE Junliang, GE Xiaohu

(School of Electronic Information and Communication, Huazhong University of Science and Technology, Wuhan 430074, China)

DOI: 10.12142/ZTECOM.202504006

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251209.1014.002.html>,
published online December 09, 2025

Manuscript received: 2025-11-11

Abstract: Fiber-to-the-Room (FTTR) has emerged as the core architecture for next-generation home and enterprise networks, offering gigabit-level bandwidth and seamless wireless coverage. However, the complex multi-device topology of FTTR networks presents significant challenges in identifying sources of network performance degradation and conducting accurate root cause analysis. Conventional approaches often fail to deliver efficient and precise operational improvements. To address this issue, this paper proposes a Transformer-based multi-task learning model designed for automated root cause analysis in FTTR environments. The model integrates multidimensional time-series data collected from access points (APs), enabling the simultaneous detection of APs experiencing performance degradation and the classification of underlying root causes, such as weak signal coverage, network congestion, and signal interference. To facilitate model training and evaluation, a multi-label dataset is generated using a discrete-event simulation platform implemented in MATLAB. Experimental results demonstrate that the proposed Transformer-based multi-task learning model achieves a root cause classification accuracy of 96.75%, significantly outperforming baseline models including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Random Forest, and eXtreme Gradient Boosting (XGBoost). This approach enables the rapid identification of performance degradation causes in FTTR networks, offering actionable insights for network optimization, reduced operational costs, and enhanced user experience.

Keywords: FTTR; root cause analysis; Transformer mechanisms; Wi-Fi; multi-task learning

Citation (Format 1): YU W C, LIU Y, ZHANG J X, et al. Root cause analysis of poor FTTR quality based on Transformer mechanisms [J]. *ZTE Communications*, 2025, 23(4): 37 – 47. DOI: 10.12142/ZTECOM.202504006

Citation (Format 2): W. C. Yu, Y. Liu, J. X. Zhang, et al., “Root cause analysis of poor FTTR quality based on Transformer mechanisms,” *ZTE Communications*, vol. 23, no. 4, pp. 37 – 47, Dec. 2025. doi: 10.12142/ZTECOM.202504006.

1 Introduction

Fiber-to-the-Room (FTTR)^[1] is a key framework for next-generation home and enterprise networks, with global deployment accelerating due to its broad application prospects. Built on an all-optical fiber network architecture, this technology delivers ultra-high bandwidth (ranging from 1 Gbit/s to 10 Gbit/s), low latency, high stability, and seamless coverage to end users. It effectively supports bandwidth- and performance-intensive applications such as 4K/8K ultra-high-definition video streaming, cloud gaming, virtual reality/augmented reality (VR/AR), smart home systems, and remote work and education^[2]. As global fixed broadband networks progress toward gigabit and ten-gigabit capabilities, the deployment scale of FTTR continues to expand, accompanied by a steady increase in market penetration. This trend under-

scores FTTR as a critical direction in the evolution of next-generation broadband access technologies.

In December 2020, the European Telecommunications Standards Institute (ETSI) included FTTR in the fifth-generation fixed network (F5G) standard system^[3], reflecting international recognition of its technological innovation and standardization progress. Currently, ten-gigabit access solutions that integrate 50G passive optical networks (PON) and FTTR are progressing through pilot testing and commercial deployment stages. Many service providers are actively promoting this technology as a strategic initiative to enhance user experience, capture high-end market segments, and strengthen competitive advantage. Furthermore, as end-users increasingly demand high-quality indoor coverage and stable network performance, achieving high reliability, efficient operation and maintenance, and consistently superior user experiences in FTTR networks has become a central focus for both academic research and industrial development.

The rapid expansion of FTTR networks, coupled with their highly complex topology, which includes primary gateways,

This work is supported in part by the National Key R&D Program of China under Grant No. 2024YFE0200504, NSFC key international joint project under Grant No. 62120106007, and Interdisciplinary Research Program of HUST under Grant No. 2024JCJY022.

secondary gateways, optical splitters, fiber links, and Wi-Fi environments, has introduced new operational challenges. First, fault localization is challenging. Issues such as slow network speed, connection interruptions, and high latency may originate from multiple sources, including abnormal optical power, hardware or software failures, configuration errors, Wi-Fi interference, or user-side problems. Actually, PON-side issues do not directly cause Wi-Fi performance degradation but rather affect the performance of the entire end-to-end link. For instance, optical power attenuation can lead to high bit error rates and packet retransmission in the link, while optical line terminal (OLT) congestion can result in increased latency and packet loss. These issues degrade the Quality of Service (QoS) for data flows. Since Wi-Fi is the last segment of the network connecting to users, problems occurring in intermediate links are sometimes erroneously attributed to Wi-Fi performance degradation. Moreover, faults often exhibit complex interdependencies across multi-level devices, rendering manual inspection inefficient and error-prone. Second, root cause analysis (RCA) is time-intensive. Traditional approaches that rely on alarms and expert knowledge struggle to quickly and accurately identify the true cause amid vast volumes of multi-source data, such as performance metrics, alarm records, configuration logs, and environmental information. This results in prolonged mean time to repair (MTTR), which negatively affects user experience and operator profitability. Third, intelligent capabilities are limited. Current operational systems lack automated and intelligent analysis functions, including root cause inference, making them ill-suited for the efficient and precise management of large-scale FTTR networks. Therefore, establishing an automated operational decision-making cycle has become an urgent and critical priority.

The accelerating pace of enterprise digital transformation, combined with the deep integration of technologies such as cloud computing, big data, and artificial intelligence, has led to a significant increase in the complexity of intelligent systems. During daily business operations and the maintenance of core systems, organizations may experience system failures or performance degradation. To ensure continuous and stable system operation, it is essential to accurately identify the root cause of issues—not only locating the fault but also understanding the underlying reasons—to develop effective preventive strategies. This process is known as RCA, which focuses on employing systematic methodologies to uncover the fundamental factors behind problems, rather than simply addressing surface-level symptoms. Currently, RCA methods can be broadly classified into two categories: data-driven and causality-driven approaches. Due to the high complexity involved in achieving complete causal inference, data-driven methods, such as association rule mining, machine learning, deep learning, intelligent agents, and knowledge graphs, have become the dominant approach in practice.

The data utilized for RCA primarily consists of three types

of heterogeneous data: 1) Location-time data: This type records the physical or logical location and associated timing of fault occurrences, thereby supporting the identification of failure propagation paths. For example, in FTTR networks, this includes AP deployment topology, user mobility trajectories, and timestamps of roaming events; 2) Physical data: This category captures the physical state of the system. Wireless metrics such as the received signal strength indicator (RSSI), the signal-to-noise ratio (SNR), and air interface utilization fall into this category; 3) Log and behavioral data: This type enables in-depth causal inference and includes examples such as device kernel logs, 802.11 protocol packet captures, and user authentication records. However, the heterogeneity of multi-source data and the presence of temporal delays remain key challenges. There is an urgent need to achieve efficient feature alignment and real-time analytical capabilities.

Among data-driven methods, association rule mining aims to identify abnormal combinations of attributes to locate root causes. For example, the classification based on associations (CBA) algorithm utilizes class association rules (CARs) to identify fault causes, demonstrating high accuracy in applications such as fault diagnosis in chiller systems^[4]. However, the effectiveness of this method heavily depends on the setting of thresholds for minimum support and minimum confidence. Its performance tends to be unstable across different datasets, and it is prone to combinatorial explosion in high-dimensional scenarios^[5–6].

Machine learning methods are extensively employed in environments with labeled data, where they demonstrate strong performance. Neural network classifiers and k-nearest neighbors (KNN) algorithms can construct feature vectors from historical data to train supervised classification models, enabling accurate fault attribution^[7]. However, this approach is heavily dependent on the availability of large quantities of high-quality labeled samples. In unsupervised or semi-supervised scenarios, the interpretability and performance of these methods are often constrained.

Deep learning methods enable automatic feature extraction and pattern recognition from high-dimensional and complex data. They demonstrate superior performance in multimodal fusion and sequence modeling, allowing the application of hybrid architectures such as CNN-BiLSTM-Attention for processing alarm sequences or leveraging Transformer-based models to integrate multi-source information for fault diagnosis^[8–10]. While deep learning is highly sensitive to nonlinear relationships and latent patterns, its effectiveness in causal inference remains dependent on the comprehensiveness of feature extraction and the quality of the input data. Furthermore, model training typically demands large volumes of labeled data and significant computational resources^[11–12].

In recent years, knowledge graphs and intelligent agent methods have been gradually applied to RCA. Knowledge graphs enable interpretable reasoning through structured

causal networks and probabilistic models. Agents built on large language models (LLMs), such as ReAct and RCagent, utilize interactive tool calls and dynamic reasoning loops to cope with complex and dynamic fault scenarios, significantly enhancing system interpretability and real-time responsiveness^[13-14]. These methods exhibit complementary characteristics. Association rule mining is suitable for extracting highly interpretable rules. Machine learning is effective for well-labeled classification tasks. Deep learning excels at handling high-dimensional and complex patterns. Knowledge graphs and agent-based methods enhance causal reasoning and interactive adaptability, thereby improving the interpretability and dynamic performance of RCA.

This paper, based on artificial intelligence methods, addresses the challenge of Wi-Fi quality degradation in FTTR networks caused by complex factors. The main contributions are as follows:

- 1) To tackle the difficulty of locating Wi-Fi-side quality degradation and performing RCA in FTTR networks, a multi-task learning model based on Transformer mechanisms is proposed, enabling simultaneous AP localization and root cause type identification.
- 2) A discrete-event simulation platform using MATLAB is designed and implemented to simulate various network degradation scenarios (e.g., weak coverage, congestion, and interference), generating a multi-label dataset for model training and validation.
- 3) Experimental results show that the proposed Transformer-based multi-task learning model achieves an accuracy of

96.75% in root cause classification tasks, significantly outperforming baseline models such as LSTM, GRU, Random Forest, and XGBoost. This demonstrates its superiority in temporal feature extraction and complex pattern recognition, highlighting its high practical value.

The rest of this paper is organized as follows. Section 2 introduces the system model, including the centralized/cloud wireless-optical access network (C-WAN)-based FTTR architecture, the RCA framework for performance degradation, and the channel access mechanism. Section 3 elaborates on the Transformer-based multi-task RCA algorithm, including problem modeling, model structure, and loss function design. Section 4 validates the effectiveness of the proposed method through simulation experiments, including dataset generation, baseline comparisons, and ablation studies. Future research directions are outlined in Section 5. Finally, Section 6 concludes the paper.

2 System Model

This section introduces the FTTR system model based on the C-WAN, including its architecture, the framework for RCA of performance degradation, and the channel access mechanism.

2.1 FTTR Network Architecture Based on C-WAN

As shown in Fig. 1, the C-WAN system is functionally divided into three planes: the management plane, the control plane, and the data plane. The management plane is responsible for monitoring device status, maintaining topology infor-

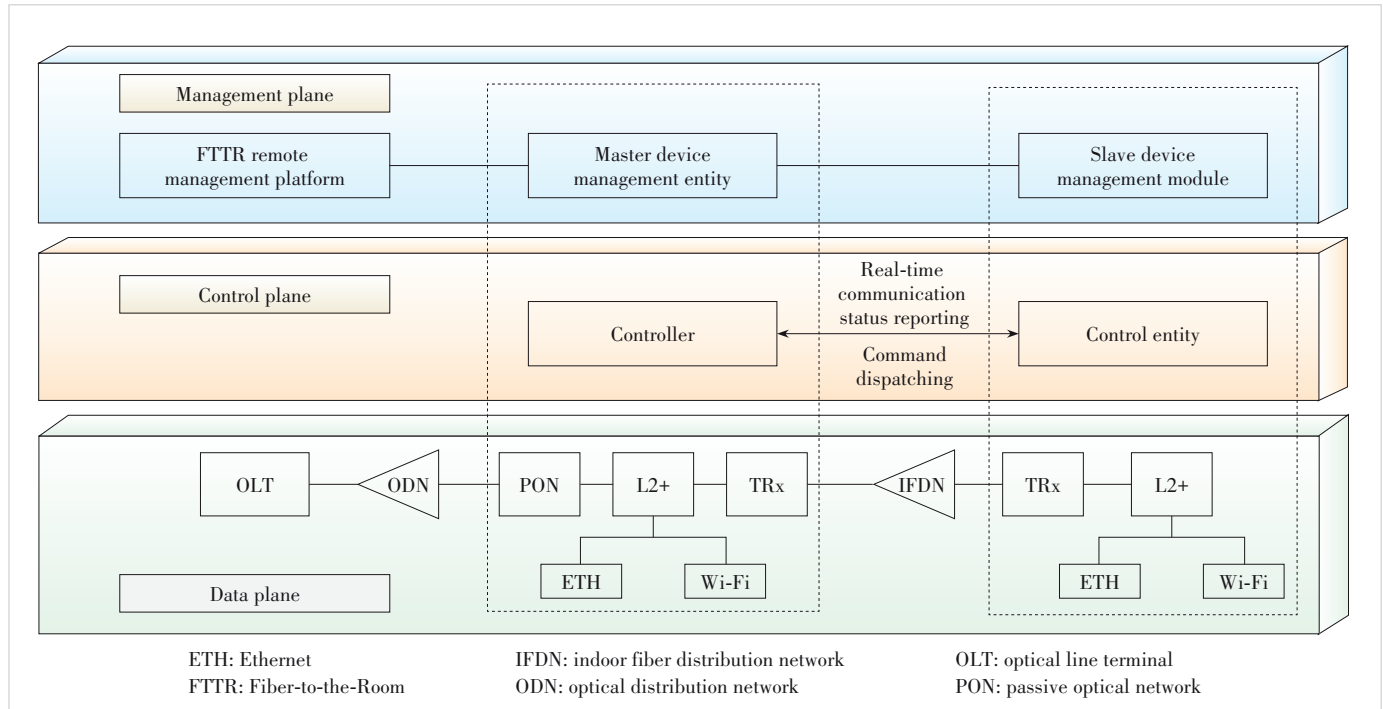


Figure 1. Centralized/cloud wireless-optical access network architecture

mation, and interfacing with operator platforms to achieve operational and maintenance objectives. The control plane primarily collects information from both air interfaces and optical links, enabling intelligent allocation and decision-making for optical-wireless resources through unified scheduling. The data plane consists of multiple devices responsible for the actual forwarding of user packets and provides transmission channels for operation and control signaling. Within this architecture, the OLT can gather various types of status information from all APs and possesses strong computational capabilities. This makes it well-suited for centralized analysis of network quality degradation causes. Detailed explanations will be provided in subsequent sections.

2.2 Framework for Root Cause Analysis of FTTR Performance Degradation

In FTTR networks where dense user and AP coexistence is common, the causes of performance degradation are complex and involve factors such as air interface collisions, channel contention, and uneven resource allocation. Therefore, this paper establishes an RCA framework for FTTR performance degradation based on the C-WAN architecture. This framework leverages real-time status data collected by the OLT, including channel utilization, collision counts, packet error rates, modulation and coding scheme (MCS), etc., and the low-latency optical fiber links to identify key factors contributing to network quality deterioration through feature extraction and correlation analysis. The analysis results can be used to guide resource re-configuration and transmission strategy adjustments, thereby forming a closed-loop operational mechanism of “detection-analysis-optimization”. This enhances network reliability and service quality.

The system model consists of a centralized OLT connected to multiple APs via optical fibers. Each AP, in turn, establishes wireless local area network (WLAN) connections with multiple stations (STAs). Leveraging low-latency optical links, the centralized OLT can promptly collect and aggregate multi-source status data generated by all APs and STAs across the network. This enables the construction of a multi-dimensional performance indicator system spanning the physical layer, link layer, and network layer. Furthermore, thanks to the edge computing module deployed at the OLT, AI models can be deployed to extract and integrate high-dimensional features. This capability allows for the identification of potential root causes of degradation, such as roaming issues, interference, collisions, congestion, and weak coverage. Based on the analysis results, optimization strategies can be formulated and delivered to the corresponding APs for execution via the C-WAN control mechanism. This architecture fully leverages the centralized advantages of C-WAN, significantly enhancing the service assurance capabilities of FTTR networks in high-density environments.

2.3 FTTR Network Channel Access Mechanism

In FTTR systems, the channel access mechanism comprises both PON and Wi-Fi components, aiming to achieve efficient and low-collision data transmission.

On the PON side, the data transmission mechanism includes downlink broadcast and uplink time division multiple access (TDMA). For downlink transmission, the OLT sends data via broadcasting, and each optical network unit (ONU) filters and receives data based on its logical identifier. For uplink transmission, multiple ONUs share the same fiber channel, and the OLT centrally schedules time slot allocations to prevent collisions. The OLT uses Gate control frames to assign transmission windows to ONUs, which then send data and Report frames within designated time slots to report their buffer status. The transmission time slots have a minimum granularity to ensure sufficient processing time for ONUs. For example, in EPON, the basic unit is 16 ns, and the minimum allocation unit is 1 024 such basic time slots (i.e., 16.384 μ s). Additionally, the system periodically executes discovery and registration procedures to maintain the active status of ONUs and prevent watchdog timer timeout. Control frames (such as Gate and Report) are assigned a higher transmission priority than data frames.

In Wi-Fi networks, the enhanced distributed channel access (EDCA) mechanism is employed to enable service differentiation and QoS management. This mechanism defines four access categories (ACs): AC_VO, AC_VI, AC_BE, and AC_BK. Each AC corresponds to specific contention parameters, including minimum contention window (CW_{min}), maximum contention window (CW_{max}), arbitration inter-frame space number (AIFSN), and transmission opportunity limit (TXOPLimit). The AIFS duration is calculated as follows:

$$\text{AIFS}[\text{AC}] = \text{SIFSTime} + \text{AIFSN}[\text{AC}] \times \text{SlotTime} \quad (1).$$

In the EDCA mechanism, SIFSTime and SlotTime are physical layer parameters. Each AC independently executes a backoff procedure. When the channel remains idle for the duration of the AIFS, a random backoff timer is triggered. If an internal collision occurs, the higher-priority AC gains transmission rights, while the lower-priority AC must double its backoff window. Once the transmission is successful, the contention window is reset to the CW_{min} value. If the transmission fails, the window continues to double until CW_{max} is reached.

3 Transformer-Based Multi-Task Root Cause Analysis Algorithm

3.1 Problem Formulation

The overall network architecture of FTTR comprises two major components: the PON and Wi-Fi access. In practical deployments, the PON section utilizes optical fiber as the trans-

mission medium, offering high bandwidth, low latency, strong anti-interference capabilities, and generally stable operation with a low probability of network quality degradation. In contrast, the Wi-Fi section operates in open wireless frequency bands, making it highly susceptible to environmental interference, diverse terminal behaviors, physical obstructions, and other factors. Consequently, Wi-Fi has become the primary source of overall network quality degradation. Relevant statistics indicate that over 80% of user experience issues originate from the Wi-Fi side. Building on this, this paper focuses on analyzing and modeling Wi-Fi-related network quality issues in FTTR systems.

The analysis of FTTR network quality degradation can be formulated as a multi-task time series classification problem. Consider a network system comprising n APs. At each time step t , the system monitors N_f feature metrics, forming a feature vector $\mathbf{x}_t \in \mathbb{R}^{N_f}$.

Given an observation time window of length L , the model input can be represented as a feature matrix:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L]^T \in \mathbb{R}^{L \times N_f} \quad (2),$$

where each feature vector $\mathbf{x}_t, t \in [1, \dots, L]$, includes three types of metrics: 1) AP-specific metrics: $\phi_{AP_i}^{(j)}$, where $i \in \{1, 2, \dots, n\}$ denotes the AP index, and j represents the metric type (e.g., number of STAs, channel status); 2) AP performance metrics: $\psi_{AP_i}^{(k)}$, reflecting the performance of each AP (e.g., throughput, interference level); 3) Global network metrics: $\omega^{(m)}$, describing the overall network state (e.g., total user count, average channel utilization).

The model outputs two vectors, Root Cause Localization (a root cause probability vector $\mathbf{Y}_{AP} \in [0, 1]^n$ indicating the probability that each AP is the root cause AP) and Root Cause Type (a root cause type probability distribution $\mathbf{Y}_{Type} \in [0, 1]^C$,

where C is the number of root cause types). The root cause types considered in this paper include Normal, Weak Coverage, Congestion, Collision, Roaming, and Interference.

The model needs to learn a mapping function F as expressed in Eq. (3):

$$F(\mathbf{X}; \Theta) = [\mathbf{Y}_{AP}, \mathbf{Y}_{Type}] \quad (3),$$

where Θ represents the parameters of the function, consisting of the Transformer-based multi-task learning model and its output heads.

3.2 Model Architecture: Transformer-Based Multi-Task Learning Model

The core architecture of the model is illustrated in Fig. 2. Each feature vector is mapped from dimension N_f to dimension N_d through a fully connected layer. Here, \mathbf{W}_p is the weight matrix, and \mathbf{b}_p is the bias vector.

$$\tilde{H}^{(0)} = \mathbf{X} \cdot \mathbf{W}_p + \mathbf{b}_p \quad (4),$$

$$\mathbf{H}^{(0)} = \tilde{H}^{(0)} + \mathbf{P} \quad (5).$$

Since the Transformer mechanism does not employ recurrent or convolutional structures, it cannot inherently capture the sequential relationships among elements in the sequence. The positional encoding \mathbf{P} , a tensor with the same shape as $\mathbf{H}^{(0)}$, is learned and added to $\mathbf{H}^{(0)}$ to enable the model to utilize the temporal order information of the time steps in the sequence.

The Transformer encoder consists of N_L identical layers stacked together, each containing a multi-head self-attention mechanism and a feed-forward neural network. The multi-head self-attention mechanism allows any time step in the sequence to directly attend to all other time steps, effectively

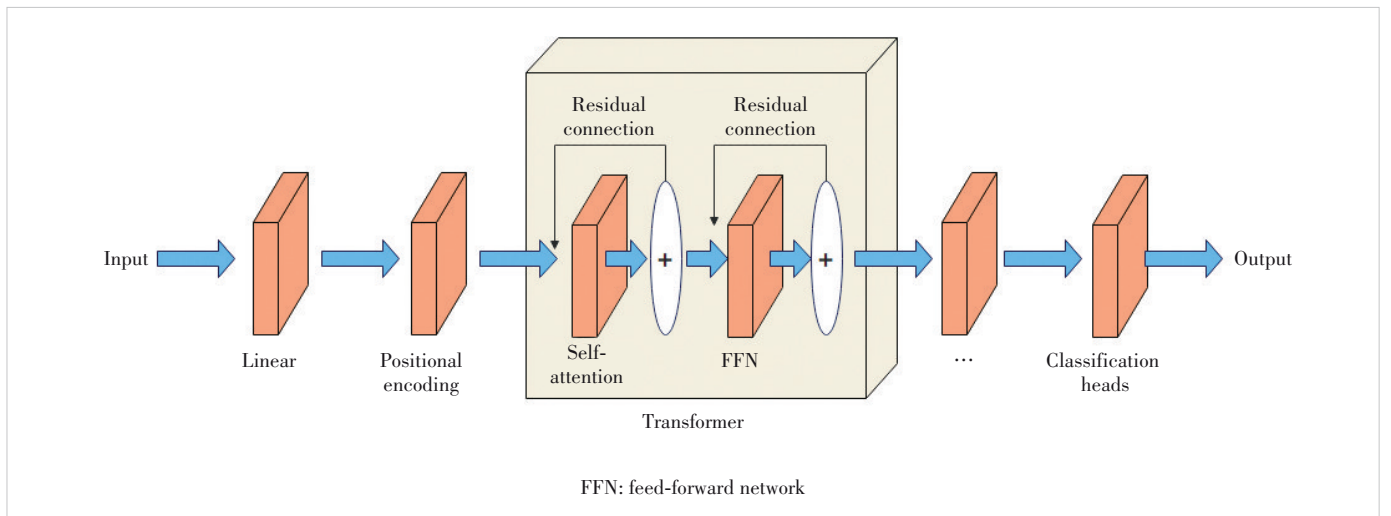


Figure 2. Multi-task Transformer mechanism

capturing long-term dependencies and interactions among features. This is crucial for analyzing the temporal causal relationships in network state metrics. The feed-forward neural network primarily applies non-linear transformations to the outputs of the self-attention layer, enhancing the model's expressive power. Each sub-layer incorporates residual connections and layer normalization, which help mitigate gradient vanishing issues in deep networks and accelerate training.

$$H^{(N_L)} = \text{TransformerEncoder}(H^{(0)}) \quad (6)$$

After processing through N_L layers of the Transformer encoder, we obtain a sequence $H^{(N_L)}$ rich in contextual information. The output of the last time step, h_{final} , is used as a comprehensive representation of the entire input sequence. It is fed into two separate classification heads to produce the model outputs Y_{AP} and Y_{Type} .

$$Y_{\text{AP}} = \text{softmax}(h_{\text{final}} W_{\text{AP}} + b_{\text{AP}}) \quad (7)$$

$$Y_{\text{Type}} = \text{softmax}(h_{\text{final}} W_{\text{Type}} + b_{\text{Type}}) \quad (8)$$

3.3 Loss Function

The total loss function is defined as the weighted sum of the two task-specific losses:

$$L_{\text{total}} = \alpha L_{\text{AP}} + \beta L_{\text{Type}} \quad (9)$$

where α and β are the weights for the two tasks. Both losses employ the cross-entropy function. The objective of model training is to minimize this total loss function.

3.4 Computational Cost Analysis

The computational cost of the model primarily originates from the initial projection layer, the Transformer encoder, and the output layer. The following analysis is based on computational complexity, where the sequence length is L , the model dimension is N_d , the input feature dimension is N_f , and the number of Transformer layers is N_L .

The operation of the initial projection layer involves matrix multiplication $X \cdot W_p$ and bias addition. Its complexity is $O(L \cdot N_f \cdot N_d)$. Since N_f and N_d are fixed dimensions, the complexity is linearly related to the sequence length L .

The complexity of each Transformer encoder layer is determined by the multi-head self-attention and the feedforward neural network.

1) Multi-head self-attention: The complexity of computing the attention matrix QK^T is $O(L^2 \cdot d_k)$, where $d_k = N_d/h$ (with h being the number of heads). Since there are h heads, the total complexity is $O(h \cdot L^2 \cdot d_k) = O(L^2 \cdot N_d)$, because $d_k = N_d/h$.

2) Feedforward neural network: The complexity of the two

linear transformations is $O(L \cdot N_d \cdot d_{ff})$. Since $d_{ff} \propto N_d$, the complexity is $O(L \cdot N_d^2)$.

Thus, the total complexity of each Transformer layer is $O(L^2 \cdot N_d + L \cdot N_d^2)$. The entire encoder has N_L layers, so the total complexity is $O(N_L \cdot L^2 \cdot N_d + N_L \cdot L \cdot N_d^2)$.

The operations of the output layer involve matrix multiplication and softmax for two classification heads. The complexity is $O(C_{\text{AP}} \cdot N_d + C_{\text{Type}} \cdot N_d)$. Since C_{AP} and C_{Type} are small and constant, this part of the complexity is negligible.

In conclusion, the total computational cost is:

$$O(L \cdot N_f \cdot N_d + N_L \cdot L^2 \cdot N_d + N_L \cdot L \cdot N_d^2 + C_{\text{AP}} \cdot N_d + C_{\text{Type}} \cdot N_d) \quad (10)$$

Based on the above analysis, the total computational cost of the model is primarily dominated by the Transformer encoder, i.e., $O(N_L \cdot L^2 \cdot N_d + N_L \cdot L \cdot N_d^2)$. The computational cost mainly depends on the sequence length L and the model dimension N_d .

4 Simulation

4.1 Dataset Generation

To train and evaluate an AI model capable of accurately determining the root causes of WLAN network issues, we developed a discrete-event simulation platform based on MATLAB. This platform simulates key mechanisms of the IEEE 802.11 protocol's MAC and physical layers to replicate various network quality degradation scenarios. The architecture of the simulator and the detailed process for generating a multi-label dataset are described below.

Our simulator follows an object-oriented design philosophy, with its core classes and interactions illustrated in Fig. 3.

The Simulator module is the core scheduling system of the simulation, responsible for maintaining the global clock and a list of APs. It advances the simulation by scanning and executing the earliest occurring events. The AP module models the behavior of real APs, maintaining a state machine that includes states such as BACK_OFF, HOLD, and SEND. The Callback module contains functions scheduled to execute at specific future time points, handling events such as the end of backoff, the start of frame transmission, and acknowledge (ACK) character timeout checks. The Events module records the details of each transmission attempt, including start time, end time, and status. The Mobile module simulates STA mobility and roaming behavior, tracking the positions of STAs and APs. It calculates the received signal strength based on real-time distance and a path loss model, triggering roaming decisions when the signal strength falls below a certain threshold. The ChannelQuality module calculates the received signal power, aggregates interference from co-channel APs and external sources, and computes the signal-to-interference-plus-noise ratio (SINR). It also determines the packet error

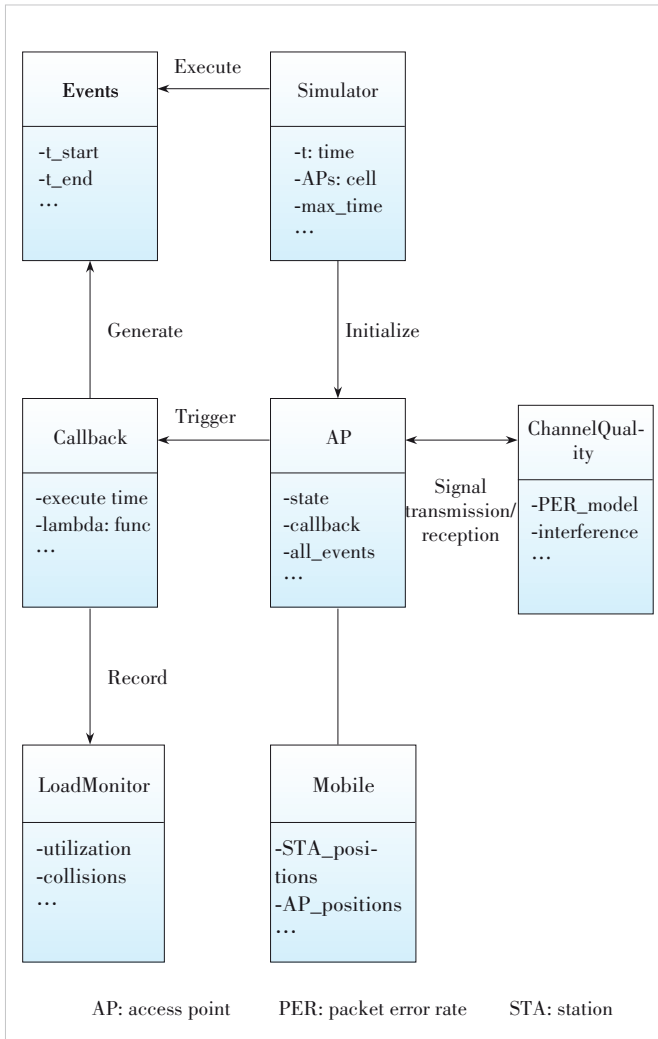


Figure 3. Simulator modules: core classes and interactions

rate (PER) using a PER-SINR mapping model, taking into account different modulation and coding scheme (MCS) levels. The LoadMonitor module monitors network load conditions, collecting statistics such as channel utilization, queue length, and collision count for each AP, and subsequently calculates the collision rate.

Data generation is achieved through multiple simulation runs, each configured with specific parameters to simulate particular problem scenarios. Roaming issues are simulated using STA mobility models, where STAs continuously move between two APs; Coverage deficiencies are simulated by setting low AP transmission power or placing STAs at the edge of coverage areas; Congestion issues are simulated by increasing the number of APs or raising the STA's transmission rate to elevate the load; Interference issues are simulated by adjusting AP topology to cause mutual interference or by introducing external APs outside the FTTR system; Severe collision scenarios are simulated by configuring small contention windows or a large number of STAs.

The dataset generation process integrates the previously mentioned modules. The first step is simulation initialization, where parameters such as the number of APs, number of STAs, physical locations, transmission power, and traffic load are configured. All AP, Mobile, ChannelQuality, and LoadMonitor modules are initialized.

During the simulation, the simulator employs an event-driven time-step advancement mechanism. At each sampling interval, it performs unified sampling of all APs to create time-stamped data samples. Each sample includes the following feature sets:

1) Network load features: the total number of STAs, the number of STAs currently associated with each AP, and system load imbalance degree (used to quantify load distribution differences among APs);

2) Channel state features: channel number occupied by each AP, the number of co-channel interfering APs, SINR for each AP, and RSSI for each AP;

3) Performance metrics: total downlink throughput per AP, maximum communication delay, frame retransmission rate, channel utilization, and the MAC-layer data frame collision rate;

4) Roaming behavior features: the number of roaming event triggers and that of failed roaming attempts per AP.

Each sample is automatically annotated with two labels. The root cause label identifies the primary network problem. The categories include roaming issues, coverage deficiency, congestion, interference, severe collisions, and normal conditions. The labeling logic is determined by the simulation configuration. The problem AP label identifies the device ID of the AP responsible for the root cause. For coverage deficiency and roaming issues, this label is the ID of the currently serving AP; for interference issues, it is the ID of the primary interfering AP; for congestion and severe collision scenarios, it is the ID of the overloaded AP. Finally, the feature vectors of all samples and their corresponding labels are stored in a matrix form to create the final dataset for model training.

In the subsequent experiment, using the aforementioned method, we conducted simulations for six different scenarios, each simulated 300 times with a time-step of 300 s. By changing the random number seeds, a dataset comprising 540 000 samples with 39 features was constructed. The dataset was partitioned into training, validation, and test sets in a 70%, 15%, and 15% ratio, respectively. Both the root cause label and problem AP label exhibit a uniform distribution.

4.2 Baseline Model Performance Comparison

The dataset generated using the aforementioned method was used to train the models. As shown in Fig. 4, as the number of training iterations increases, the loss gradually decreases, and the accuracy gradually improves, eventually stabilizing.

To comprehensively evaluate the performance of the proposed Transformer-based multi-task learning root cause analy-

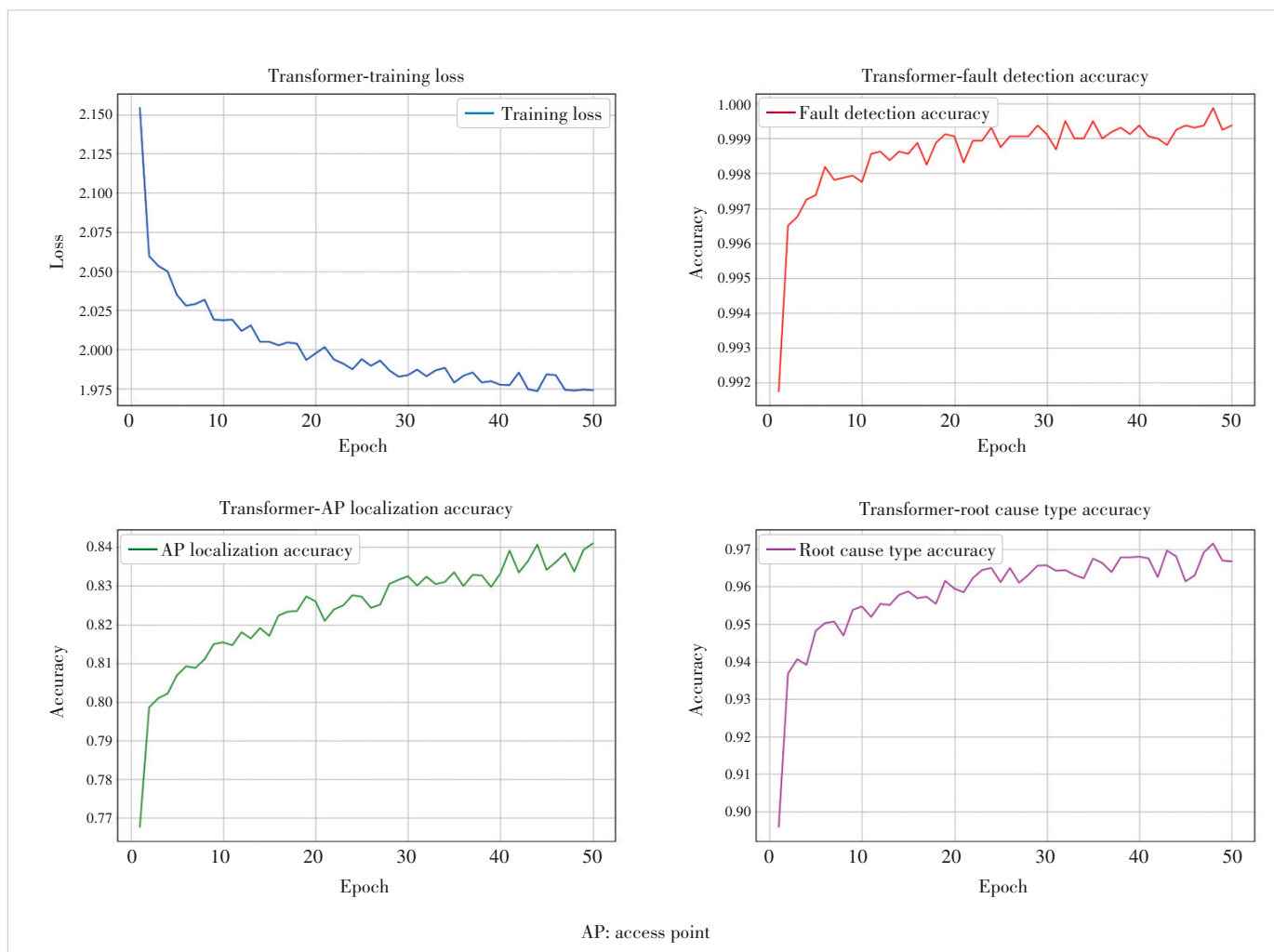


Figure 4. Baseline model performance comparison: training results

sis algorithm for FTTR networks, we compared it against several representative baseline models, including classical sequence models and traditional machine learning methods. The selected baseline models are two sequence models: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), and two traditional machine learning models: Random Forest and eXtreme Gradient Boosting (XGBoost).

As shown in Fig. 5, the proposed Transformer-based multi-task learning model achieved the highest accuracy in both AP localization and root cause type identification tasks, demonstrating its excellent capability in temporal feature extraction and complex pattern discrimination.

- AP localization task: The Transformer-based multi-task learning model achieved an accuracy of 80.92%, outperforming LSTM (79.54%), GRU (79.02%), Random Forest (74.64%), and XGBoost (76.29%). This indicates that the self-attention mechanism is superior to recurrent structures and tree models in capturing global dependencies. However, the overall accuracy rates were relatively low, primarily because

when network quality degrades, non-root-cause APs may exhibit similar features. For example, if AP1 is interfered with by AP2, AP2 may also experience interference from AP1, resulting in highly similar features between the two APs.

- Root cause type classification task: The Transformer-based multi-task learning model achieved an accuracy of 96.75%, surpassing all baseline models. Traditional machine learning methods performed significantly worse in this task, highlighting the advantage of deep sequence models in fine-grained, multi-category classification scenarios.

In summary, the Transformer-based multi-task learning model is well-suited for root cause analysis in FTTR network quality degradation scenarios.

4.3 Ablation Study on Transformer Mechanisms

To conduct an in-depth analysis of the contributions of key modules in the Transformer-based multi-task learning model, we performed an ablation study to systematically evaluate the individual effects of positional encoding and the multi-head attention

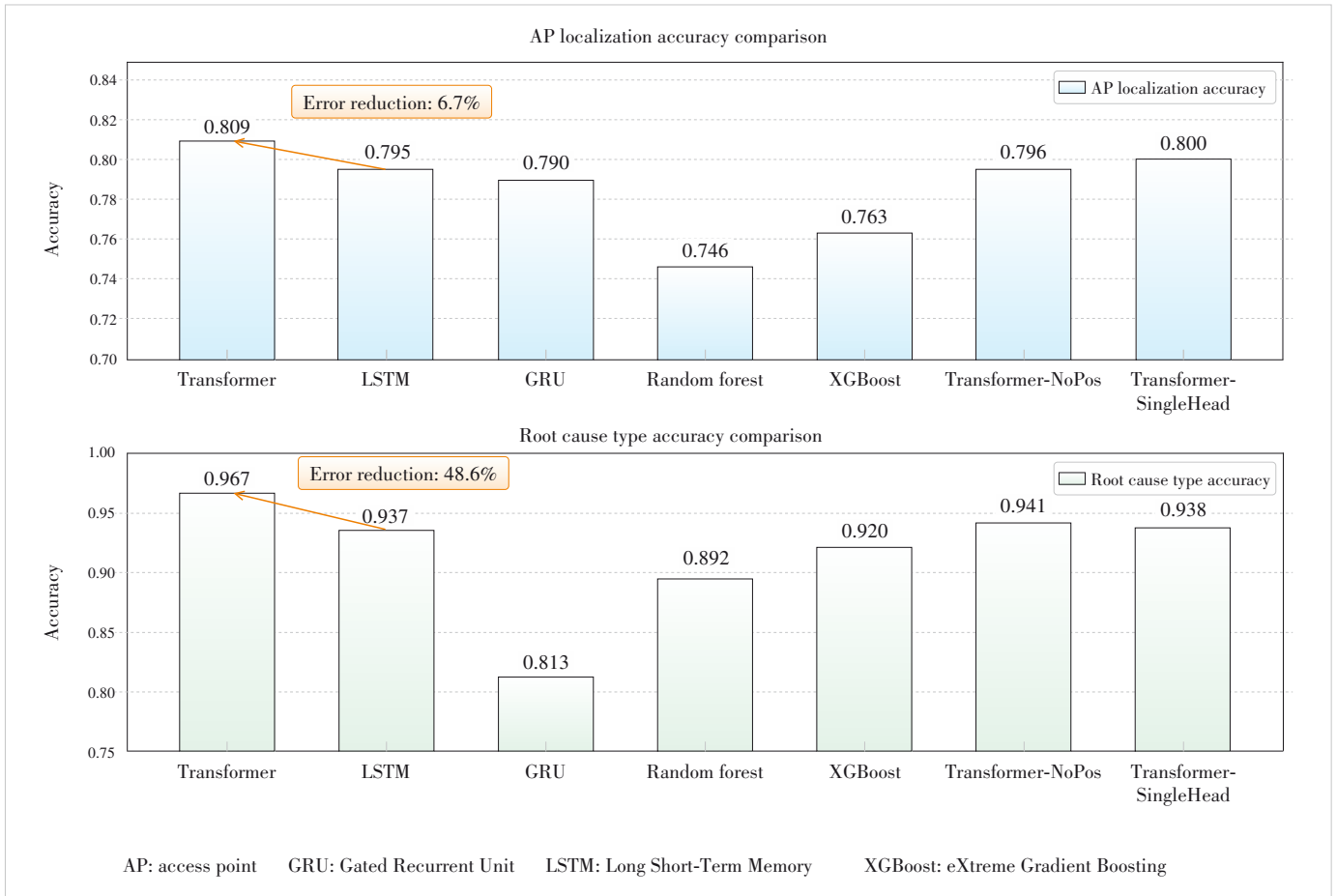


Figure 5. Accuracy comparison of AP localization and root cause type identification

mechanism. Specifically, we designed two variant models. 1) Transformer-NoPos: The positional encoding module was removed; 2) Transformer-SingleHead: The multi-head attention mechanism was replaced with a single-head attention mechanism.

Upon removing positional encoding, the AP localization accuracy dropped by 1.35%, and the root cause type classification accuracy decreased by 2.63%. These results indicate that encoding sequential information is essential for fault sequence interpretation. Without explicit positional cues, the model has difficulty distinguishing the temporal order of events. The relatively moderate performance degradation is mainly due to the use of simulation-generated data, where each scenario is pre-configured with a specific fault type. For instance, in the coverage deficiency scenario, the entire simulation consistently exhibits coverage-related issues, and the features exhibit minimal temporal variation.

Replacing multi-head attention with single-head attention led to performance declines in both tasks. The root cause classification accuracy decreased by 2.95%, confirming that the multi-head structure enhanced the model's expressive capacity and robustness by integrating diverse features from mul-

tiples representation subspaces.

Overall, the ablation study demonstrates that both positional encoding and the multi-head attention mechanism are critical components of the Transformer-based multi-task learning model, thus significantly contributing to its accuracy and generalization performance in root cause analysis tasks.

5 Future Work and Challenges

The current model faces two major challenges. First, it relies on simulated data with inherent limitations, resulting in insufficient generalization capability and credibility due to the absence of real data; second, it performs poorly in root-cause AP localization. Accordingly, we will address these two aspects separately by elaborating on the current issues and proposing future improvement plans.

Currently, our experiments rely on simulated data. There are significant differences between real data and simulated data, among which we consider the following to be particularly important: 1) The simulated data does not emulate real queues, thus lacking scenarios of overflow-induced packet loss, and the measured latency data does not include waiting delays in queues; 2) The simulation of actual service flows is

inadequate, as our packet transmission process is relatively stable and lacks simulation of burst traffic; 3) Transmission control protocol (TCP) mechanisms (such as sliding windows, congestion control) are difficult to simulate. Therefore, we plan to use real data in the future to enhance the credibility of our model.

Through collaboration with the industry, we have currently set up an FTTR test environment. In the future, we will first integrate a timer within the AP to periodically measure the parameters and metrics we need, and transmit the data to a PC via Ethernet (without interfering with Wi-Fi data transmission). Subsequently, we will attempt to label the data based on real environmental data, which will be a challenging yet critical task. We cannot simply label data based on specific metrics (for example, marking it as a “coverage issue” when the RSSI falls below a certain threshold), as this would cause the model to learn our labeling methodology, which is not the desired outcome. Therefore, the rationality of the labeling approach will directly impact the final accuracy and generalization performance of our model. Finally, we will retrain and fine-tune our model using real environmental data to enhance its practical utility.

Regarding root-cause AP localization, our analysis reveals that localization performance is the poorest in congestion scenarios. Further data investigation indicates that this is due to limitations inherent in the simulated data itself. Specifically, the simulation fails to accurately model TCP-related mechanisms and does not account for queuing delays caused by Wi-Fi queues. As a result, the round trip time (RTT), which should ideally reflect feature differences between root-cause and non-root-cause APs in congestion scenarios, fails to exhibit discriminative characteristics. In addition to utilizing real-world data, we will explore the following approaches in future work to further enhance model performance in root-cause AP localization tasks: 1) incorporating device-specific attention mechanisms, 2) introducing additional highly discriminative features, and 3) experimenting with models trained on single-AP data.

6 Conclusions

This paper addresses the challenges of identifying the source of quality degradation and analyzing the complex root causes in FTTR networks by proposing a multi-task root cause analysis model based on the Transformer mechanisms. The model is capable of simultaneously detecting faulty APs and determining the types of root causes. Experimental results show that the model achieves strong performance on a simulated dataset, with a root cause classification accuracy of 96.75%, surpassing several traditional machine learning and deep learning baselines. Ablation studies further confirm the critical role of positional encoding and the multi-head attention mechanism in enhancing model performance. However, during network quality degradation, non-root-cause APs may

display similar symptoms, leading to relatively lower accuracy in fault localization. Additionally, the current study is based on simulated data, where temporal feature variations may not fully capture real-world dynamics. Future work will focus on validating the model’s generalization using real-world network data, exploring online learning mechanisms, improving interpretability, and integrating technologies such as knowledge graphs to enhance the transparency and reliability of the reasoning process. This study offers an effective technical solution for intelligent operation and maintenance in FTTR networks, contributing to improved network reliability and user satisfaction.

References

- [1] Broadband Development Alliance. Fiber-to-the-Room (FTTR) white paper [R]. Beijing, China, 2022
- [2] PFEIFFER T, DOM P, BIDKAR S, et al. PON going beyond FTTH (invited tutorial) [J]. *Journal of optical communications and networking*, 2022, 14(1): A31 – A40. DOI: 10.1364/JOCN.439241
- [3] ETSI GR F5G 001 V1.1.1. Fifth Generation Fixed Network (F5G); F5G Generation Definition [R]. Sophia Antipolis, France: ETSI, 2020
- [4] LIU J Y, SHI D L, LI G N, et al. Data-driven and association rule mining-based fault diagnosis and action mechanism analysis for building chillers [J]. *Energy and buildings*, 2020, 216: 109957. DOI: 10.1016/j.enbuild.2020.109957
- [5] CHEN W C, TSENG S S, WANG C Y. A novel manufacturing defect detection method using association rule mining techniques [J]. *Expert systems with applications*, 2005, 29(4): 807 – 815. DOI: 10.1016/j.eswa.2005.06.004
- [6] SIM H, CHOI D, KIM C O. A data mining approach to the causal analysis of product faults in multi-stage PCB manufacturing [J]. *International journal of precision engineering and manufacturing*, 2014, 15(8): 1563 – 1573. DOI: 10.1007/s12541-014-0505-8
- [7] MA Q P, LI H Y, THORSTENSON A. A big data-driven root cause analysis system: application of machine learning in quality problem solving [J]. *Computers & industrial engineering*, 2021, 160: 107580. DOI: 10.1016/j.cie.2021.107580
- [8] JAVANBAKHT N, NESHASTEGARAN A, IZADI I. Alarm-based root cause analysis in industrial processes using deep learning [EB/OL]. (2022-03-21) [2025-09-16]. <https://arxiv.org/abs/2203.11321>
- [9] LIU P, XU H W, OUYANG Q Y, et al. Unsupervised detection of microservice trace anomalies through service-level deep Bayesian networks [C]// *Proceedings of IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2020: 48 – 58. DOI: 10.1109/ISSRE5003.2020.00014
- [10] MAMPAKA M M, SUMBWANYAMBE M. Poor data throughput root cause analysis in mobile networks using deep neural network [C]// *The 2nd Wireless Africa Conference (WAC)*. IEEE, 2019: 1 – 6. DOI: 10.1109/AFRICA.2019.8843409
- [11] ZHAO M H, ZHONG S S, FU X Y, et al. Deep residual shrinkage networks for fault diagnosis [J]. *IEEE transactions on industrial informatics*, 2020, 16(7): 4681 – 4690. DOI: 10.1109/TII.2019.2943898
- [12] CHOI M, KIM T, LEE J P, et al. An empirical study on root cause analysis and prediction of network failure using deep learning [C]// *Proceedings of International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2021: 741 – 746. DOI: 10.1109/ICTC52510.2021.9621097
- [13] ROY D, ZHANG X C, BHAVE R, et al. Exploring LLM-based agents for

root cause analysis [C]//The 32nd ACM International Conference on the Foundations of Software Engineering. ACM, 2024: 208 – 219. DOI: 10.1145/3663529.3663841

- [14] WANG Z F, LIU Z C, ZHANG Y Y, et al. RCAgent: cloud root cause analysis by autonomous agents with tool-augmented large language models [C]//The 33rd ACM International Conference on Information and Knowledge Management. ACM, 2024: 4966 – 4974. DOI: 10.1145/3627673.3680016

Biographies

YU Weichao received his BS degree in communication engineering from the School of Electronic Information and Communications, Huazhong University of Science and Technology, China in 2024, where he is currently pursuing his MS degree. His research interests include FTTR, Wi-Fi, and artificial intelligence.

LIU Yang received his BS degree in electronics science and technology from Huazhong University of Science and Technology, China in 2018, where he is currently pursuing his PhD degree. His research interests include Wi-Fi 7 and FTTR technologies.

ZHANG Junxiong received his ME degree in electronic information engineer-

ing from Huazhong University of Science and Technology, China in 2021. He is currently pursuing his PhD degree at the same institution. His research interests include Det-Wi-Fi and FTTR technologies.

YE Junliang received his BS degree in communication engineering from China University of Geosciences in 2011, and PhD degree from Huazhong University of Science and Technology, China in 2018. His research interests include heterogeneous networks, stochastic geometry, mobility-based access models of cellular networks, millimeter wave communications, and next-generation wireless communications.

GE Xiaohu (xhge@mail.hust.edu.cn) received his PhD degree in communication and information engineering from Huazhong University of Science and Technology (HUST), China in 2003. He was a researcher at Ajou University, Republic of Korea, and Politecnico di Torino, Italy from 2004 to 2005. He has been with HUST since 2005 and is currently a full professor at the School of Electronic Information and Communications, HUST. He is also an adjunct professor with the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. His research interests include mobile communications, traffic modeling in wireless networks, green communications, and interference modeling in wireless communications. He was the recipient of the Best Paper Award at IEEE Globecom 2010. Prof. GE is the Chinese representative for the International Federation for Information Processing (IFIP). He serves as an associate editor for *IEEE Wireless Communications*, *IEEE Transactions on Vehicular Technology*, and *IEEE Access*.

QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer for FTTR Networks



CAI Jinhan, ZAN Mingyuan, SHEN Gangxiang

(School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China)

DOI: 10.12142/ZTECOM.202504007

<https://kns.cnki.net/kcms/detail/34.1294.tn.20251126.1148.002.html>,
published online November 26, 2025

Manuscript received: 2025-09-17

Abstract: As Fiber-to-the-Room (FTTR) networks proliferate, multi-device deployments pose significant energy consumption challenges. This paper proposes a Quality of Service (QoS)-aware energy-saving scheme based on a multi-threshold buffer energy saving (MBES) scheme to reduce consumption while ensuring energy QoS. MBES leverages the centralized control of the main fiber unit (MFU) and the wireless-state awareness of subordinate fiber units (SFUs) for synergistic fiber-wireless energy savings. The scheme assigns independent, dynamic buffer thresholds to service queues on SFUs, enabling low-latency reporting for high-priority traffic while accumulating low-priority data to extend sleep cycles. At the MFU, a coordinated scheduling algorithm accounts for Wi-Fi access delay and creates an adaptive closed-loop control by adjusting SFUs' buffer thresholds based on end-to-end delay feedback. Simulation results show that, while satisfying strict latency requirements, MBES achieves a maximum energy saving of 17.75% compared with the no energy saving (NES) scheme and provides a superior trade-off between latency control and energy efficiency compared with the single-threshold buffer energy saving (SBES) scheme.

Keywords: Fiber-to-the-Room; power consumption; energy consumption; energy saving

Citation (Format 1): CAI J H, ZAN M Y, SHEN G X. QoS-aware energy saving based on multi-threshold dynamic buffer for FTTR networks [J]. *ZTE Communications*, 2025, 23(4): 48 – 64. DOI: 10.12142/ZTECOM.202504007

Citation (Format 2): J. H. Cai, M. Y. Zan, G. X. Shen, "QoS-aware energy saving based on multi-threshold dynamic buffer for FTTR networks," *ZTE Communications*, vol. 23, no. 4, pp. 48 – 64, Dec. 2025. doi: 10.12142/ZTECOM.202504007.

1 Introduction

The rapid development of emerging services such as ultra-high-definition video, immersive extended reality (XR), smart education, and cloud-edge collaboration has presented unprecedented challenges to indoor access networks in terms of bandwidth, latency, and connection stability^[1]. To meet these demands, a next-generation indoor fiber access architecture, Fiber-to-the-Room (FTTR), has emerged^[2]. Through the centralized control by a main fiber unit (MFU) and the distributed coverage by multiple subordinate fiber units (SFUs), FTTR integrates the functions of a passive optical network (PON) optical network unit (ONU) with Wi-Fi access points (APs), achieving direct fiber connections and high-performance Wi-Fi coverage across the whole home^[3]. This architecture effectively overcomes the signal attenuation, insufficient coverage, and severe interference issues of traditional single-point Wi-Fi in multi-room, multi-device scenarios, pro-

viding a transmission guarantee with high bandwidth, low latency, and high reliability for emerging services^[4]. However, while FTTR enhances network performance, the significant increase in access devices leads to a sharp rise in system energy consumption. In common low-load or no-traffic scenarios, devices continue to power most of their modules, causing significant energy waste. This issue has become a key bottleneck hindering the green and large-scale deployment of FTTR^[5].

Existing research on energy savings in access networks primarily focuses on two separate domains: PONs and Wi-Fi. While research specifically targeting the FTTR architecture remains scarce, findings in these related areas offer important references for its energy-saving design. In the context of PONs, successive standards from Gigabit-capable PON (GPON) and 10-Gigabit-capable PON (XG-PON) to 50-Gigabit-capable PON (50G-PON) have introduced sleep mechanisms for ONUs, such as the doze mode, the cyclic sleep mode, and the watchful sleep mode (WSM), among which WSM, in particular, has become a key reference for standard energy-saving solutions in FTTR^[6–9]. Significant research has focused on optimizing ONU sleep. For instance, YAN et al.^[10] proposed uplink-centric and

This work was partially supported by the National Nature Science Foundation of China under Grant No. W2411058.

downlink-centric scheduling to achieve a trade-off between latency and energy savings. SHI et al.^[11] designed a differentiated ONU sleep-scheduling scheme based on service-level agreements to meet multi-level Quality of Service (QoS) demands. ZHANG et al.^[12] introduced a modular energy-saving method that reduces power consumption by dynamically deactivating internal ONU transceiver modules. Other works have explored aspects such as adaptive sleep-mode optimization^[13], the integration of sleep mechanisms with dynamic bandwidth allocation^[14], power-consumption modeling for various PON technologies^[15–16], and energy-efficiency analysis for multi-level ONU power states with intelligent sleep-duration adjustments^[17–18]. These studies provide valuable insights into the trade-off between energy consumption and QoS in PON systems. In the domain of Wi-Fi networks, energy-saving mechanisms primarily concentrate on stations (STAs). The Wi-Fi 6 standard introduced the Target Wake Time (TWT) mechanism, which allows STAs to negotiate with APs to enter the sleep mode during non-essential periods^[19]. LORINCZ et al.^[20] and PACK et al.^[21] proposed STA sleep-control methods based on load and adaptive intervals, effectively improving the balance between energy consumption and performance. In contrast, research on the energy consumption of APs remains relatively limited. However, existing works have explored aspects such as component-level power consumption^[22], power characteristics under different configurations^[23], and traffic-based power-prediction modeling^[24], providing support for the design and management of energy-saving wireless networks.

In converged fiber-wireless networks, studies have shown that a lack of coordination between the independent energy-saving mechanisms of PONs and Wi-Fi can lead to degraded overall performance^[25]. To address this, researchers have proposed various joint energy-saving strategies, including threshold-based ONU sleep with wireless traffic aggregation^[26], dynamic scheduling combined with wireless rerouting^[27], and coordinated optimization of Wi-Fi access and PON sleep modes^[28]. These methods reduce energy consumption while, to a certain extent, guaranteeing latency and throughput. Despite advances in independent and partially joint energy-saving optimizations for PONs and Wi-Fi, these approaches face notable limitations in FTTR scenarios. First, many schemes rely on a single load threshold, which makes it difficult to differentiate services with diverse latency requirements (e.g., voice versus video). Second, they do not fully exploit the MFU's centralized control capabilities or its advantage in monitoring the Wi-Fi-side status.

To fill this gap, this paper proposes an energy-saving scheme based on a multi-threshold buffer energy saving (MBES) scheme for FTTR networks, which achieves a balance between energy consumption and service performance through differentiated caching at the SFU side and centralized scheduling at the MFU side. The main contributions of this paper are as follows. First, we propose an FTTR uplink energy-saving framework

that reduces system power consumption while meeting diverse QoS requirements. Second, we design an SFU-side multi-threshold caching mechanism that differentiates services by priority to balance latency and energy saving. Third, we develop an MFU-side centralized scheduling algorithm that adaptively adjusts buffer thresholds based on Wi-Fi delay feedback, enabling QoS-aware closed-loop energy control.

The remainder of this paper is organized as follows. Section 2 introduces the FTTR network architecture and the uplink QoS assurance mechanism. Section 3 establishes the power consumption models for the SFU and MFU, along with the system energy consumption model. Section 4 elaborates on the proposed MBES strategy, detailing the SFU's multi-threshold wake-up mechanism and the MFU's cooperative scheduling algorithm. Section 5 evaluates and analyzes the performance of the strategy through simulations. Conclusions are drawn in Section 6.

2 FTTR Architecture and Uplink QoS Mechanisms

This section introduces the FTTR system architecture, focusing on its network structure, uplink data transmission process, and QoS assurance mechanisms. This provides the foundation for the design of the energy-saving strategy presented later in this paper.

2.1 FTTR Network Architecture

Fig. 1 illustrates the basic architecture of an FTTR network. The network consists of an MFU, an indoor fiber distribution network (IFDN), and one or more SFUs. Serving as the gateway and control center of the home network, the MFU connects to the external optical line terminal (OLT) through the PON protocol. Meanwhile, it manages SFUs deployed in different rooms via the IFDN using indoor fiber protocols such as G.647^[9]. At the data flow level, the MFU receives downstream data from the OLT and distributes it to the SFUs. For upstream data from SFUs, the MFU either forwards the traffic to the target SFU or transmits it to the OLT according to the destination address. Compared with traditional optical network terminals (ONTs), the MFU can not only receive broadcast information from the OLT but also provide centralized control of the entire FTTR network, including bandwidth allocation, roaming management, and energy-saving control for SFUs. As the terminal devices of the network, SFUs integrate the functionalities of ONUs and wireless APs. Each SFU is equipped with a G.647 module and Wi-Fi/Ethernet modules, enabling the conversion between optical and wireless frames. Specifically, the SFU encapsulates uplink data from user terminals into G.647 frames and transmits them to the MFU, and decapsulating downstream G.647 data from the MFU and delivering it to STAs via wireless links.

2.2 Uplink Transmission and Associated QoS Assurance in FTTR

In the FTTR architecture, all uplink data transmissions are

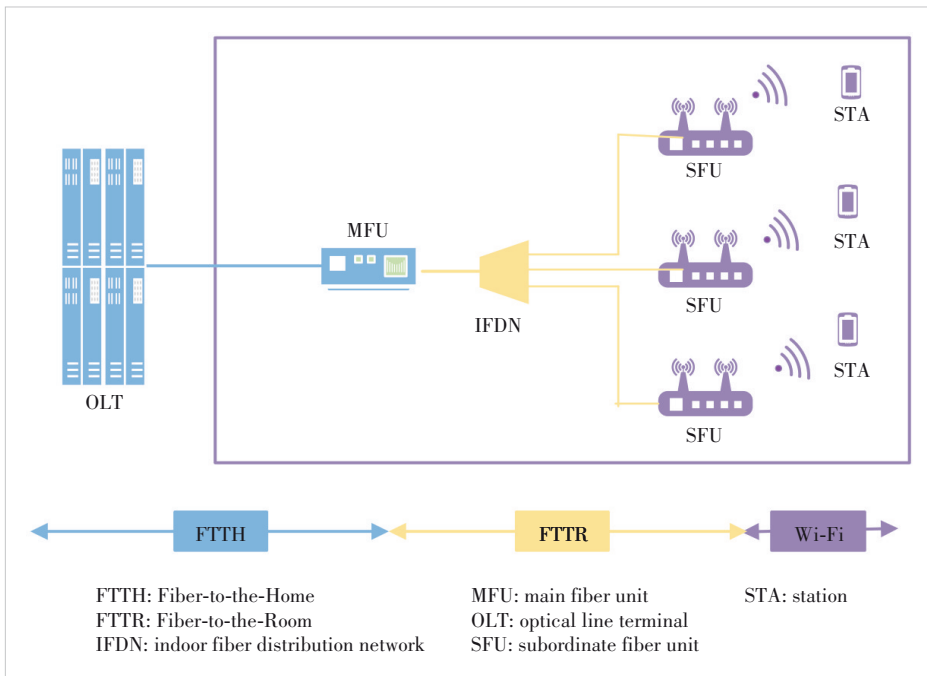


Figure 1. FTTR network architecture

centrally controlled by the MFU. When service data from an STA reaches its associated SFU through a Wi-Fi link, the SFU provides differentiated QoS by classifying packets into transmission container (T-CONT) queues according to service priorities. The four Wi-Fi access categories (ACs)—voice (VO), video (VI), best effort (BE), and background (BK)—are mapped to four T-CONT queues (T-CONT 1 – 4) with different scheduling attributes. Among them, T-CONT 1 is allocated fixed bandwidth to serve the highest-priority services; T-CONT 2 is assigned assured bandwidth, while T-CONT 3 and T-CONT 4 are used to carry low-priority services.

After packets enter the T-CONT queues, their uplink transmission is fully scheduled by the MFU through a dynamic bandwidth allocation (DBA) mechanism, in which the status-report-based DBA (SR-DBA) process is commonly adopted. This scheduling operates in a periodic cycle, as illustrated in Fig. 2. Specifically, each cycle begins with the MFU broadcasting a bandwidth map (BWmap), which specifies the exact timeslots and lengths for uplink transmissions of the T-CONT queues at each SFU. Upon receiving authorization, an SFU transmits data and appends dynamic bandwidth report upstream (DBRu) information to the data frame to indicate the real-time queue status. Based on the collected DBRu information from all SFUs, the MFU executes its DBA algorithm to generate a new BWmap allocation for the next cycle, thereby completing one scheduling loop and initiating the next cycle.

Although the SR-DBA mechanism can effectively guarantee bandwidth utilization and QoS, its inherent energy efficiency issues have also become evident. In common home network scenarios characterized by significant spatiotemporal traffic imbal-

ance, many SFUs may carry only a small amount of bursty traffic or even remain idle for long periods. However, under the current SR-DBA process, an SFU must frequently wake up its power-hungry optical transmitter to send DBRu reports, even when its queues are empty, so as to maintain its presence in the scheduling system. This communication leads to unnecessary power consumption and considerable energy waste, which constitutes the key problem addressed in this paper.

3 FTTR Device Power and System Energy Consumption Models

This section establishes the power consumption models of the SFU and MFU, as well as the overall system energy model, providing a quantitative foundation for the energy-saving strategy discussed in Section 4.

discussed in Section 4.

3.1 SFU's Power Consumption Model

The data transmission and reception of an SFU are influenced by the operating states of both its optical and wireless modules. Treating the SFU as a single entity for energy control makes it difficult to balance energy efficiency and communication latency. Therefore, this paper adopts a modular approach to model SFU power consumption. As shown in Fig. 3, the overall power consumption of an SFU is decomposed into three core components: the base module, the optical module, and the Wi-Fi module. The base module encompasses critical subsystems, including the processor, buffer, and power supply, all of which

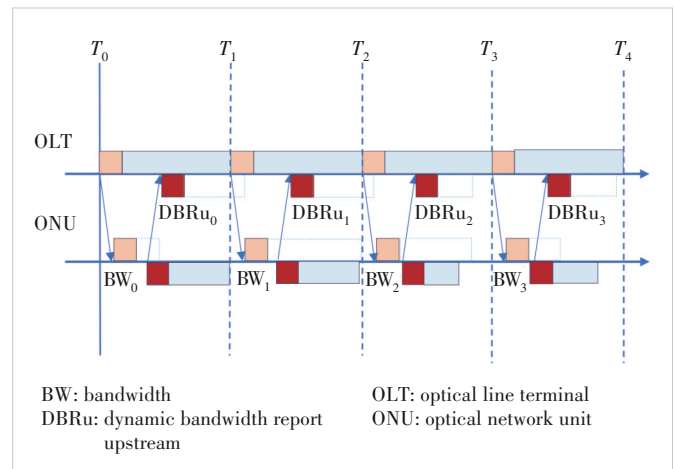


Figure 2. Status-report-based dynamic bandwidth allocation process

are indispensable for sustaining the device's core functionality. The optical module consists of the optical transmitter, receiver, and control circuits, which handles data transmission over the fiber link. The Wi-Fi module comprises the Wi-Fi control unit and radio frequency (RF) units, with the control unit responsible for multiple-input multiple-output (MIMO) operations and transmission power adjustment. Based on the modular division, the total power consumption of an SFU can be expressed as:

$$P_{\text{SFU}} = P_{\text{base}} + P_{\text{PON}} + P_{\text{WiFi}} \quad (1),$$

where P_{base} is the power consumption of the base module, P_{PON} is the power consumption of the optical module dominated by the transmitter and receiver, and P_{WiFi} is that of the Wi-Fi module, primarily determined by the number of RF units as well as their transmission power and maximum data rate.

The base module of the SFU is configured to remain continuously active in energy-saving design. Keeping this module active allows its internal clock synchronization circuits to respond promptly to wake-up requests, thereby minimizing synchronization latency with the MFU, which is essential for maintaining QoS. In addition, the buffer units inside the base module must operate continuously to process uplink data, preventing packet loss and retransmission that would occur if the module were turned off. More importantly, the power consumption of the base module is much lower than that of the high-power optical and Wi-Fi modules. Thus, considering it as a primary energy-saving target is suboptimal, given the marginal potential gains.

The power consumption of the optical module is mainly determined by its transmitter and receiver, which can be expressed as:

$$P_{\text{PON}} = v_{\text{T}} \cdot P_{\text{Tx}} + v_{\text{R}} \cdot P_{\text{Rx}} \quad (2),$$

where P_{Tx} and P_{Rx} are the power consumptions of the transmitter and receiver, respectively, and v_{T} and v_{R} are binary variables (0 or 1) indicating the on/off states of the transmitter and receiver. In FTTR systems, optical-side energy saving is mainly achieved by turning on/off the transceiver. When no uplink traffic is present for a long period, the SFU turns off the transmitter and periodically activates the receiver to detect MFU downlink data or wake-up instructions, thereby reducing power consumption.

The power consumption of the Wi-Fi module consists of the RF module and dynamic power consumptions, shown as:

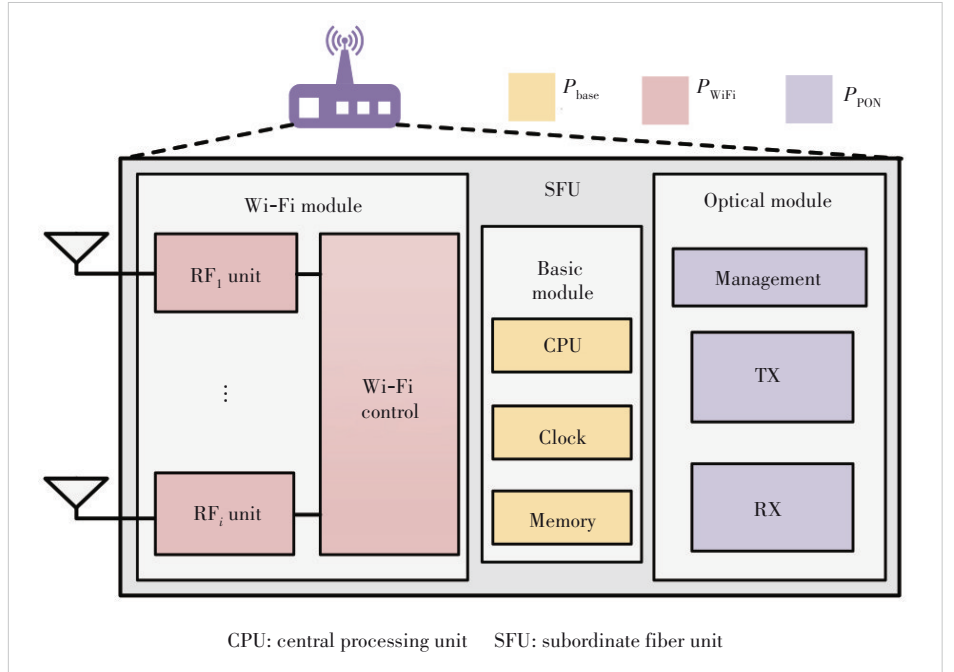


Figure 3. Functional modules and power consumption breakdown of the SFU

$$P_{\text{WiFi}} = P_{\text{RF}} + P_{\text{D}} \quad (3),$$

where P_{RF} is the RF module power consumption and P_{D} is the dynamic power consumption. Furthermore, P_{RF} is determined by the base power consumption and transmission power, given by:

$$P_{\text{RF}} = P_{\text{RF_Base}} + P_{\text{RF_Tx}} \quad (4),$$

where $P_{\text{RF_Base}}$ is the base power consumption of the RF module and $P_{\text{RF_Tx}}$ is the transmission power of the RF unit. $P_{\text{RF_Tx}}$ is an adjustable parameter that determines the coverage of Wi-Fi signals. In practice, the received signal strength indicator (RSSI) at the STA side is commonly used as the evaluation metric. When the RSSI exceeds a predefined threshold, the SFU is deemed capable of maintaining stable communication with the STA at the current transmission power.

Following the RF link power consumption model used in Ref. [5], the transmission power of the i -th RF unit, $P_{\text{RF_Tx}_i}$ (dBm), first undergoes antenna path loss σ (dB), then combines with antenna gain G_A (dBi), and subsequently passes through free-space path loss L_p (dB) and obstacle loss L_b (dB) before finally reaching the STA receiver. The received RSSI can be expressed as:

$$\text{RSSI} = P_{\text{RF_Tx}_i} - \sigma + G_A - L_p - L_b \quad (5).$$

The free-space path loss L_p is calculated as:

$$L_p = 32.4 + 20\log_{10} D + 20\log_{10} M \quad (6),$$

where D is the distance between the SFU and STA (km), and M

is the frequency (MHz). The obstacle loss L_b is given by:

$$L_b = \sum_{k=1}^K (j_k \cdot W_k) \quad (7)$$

where W_k (dB) denotes the attenuation caused by the k -th type of obstacles, and j_k is the number of such obstacles. Based on the above model, by setting the RSSI in Eq. (5) as the target threshold, the minimum transmit power required to ensure communication quality can be derived as:

$$P_{\text{RF_Tx}} = \text{RSSI} + \sigma - G_A + (32.4 + 20 \log_{10} D + 20 \log_{10} M) + \sum_{k=1}^K (j_k \cdot W_k) \quad (8)$$

Considering that both the SFU and STA may incorporate multiple RF units, the overall RF transmit power $P_{\text{RF_Tx}}$ is obtained by summing the transmit power of each unit:

$$P_{\text{RF_Tx}} = \sum_{i=1}^{N_{\text{SS}}} P_{\text{RF_Tx}i} \quad (9)$$

where N_{SS} is the number of RF units. Similarly, the total base power consumption of the RF module, $P_{\text{RF_Base}}$, is calculated as the sum of the base power of all RF units. In practice, the transmit power is also constrained by the effective isotropic radiated power (EIRP). The limit power P_L is expressed as:

$$P_L \geq P_{\text{RF_Tx}} - \sigma + G_A \quad (10)$$

Considering the EIRP constraint, the final transmit power $P_{\text{RF_Tx}i}$ is given by

$$P_{\text{RF_Tx}} = \begin{cases} \sum_{i=1}^{N_{\text{SS}}} P_{\text{RF}i}, & \sum_{i=1}^{N_{\text{SS}}} P_{\text{RF_Tx}i} \leq (P_L + \sigma - G_A) \\ P_L + \sigma - G_A, & \sum_{i=1}^{N_{\text{SS}}} P_{\text{RF_Tx}i} > (P_L + \sigma - G_A) \end{cases} \quad (11)$$

Thus, by substituting the EIRP-constrained transmit power of Eq. (11) into Eq. (5), the actual received signal strength at the m -th STA, RSSI_m , can be obtained.

The dynamic power consumption of the Wi-Fi module in the SFU, P_D , mainly arises from protocol processing at the baseband and media access control (MAC) layers, and is positively correlated with the actual data transmission rate. A higher data rate increases the processing load for modulation/demodulation and coding/decoding on the chip, thereby leading to a higher power consumption. For a single STA, its actual transmission rate R_m is constrained by the maximum negotiated link rate R_m^{max} with the SFU and varies dynamically with the current RSSI_m . This nonlinear relationship between rates and RSSI can be approximated by a sigmoid function^[29].

$$R_m = \frac{R_m^{\text{max}}}{1 + e^{\frac{-(120 + \text{RSSI}_m) - c_1}{c_2}}} \quad (12)$$

where parameters c_1 and c_2 are set according to experimental data in Ref. [29] to model the growth characteristic of throughput as signal strength varies. Accordingly, the total SFU transmission rate R_{total} at a given time is the sum of the rates of its connected STAs.

$$R_{\text{total}} = \sum_{m=1}^M R_m \quad (13)$$

The dynamic power consumption of the SFU is then calculated as

$$P_D = P_{\text{max}} \cdot \alpha_D \cdot \frac{R_{\text{total}}}{R_{\text{max}}} \quad (14)$$

where P_{max} is the total power consumption of the device at maximum load, and α_D is the proportion of maximum dynamic power relative to total power. Therefore, $P_{\text{max}} \cdot \alpha_D$ represents the upper bound of dynamic power consumption. $R_{\text{max}} = \sum_{m=1}^M R_m^{\text{max}}$ is the sum of the maximum negotiated link rates of all STAs, representing the maximum achievable total rate.

3.2 MFU's Power Consumption Model

Following the modeling approach of the SFU, the total power consumption of the MFU in an FTTR system, P_{MFU} , can also be divided into base power consumption $P_{\text{M_Base}}$ and dynamic power consumption $P_{\text{M_Dynamic}}$:

$$P_{\text{MFU}} = P_{\text{M_Base}} + P_{\text{M_Dynamic}} \quad (15)$$

As the control center of the network, the MFU is usually required to remain continuously active to manage all SFUs, and its power characteristics differ from those of the SFU. Its base power consumption is a fixed value, while its dynamic power consumption mainly depends on the number of active SFUs under its management, calculated as:

$$\begin{cases} P_{\text{M_Base}} = P_{\text{M_max}} \cdot \alpha_{\text{M_Base}} \\ P_{\text{M_Dynamic}} = P_{\text{M_max}} \cdot \alpha_{\text{M_Dynamic}} \cdot \frac{N_{\text{active}}}{N} \end{cases} \quad (16)$$

where $P_{\text{M_max}}$ is the maximum power consumption of the MFU, and $\alpha_{\text{M_Base}}$ and $\alpha_{\text{M_Dynamic}}$ are the proportional coefficients of the base power and the maximum dynamic power relative to the maximum total power consumption, respectively. The model assumes that power consumption increases linearly with the number of active SFUs, N_{active} . When all SFUs are active, the dynamic power consumption reaches its upper limit, $P_{\text{M_max}} \cdot \alpha_{\text{M_Dynamic}}$.

3.3 FTTR System Energy Consumption Model

The total energy consumption of the FTTR system, E_{FTTR} , consists of the energy consumption of the MFU and the sum of all SFUs, which is expressed as:

$$E_{\text{FTTR}} = E_{\text{MFU}} + \sum_{i=1}^N E_{\text{SFU}}^i \quad (17),$$

where N denotes the set of SFUs in the system. As the MFU must remain active as the central network controller, its energy consumption can be expressed as the accumulation of its instantaneous power consumption P_{MFU} over the total operational time T_{tot} :

$$E_{\text{MFU}} = \int_0^{T_{\text{tot}}} P_{\text{MFU}} dt = P_{M_Base} \cdot T_{\text{tot}} + \int_0^{T_{\text{tot}}} P_{M_Dynamic} dt \quad (18).$$

Unlike the MFU, the SFU is the primary target of energy-saving strategies, as its operating state dynamically switches according to the employed mechanism. Therefore, the energy consumption of an SFU must be calculated by summing its consumption in different states. For a single SFU, based on the on or off state of its optical transceivers, three states are defined: active, doze, and sleep. The energy consumption of the i -th SFU is given by:

$$E_i = E_i^a + E_i^d + E_i^s + E_i^o \quad (19),$$

where E_i^a , E_i^d , E_i^s , and E_i^o represent the energy consumed in the active, doze, sleep, and state-transition modes, respectively. When the SFU is active, all modules remain on; in the doze state, the transmitter is turned off; and in the sleep state, both the transmitter and receiver are turned off. The energy consumption in each state can be calculated as:

$$E_i^a = (P_{\text{base}} + P_{\text{PON}} + P_{\text{WiFi}}) \cdot T_i^a = (P_{\text{base}} + P_{\text{Tx}} + P_{\text{Rx}} + P_{\text{WiFi}}) \cdot T_i^a \quad (20),$$

$$E_i^d = (P_{\text{base}} + P_{\text{Rx}} + P_{\text{WiFi}}) \cdot T_i^d \quad (21),$$

$$E_i^s = (P_{\text{base}} + P_{\text{WiFi}}) \cdot T_i^s \quad (22),$$

where T_i^a , T_i^d , and T_i^s denote the time durations of the SFU in the active, doze, and sleep states, respectively. In addition, when an SFU switches from sleep to the active state for data transmission, the processes of clock recovery, synchronization with the MFU, and laser activation are required. The transition overhead time is calculated as:

$$T_i^o = \max \{T_i^{\text{rec}} + T_i^{\text{sync}}, T_i^{\text{laser}}\} \quad (23),$$

where T_i^{rec} is the clock recovery time, T_i^{sync} is the synchronization time, and T_i^{laser} is the laser activation time. For downlink reception, the receiver must be woken up earlier, corresponding

to $T_{\text{downoverhead}} = T_i^{\text{rec}} + T_i^{\text{sync}}$. For uplink transmission, the transmitter must be activated in advance, corresponding to $T_{\text{upoverhead}} = T_i^{\text{laser}}$. Assuming the instantaneous power consumption during the wake-up process equals the active-state power of the corresponding modules, the transition energy is

$$E_i^o = (P_{\text{Tx}} + P_{\text{Rx}}) \cdot T_i^o \quad (24).$$

In summary, by substituting Eqs. (18) – (24) into Eq. (17), the total energy consumption of the FTTR system can be expressed as:

$$E_{\text{FTTR}} = (P_{M_Base} \cdot T_{\text{tot}} + \int_0^{T_{\text{tot}}} P_{M_Dynamic} dt) + \left[(P_{\text{base}} + P_{\text{Tx}} + P_{\text{Rx}} + P_{\text{WiFi}}) \cdot T_i^a + (P_{\text{base}} + P_{\text{Rx}} + P_{\text{WiFi}}) \cdot T_i^d + (P_{\text{base}} + P_{\text{WiFi}}) \cdot T_i^s + (P_{\text{Tx}} + P_{\text{Rx}}) \cdot T_i^o \right] \quad (25).$$

4 QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer

Based on the FTTR system and energy consumption models described in the previous section, this section elaborates on the MBES scheme. It first analyzes the limitations of existing energy-saving mechanisms and then presents the key concept and implementation of MBES, including the SFU-side multi-threshold wake-up mechanism and the MFU-side cooperative scheduling algorithm.

4.1 Limitations of Existing Energy-Saving Mechanisms in FTTR

The current G.fast standard for FTTR networks inherits WSM from the XG-PON protocol for energy management. WSM employs an instant wake-up strategy: whenever traffic arrives at the Wi-Fi interface of an SFU, the device immediately transitions from the low-power mode to the full-power mode. However, this approach exhibits two major limitations in addressing the spatially and temporally unbalanced traffic patterns commonly observed in FTTR networks. First, the strategy is coarse-grained and lacks differentiated QoS support. In scenarios with only small amounts of low-priority traffic, frequent instant wake-ups lead to unnecessary energy consumption. A buffer-based delayed wake-up mechanism, where wake-up is triggered only after traffic accumulates beyond a certain threshold, can mitigate this issue. Nevertheless, traditional single-threshold schemes risk degrading high-priority services (e.g., voice and video), which may experience excessive latency while waiting for low-priority packets to fill the buffer, thereby failing to meet stringent QoS requirements. Second, the strategy lacks cross-domain coordination and adaptive capability. Conventional PON energy-saving decisions rely solely on optical link status, overlooking conditions on the Wi-Fi side. One of the key advan-

tages of FTTR architecture is that the SFU, as a converged node for optical and wireless domains, can collect cross-domain metrics such as Wi-Fi access latency and relay this information to the MFU for centralized analysis^[3]. However, the existing WSM fails to leverage this architectural benefit and cannot dynamically adjust energy-saving parameters based on end-to-end service performance, making it difficult to achieve an optimal balance between energy efficiency and QoS assurance.

In summary, existing schemes are primarily single-threshold-based, lack cross-domain coordination, and seldom incorporate end-to-end latency feedback for closed-loop control. These limitations motivate the design of MBES, which differentiates queue priorities, leverages the MFU's centralized control capabilities, and dynamically adjusts threshold settings based on latency feedback.

4.2 SFU's Multi-Threshold Wake-Up Mechanism

On the SFU side, a differentiated wake-up mechanism based on multi-threshold buffering is adopted. This mechanism performs fine-grained buffer management and wake-up control according to QoS. Specifically, an independent buffer threshold is configured for each T-CONT queue, which corresponds to the four service categories: VO, VI, BE, and BK. High-priority services (e.g., VO) are assigned lower thresholds so that even a small amount of data can trigger the optical transmitter to wake up and request bandwidth, thereby meeting their low-latency requirements. In contrast, low-priority services (e.g., BK) are assigned higher thresholds to aggregate more packets and reduce unnecessary wake-ups, thus maximizing the energy-saving cycle.

Fig. 4 shows the process of the SFU multi-threshold wake-up mechanism. When the buffered data of any T-CONT queue reaches its preset threshold, the optical transmitter of the SFU is awakened and reports the current buffer status of all queues to the MFU to request bandwidth. After completing the data transmission, the transmitter is turned off again. For the downlink, since the MFU operates with a fixed broadcast cycle, the receiver of the SFU adopts a periodic listening strategy. It only wakes briefly when it needs to receive downlink data or signaling, while remaining off during other times to save energy.

To ensure the effectiveness of this mechanism, its operating model and constraints are constructed from three aspects: buffer-capacity constraints, transceiver switching strategy, and the relationship between buffer threshold and latency.

1) Buffer constraint

Assume there are N SFUs in the system, and each SFU supports K T-CONT queues with different priorities. For the k -th priority queue of the i -th SFU ($i \in \{1, \dots, N\}, k \in \{1, \dots, K\}$), let the current buffered data be denoted as $Q_{i,k}$, and the preset wake-up threshold as $Th_{i,k}^{base}$.

An SFU reports its buffer demand for all queues to the MFU only when at least one queue buffer reaches or exceeds its threshold; otherwise, the SFU remains in an energy-saving mode without reporting. This relationship can be expressed as:

$$R_{i,k} = \begin{cases} Q_{i,k}, \exists Q_{i,k} \geq Th_{i,k}^{base}, k = 1, \dots, K \\ 0, \text{otherwise} \end{cases} \quad (26),$$

where $R_{i,k}$ represents the buffer information of the k -th queue of the i -th SFU reported to the MFU. In addition, the configuration of all thresholds must comply with the physical buffer-capacity constraint of each SFU. Let the maximum buffer capacity of the i -th SFU be B_i . The sum of thresholds of all priority queues must not exceed this limit:

$$\sum_{k=1}^K Th_{i,k}^{base} \leq B_i \quad (27).$$

2) Transmitter shut-down strategy

The switching strategy of the transmitter follows certain strict timing constraints. We define the start and end time of the j -th bandwidth allocation cycle as $[T_{j,start}, T_{j,end}]$. Within this cycle, if the i -th SFU is assigned an uplink timeslot starting at $T_{i,j}^{start}$ with

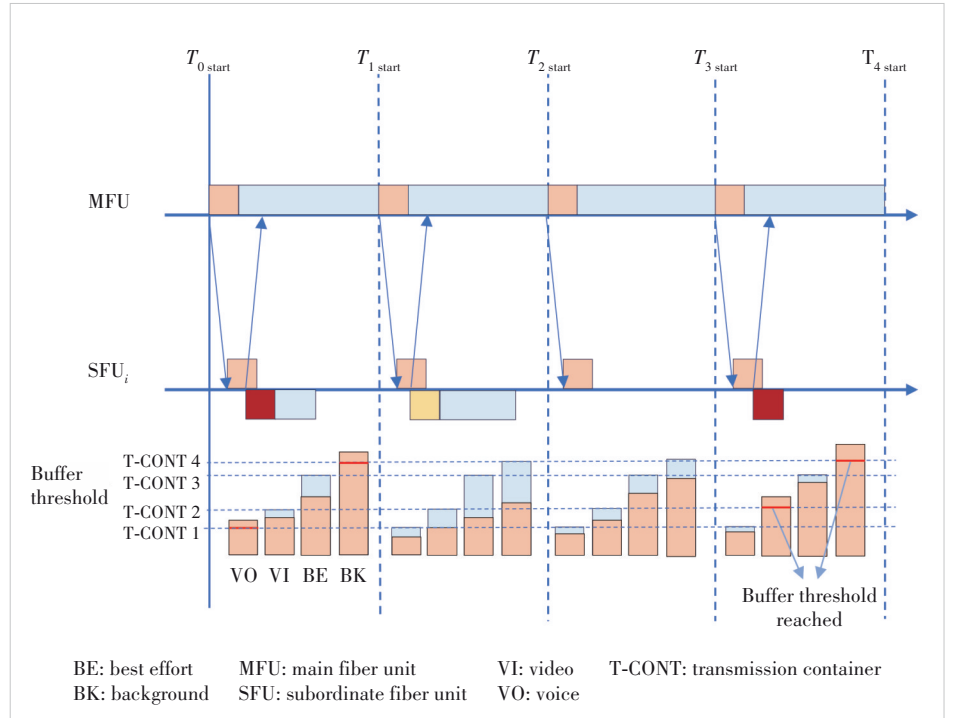


Figure 4. Process of the SFU's multi-threshold wake-up mechanism

length $\text{Len}_{i,j}$, this timeslot must be fully contained within the cycle, satisfying the following boundary conditions:

$$T_{j \text{ start}} \leq T_{i,j}^{\text{start}} \quad (28),$$

$$T_{i,j}^{\text{start}} + \text{Len}_{i,j} \leq T_{j \text{ end}} \quad (29).$$

To transmit data on time, the SFU must wake up its transmitter in advance. If the transmitter requires a start-up overhead $T_{\text{upoverhead}}$, the wake-up time is determined as:

$$T_{i,j}^{\text{wake}} = T_{i,j}^{\text{start}} - T_{\text{upoverhead}} \quad (30).$$

After completing one uplink transmission, the SFU can shut down its transmitter to save energy. However, this operation is beneficial only if the idle interval between two consecutive transmissions is sufficiently long. Specifically, the idle time from the end of transmission $j-1$ to the beginning of transmission j must be greater than or equal to the transmitter start-up overhead $T_{\text{upoverhead}}$:

$$T_{i,j-1}^{\text{start}} + \text{Len}_{i,j-1} + T_{\text{upoverhead}} \leq T_{i,j}^{\text{wake}} \quad (31).$$

In addition, to ensure fairness in energy saving among multiple SFUs, the MFU should maintain a fixed relative order in the allocation of uplink timeslots.

3) Receiver shut-down strategy

Let $T_{i,j}^0$ denote the moment when the i -th SFU parses the downlink frame from the MFU and obtains its allocated timeslot in the j -th bandwidth allocation cycle. It is calculated as:

$$T_{i,j}^0 = T_{j \text{ start}} + T_{\text{DBA}} + T_i^{\text{tran}} + T_G^{\text{proc}} \quad (32),$$

where $T_{j \text{ start}}$ is the start time of the j -th cycle, T_{DBA} is the computation overhead of DBA, T_i^{tran} is the propagation delay from the MFU to the i -th SFU, and T_G^{proc} is the processing time for the SFU to parse the downlink frame. If the duration of receiving the downlink data frame is denoted as T_{down} , the completion time of the i -th SFU receiving all downlink traffic in the j -th cycle is:

$$T_{i,j}^1 = T_{i,j}^0 + T_{\text{down}} \quad (33).$$

After completing downlink reception, the SFU decides whether to shut down its receiver based on energy-saving benefits. The prerequisite for executing the shutdown operation is that the idle interval from the completion of this reception ($T_{i,j}^1$) to the start of the next reception ($T_{i,j+1}^0$) must be greater than or equal to the receiver start-up overhead $T_{\text{downoverhead}}$:

$$T_{i,j+1}^0 - T_{i,j}^1 \geq T_{\text{downoverhead}} \quad (34).$$

4) Relationship between SFU buffer thresholds and latency

In the FTTR energy-saving scenario, the end-to-end latency

of the k -th priority packet at the i -th SFU, denoted as $T_{i,k}$, can be expressed as:

$$T_{i,k} = T_{i,k}^{\text{WiFi}} + T_{i,k}^{\text{PON}} + T_{\text{trans}} = T_{i,k}^{\text{WiFi}} + (T_{\text{report}} + T_{\text{upoverhead}} + T_i^{\text{sleep}}) + T_{\text{trans}} \quad (35),$$

where $T_{i,k}^{\text{WiFi}}$ is the latency on the Wi-Fi side caused by contention or retransmission, and T_{trans} is the optical transmission delay. Since the distance between the SFU and MFU in FTTR systems is very short, T_{trans} can be ignored. The optical-side latency $T_{i,k}^{\text{PON}}$ is further decomposed into three parts: the reporting overhead T_{report} , the actual sleep waiting time in the queue T_i^{sleep} , and the transmitter start-up overhead $T_{\text{upoverhead}}$.

To satisfy QoS requirements, $T_{i,k}$ must not exceed the latency bound D_k , i.e., $D_k \geq T_{i,k}$. When a packet arrives at an SFU, its $T_{i,k}^{\text{WiFi}}$ is already fixed. Therefore, the maximum tolerable optical-side latency can be derived and translated into the maximum allowed number of waiting cycles, denoted as $n_{i,k}$. This value represents the maximum number of scheduling cycles that a packet can wait in the SFU buffer without violating its latency bound:

$$n_{i,k} = \left\lfloor \frac{D_k - T_{i,k}^{\text{WiFi}} - T_{\text{report}} - T_{\text{upoverhead}}}{T_{\text{cycle}}} \right\rfloor \quad (36).$$

The SFU does not need to track the countdown of each packet individually. Instead, it maintains a global minimum waiting cycle, denoted as n_i , which is the smallest value among all $n_{i,k}$. This represents the “most urgent” packet in the buffer. When a new packet (the m -th arrival) enters, the SFU calculates its $n_{i,k,m}$ and updates the global counter.

$$n_i = \min \{ n_{i,k,1}, n_{i,k,2}, \dots, n_{i,k,m} \} \quad (37).$$

The counter n_i acts as a mandatory countdown timer and decreases by one at the end of each scheduling cycle. When n_i reaches zero, the SFU must wake up and request bandwidth to meet latency targets, regardless of buffer thresholds. Thus, the unified bandwidth-reporting condition that combines latency-triggered and buffer-threshold-triggered events is

$$R_{i,k} = \begin{cases} Q_{i,k}, \exists Q_{i,k} \geq Th_{i,k}, k = 1, \dots, K \\ Q_{i,k}, n_i = 0 \\ 0, \text{otherwise} \end{cases} \quad (38).$$

Under the QoS constraint, buffer thresholds are further adjusted dynamically to optimize energy saving, with $Th_{i,k}^{\text{base}}$ serving only as the initial value. The threshold directly affects the balance between energy saving and latency: a high threshold yields better energy saving but larger latency, while a low threshold does the opposite. For the i -th SFU in the j -th bandwidth allocation cycle, if the bandwidth allocated to priority k is $G_{i,j,k}$, the newly arrived data is $Q_{i,j,k}$, and the bandwidth demand

reported in the previous cycle is $R_{ij-1,k}$. The total amount of data at the beginning of the current cycle is:

$$Q_{ij,k}^{\text{total}} = R_{ij-1,k} - G_{ij,k} + Q_{ij,k} \quad (39).$$

When $Q_{ij,k}^{\text{total}} \geq Th_{i,k}^{\text{base}}$, a bandwidth report is triggered. It is evident that the time to reach the threshold depends on the packet arrival rate λ_k . To adapt the thresholds to traffic variations, the arrival rate is estimated by averaging over the previous r cycles:

$$\lambda_k = \frac{1}{r} \sum_{j=n-r+1}^n \frac{Q_{ij,k}}{t_{ij}} \quad (40),$$

where t_{ij} is the effective observation duration in cycle j . Accordingly, the adaptive threshold update mechanism driven by the arrival rate is defined as:

$$Th_{i,k} = \alpha_k \cdot \lambda_k \cdot (n_{i,k} \cdot T_{\text{cycle}}) \quad (41),$$

where α_k is the weight factor of the priority- k service, and $(n_{i,k} \cdot T_{\text{cycle}})$ is its maximum waiting time. With this design, the thresholds decrease adaptively for high-priority services or traffic bursts to suppress queueing latency, while the thresholds increase for low-priority or low-traffic services to maximize energy savings.

4.3 MFU's Cooperative Scheduling and Dynamic Threshold Adjustment

The MFU allocates bandwidth to multi-priority services based on buffer reports and Wi-Fi-side latency information from SFUs. During this process, it dynamically adjusts buffer thresholds to meet QoS requirements. Since wake-up is performed locally by SFUs, the MFU does not require additional state decisions, which reduces signaling overhead and allows it to focus on global scheduling. Specifically, the MFU first collects the bandwidth requests $R_{i,k}$ from all active SFUs. It then calculates the end-to-end latency $D_{i,k} = T_{i,k}^{\text{Wi-Fi}} + T_i^{\text{sleep}}$, which combines the Wi-Fi-side latency $T_{i,k}^{\text{Wi-Fi}}$ and the sleep-induced latency T_i^{sleep} . The MFU sorts the set $\{D_{i,k}\}$ in descending order to obtain the index mapping $\sigma(i)$, and reorders the bandwidth request set $\{R_{i,k}\}$ accordingly into $\{R_{\sigma(i),k}\}$. This ensures that SFUs with longer latency are prioritized in bandwidth allocation. For SFUs whose buffer thresholds are not reached, the MFU retains their scheduling order but does not assign bandwidth, thereby maintaining overall fairness. Bandwidth is then allocated from T-CONT 1 to T-CONT 4 in descending order of priority. Let the uplink frame length be L_f (μs), the total available uplink capacity be C (bit/s), the maximum proportions for different priorities be $\gamma_1, \gamma_2, \gamma_3$, and γ_4 , and the number of active SFUs be n .

For bandwidth-guaranteed services T-CONT 1 and T-CONT 2, each active SFU first receives a basic bandwidth $BW_{i,k}$, proportionally divided according to the service proportion and the

number of active SFUs.

$$BW_{i,1} = \frac{L_f \cdot C \cdot \gamma_1}{n} \quad (42),$$

$$BW_{i,2} = \frac{L_f \cdot C \cdot \gamma_2}{n} \quad (43).$$

For the highest-priority T-CONT 1 services, a "fixed allocation and excess grant" mechanism is adopted. If its reported demand $R_{i,1}$ exceeds the basic bandwidth, the MFU grants additional bandwidth to ensure the timely transmission of latency-sensitive traffic. Thus, the actual bandwidth $G_{i,1}$ allocated to each active SFU for T-CONT 1 is:

$$G_{i,1} = \min \left\{ \max \{ BW_{i,1}, R_{i,1} \}, \frac{L_f \cdot C - \sum_{i=1}^n BW_{i,2}}{n} \right\} \quad (44),$$

where $\frac{L_f \cdot C - \sum_{i=1}^n BW_{i,2}}{n}$ represents the upper allocation bound after deducting the basic bandwidth of T-CONT 2, ensuring that all active SFUs of T-CONT 1 obtain transmission opportunities. T-CONT 2 follows a similar allocation logic, but its actual granted bandwidth is limited by the allocation result of T-CONT 1:

$$G_{i,2} = \min \left\{ \max \{ BW_{i,2}, R_{i,2} \}, \frac{L_f \cdot C - \sum_{i=1}^n G_{i,1}}{n} \right\} \quad (45).$$

After allocating bandwidth to T-CONT 1 and T-CONT 2, the MFU distributes the remaining bandwidth to T-CONT 3 and T-CONT 4 in sequence. The granted bandwidth is subject to both the services' bandwidth cap and the residual link capacity. For T-CONT 3, the allocation is

$$\begin{cases} BW_{i,3} = \frac{L_f \cdot C \cdot \gamma_3}{n} \\ G_{i,3} = \min \left\{ BW_{i,3}, R_{i,3}, L_f \cdot C - \sum_{i=1}^n (G_{i,1} + G_{i,2}) - \sum_{i=1}^{i-1} G_{i,3} \right\} \end{cases} \quad (46).$$

Similarly, for T-CONT 4, the allocation is

$$\begin{cases} BW_{i,4} = \frac{L_f \cdot C \cdot \gamma_4}{n} \\ G_{i,4} = \min \left\{ BW_{i,4}, R_{i,4}, L_f \cdot C - \sum_{i=1}^n (G_{i,1} + G_{i,2} + G_{i,3}) - \sum_{i=1}^{i-1} G_{i,4} \right\} \end{cases} \quad (47).$$

Through this mechanism, the highest-priority services receive rigid guarantees, the second-highest priority services ob-

tain conditional guarantees, and the lowest-priority services are scheduled only when surplus bandwidth is available.

After completing the bandwidth allocation, the MFU proceeds to dynamically adjust the buffer thresholds for each SFU. This adjustment process adheres to two fundamental principles. First, the threshold should be corrected based on the deviation between the services' measured latency T_k^{measured} and its target latency deadline D_k . Second, after adjustment, the threshold of a high-priority queue must remain no greater than that of a low-priority queue to maintain the priority order. If the measured latency of a certain priority class exceeds its requirement, that is

$$\Delta_k = T_k^{\text{measured}} - D_k > 0 \quad (48),$$

and then in the next scheduling cycle, its buffer threshold is decreased:

$$Th_{i,k}^{\text{new}} = Th_{i,k}^{\text{old}} - \eta_k \cdot \Delta_k \quad (49),$$

where η_k is the step size for decreasing the threshold of priority k . To prevent the violation of priority order, a correction is applied:

$$Th_{i,k}^{\text{new}} = \min \{Th_{i,k}^{\text{new}}, Th_{i,k+1}^{\text{new}}\} \quad (50).$$

Conversely, when $\Delta_k < 0$ and this condition holds for n_Δ consecutive cycles, it indicates that the threshold is overly conservative. In this case, the threshold can be moderately increased:

$$Th_{i,k}^{\text{new}} = \min \{Th_k^{\text{max}}, Th_{i,k}^{\text{old}} - \beta_k \cdot \Delta_k\} \quad (51),$$

where Th_k^{max} is the maximum allowable buffer threshold for priority k , and β_k is the step size for increasing the threshold.

In summary, the MFU's energy-saving bandwidth allocation follows three principles. First, it allocates bandwidth among active SFUs in order of service priority, meeting latency requirements of different services. Second, within the same priority class, it prioritizes the scheduling of SFUs with higher Wi-Fi side latency to improve end-to-end latency performance. Third, it dynamically adjusts buffer thresholds based on service latency feedback, achieving a balance between QoS assurance and energy saving.

5 Simulation Results and Analyses

This section evaluates the performance of the proposed MBES using a simulation environment built on the Java platform. The MBES is compared with two other baseline strategies: the no energy saving (NES) scheme and the single-threshold buffer energy saving (SBES) scheme. The simulation scenario is configured as a home network, with average system latency and energy consumption as the primary performance metrics. This evaluation aims to assess the effectiveness of energy saving by the different strategies under service performance constraints, thereby validating the efficiency of MBES.

5.1 Parameter Setting

Given the small scale of FTTR deployments in residential settings, the number of SFUs is limited and each SFU typically serves only a few STAs. In our simulations, each SFU is therefore assumed to serve two STAs, while the total number of SFUs varies from 2 to 5. Each STA generates a constant aggregate traffic load of 120 Mbit/s, apportioned equally across four priority services—VO, VI, BE, and BK—at 30 Mbit/s per service. Traffic arrivals follow a Poisson process with an average flow duration of 1 s, and packet sizes are fixed at 1 500 bytes. On the wireless side, we model IEEE 802.11ax and use the enhanced distributed channel access (EDCA) mechanism. Detailed PHY/MAC parameter values are listed in Table 1.

For the optical link, key parameters are configured according to the G. fin protocol standard^[9]. The downlink and uplink capacities are set to 10 Gbit/s and 2.5 Gbit/s, respectively, and the physical fiber length between the MFU and SFUs is 10 m. The bandwidth allocation period is 125 μ s. In the MFU scheduling strategy, bandwidth assignment considers both real-time SFU requests and predefined service-level constraints. To reflect differentiated service priorities, the maximum proportions of the total uplink bandwidth that T-CONT 1–4 queues can occupy are set to 0.2, 0.5, 0.3, and 0.1, respectively, as shown in Table 2. Importantly, their sum is not required to equal 1, allowing the MFU to flexibly perform statistical multiplexing based on real-time traffic demand. Two baseline schemes are implemented for comparison. In NES, all SFUs remain continuously active, and the MFU applies the classic GIANT algorithm^[30], simplified by setting the service interval of all services to 1 to guarantee allocation in every cycle. In SBES, each SFU employs a single global buffer threshold to trigger its optical transmitter; once active, it follows the same allocation procedure as in NES. The MFU-side bandwidth scheduling parameters used

Table 1. Simulation scenario, traffic, and EDCA parameter settings

Scenario Settings			
Number of SFUs	2/3/4/5	Number of Stations per SFU	2
Traffic Settings			
VO data rate	30 Mbit/s	VI data rate	30 Mbit/s
BE data rate	30 Mbit/s	BK data rate	30 Mbit/s
Packet size	1 500 B	Traffic duration	1 s
EDCA Settings			
Access category	CWmin	CWmax	AIFSN
VO	3	7	2
VI	7	15	3
BE	15	1 023	6
BK	15	1 023	9

AIFSN: arbitration inter-frame space number
 BE: best effort
 BK: background
 CWmax: maximum contention window
 CWmin: minimum contention window
 EDCA: enhanced distributed channel access
 SFU: subordinate fiber unit
 VI: video
 VO: voice

Table 2. MFU bandwidth scheduling parameter settings

Scenario Settings			
MFU-SFU distance	10 m	Total uplink bandwidth	2.5 Gbit/s
SFU uplink rate	2.5 Gbit/s	Total downlink bandwidth	10 Gbit/s
Proposed Scheme (MBES)			
DBA Type	MBES	Allocation period and frame length	125 μ s
T1 total bandwidth ratio	0.2	T2 total bandwidth ratio	0.5
T3 total bandwidth ratio	0.3	T4 total bandwidth ratio	0.1
Baseline Scheme (NES/SBES)			
DBA Type	GIANT	Allocation period and frame length	125 μ s
T1 Total Bandwidth Ratio	0.2	T2 total bandwidth ratio	0.5
T3 Total Bandwidth Ratio	0.3	T4 total bandwidth ratio	0.1
SI_{\max}	1	SI_{\min}	1

DBA: dynamic bandwidth allocation
 MFU: main fiber unit
 MBES: multi-threshold buffer energy saving
 NES: no energy saving
 SBES: single-threshold buffer energy saving
 SFU: subordinate fiber unit

Table 3. Parameter settings for the power consumption model

Parameter	Value	Parameter	Value	Parameter	Value
P_{base}	1	P_{Tx}	1.2	P_{Rx}	0.8
$P_{\text{RF_base}}$	0.6	P_{\max}	6	α_D	0.2
D	5 m	σ	1 dB	G_A	3 dBi
M	5 GHz	$W_k(5G)$	15 dB	j_k	0
$P_L(5G)$	23 dBm	$R_{\max}^m(5G)$	2 402 Mbit/s	RSSI	-40 dBm
P_{M_max}	10	α_{M_Base}	0.4	$\alpha_{M_Dynamic}$	0.6
$T_{\text{upoverhead}}$	125 μ s	$T_{\text{downoverhead}}$	125 μ s		

in the simulations are summarized in Table 2.

For the energy consumption analysis, the power parameters of the SFU and MFU functional modules are listed in Table 3. These values are normalized based on data from actual products^[31–32]. The SFU's base power consumption is set to 1 unit, while its optical transmitter and receiver consume 1.2 and 0.8 units, respectively. The wake-up overhead for the optical transceiver is 125 μ s^[33]. The SFU's maximum power consumption is capped at 6 units, with dynamic power consumption accounting for up to 20% (1.2 units). The MFU's maximum power consumption is set to 10 units, with the base and maximum dynamic power consumptions accounting for 40% and 60%, respectively; it is assumed to operate at full power throughout the simulation. On the Wi-Fi side, each RF module has a base power consumption of 0.6 units. The dynamic transmission power of the Wi-Fi module is calculated according to link parameters. According to the Wi-Fi

6 standard, the simulation models Wi-Fi transmission in the 5 GHz band with a 160 MHz channel bandwidth, using modulation and coding scheme (MCS) index 11. The SFU is assumed to support 2×2 MIMO with a maximum transmission rate of 2 402 Mbit/s, and the STA parameters are configured to match. To determine the RF module's transmission power, the initial target RSSI at the STA is set to -40 dBm^[34]. The actual transmission power under the EIRP constraint and the resulting received RSSI are calculated using the channel parameters in Table 3 and Eqs. (5) – (11), which determine the final dynamic power consumption of the Wi-Fi module.

Under the MBES strategy, the buffer thresholds and latency requirements of different priority services directly influence the on/off timing of SFU transmitters, thereby affecting both energy efficiency and service latency. Table 4 summarizes the latency requirements, buffer thresholds, adjustment step sizes, and parameter settings used for the SBES comparison scheme. For high-priority services, the initial buffer thresholds are determined based on traffic arrival rates and latency requirements and are dynamically adjusted during operation according to the parameters in Table 4. For comparison, the buffer threshold in the SBES scheme is set to the sum of all priority-specific thresholds used in the multi-threshold strategy.

5.2 Result Analysis

This subsection analyzes the latency and energy consumption of the proposed MBES scheme. In the result figures, the legends NES, MBES, and SBES represent the respective schemes, while VO, VI, BE, and BK denote voice, video, best-effort, and background services, respectively.

Fig. 5 illustrates how the average system latency of the FTTR network varies with the number of SFUs. Under the NES scheme, in which SFUs actively report in every cycle and the MFU bandwidth is sufficient, latency for all service priorities remains stable, with BK service latency staying below 5 ms. In contrast, the energy-saving schemes introduce additional latency because SFUs enter a sleep state when their buffer thresholds are not reached, delaying reporting and increasing optical-side latency. Among these, the MBES scheme demonstrates superior performance over SBES: by assigning lower thresholds to high-priority services, MBES triggers earlier reporting, enabling

Table 4. Buffer threshold parameter settings

Parameter	VO	VI	BE	BK
Latency requirement/ms	2	10	30	50
Initial buffer threshold/kB	3	30	60	100
Step size (decrease/increase)/(kB/ms)	2/1	2/1	2/1	2/1
Consecutive cycle threshold (cycle)	10	5	5	5
Maximum buffer threshold/kB	100	200	500	1 000
Single threshold buffer/kB	196			
Physical buffer constraint /MB	32			

BE: best effort BK: background VI: video VO: voice

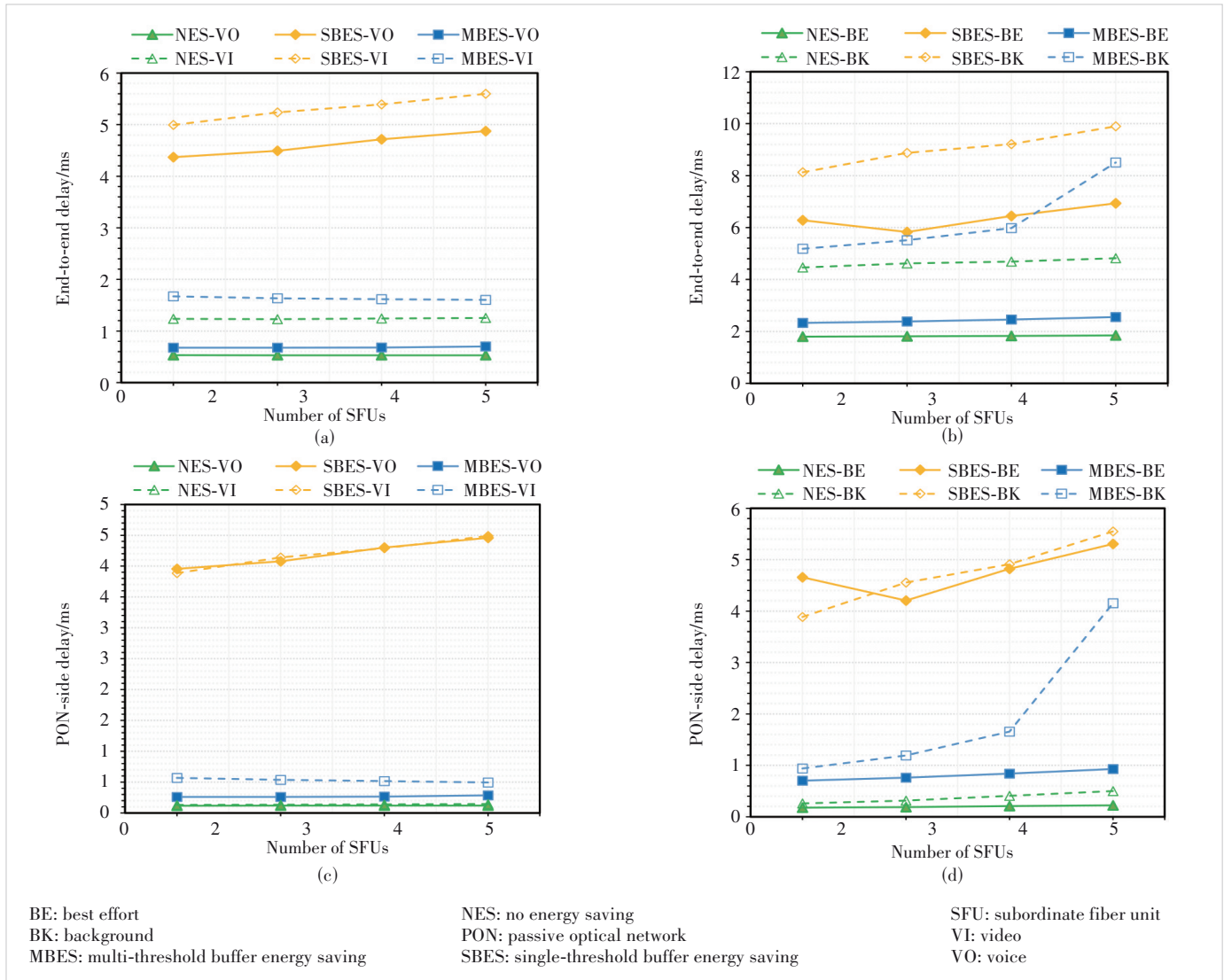


Figure 5. Average system latency under the NES, SBES, and MBES schemes (STA=2): (a) end-to-end latency of high-priority services, (b) end-to-end latency of low-priority services, (c) optical-side latency of high-priority services, and (d) optical-side latency of low-priority services

the concurrent transmission of lower-priority traffic and effectively reducing overall latency. Conversely, SBES uses a single unified buffer threshold for all services, causing SFUs to report less frequently, accumulate larger data bursts, and ultimately experience higher latency. The results in Figs. 5c and 5d show that under the energy-saving schemes, the optical-side latency of BK services increases most significantly, and this effect becomes more pronounced as the number of SFUs increases. The impact is especially severe in the SBES scheme, where BK traffic must wait for high-priority transmissions and can remain buffered for extended periods when bandwidth is insufficient. The MBES scheme alleviates this problem because frequent reporting triggered by high-priority services increases the scheduling opportunities for BK traffic. Additionally, under the SBES scheme, the optical-side latency of VO services is comparable to, and in some cases slightly higher than, that of VI services.

This is because, with a unified threshold, VO traffic often enters the MFU buffer and is reported and transmitted together with VI traffic, reducing the latency gap between the two. As the number of SFUs increases, system bandwidth allocation becomes more constrained, amplifying the reporting delays introduced by energy-saving schemes. The MBES strategy achieves a balance between energy savings and latency by applying differentiated buffer thresholds, whereas the SBES strategy, with its longer reporting cycles, not only increases queuing delays for low-priority services but also weakens the scheduling advantage of high-priority services. These findings highlight that setting appropriate buffer thresholds and dynamically adjusting them based on service priorities is essential for maintaining service performance while optimizing energy efficiency in FTTR systems.

Fig. 6 compares the total FTTR system energy consumption and the average SFU energy consumption of the NES, MBES,

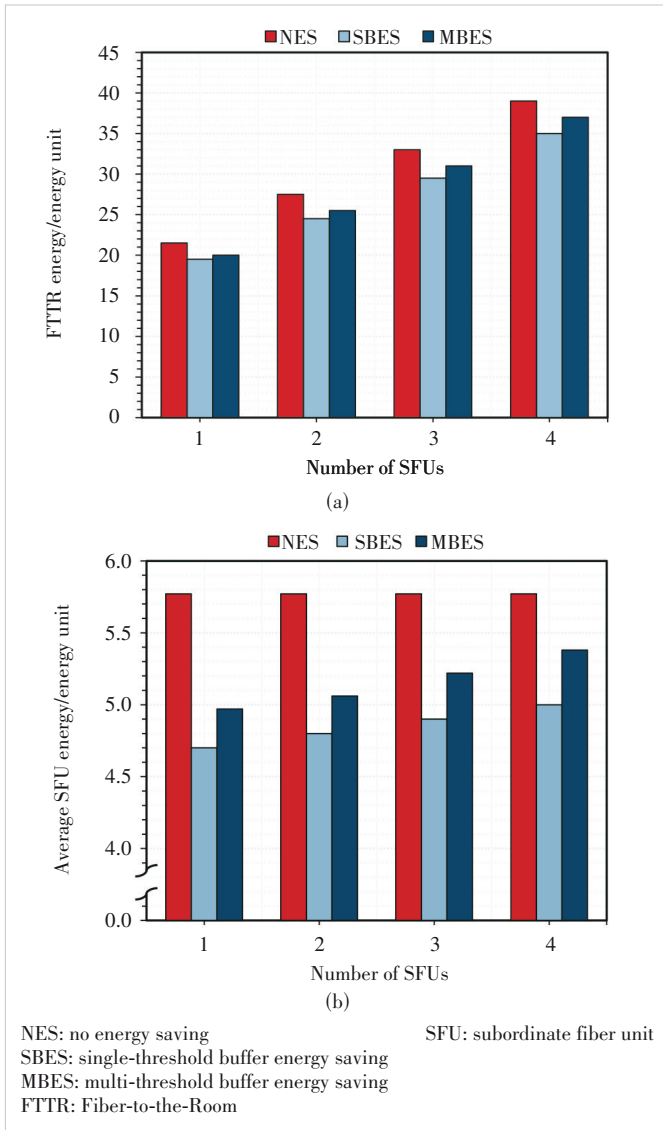


Figure 6. System energy consumption under the NES, SBES, and MBES schemes (STA=2): (a) FTTR system energy consumption and (b) average SFU energy consumption

and SBES schemes as the number of SFUs increases from 1 to 5. System energy consumption is calculated according to Eq. (25) and expressed in power units. As shown in Fig. 6a, the NES scheme exhibits the highest total energy consumption. The SBES scheme achieves the lowest consumption, reducing energy use by approximately 9.50% to 10.43% compared with NES. The MBES scheme consumes slightly more than SBES but still provides savings of about 4.58% to 7.28% compared with NES. Fig. 6b further shows that the average SFU energy consumption follows a similar trend. Consumption is the highest under the NES scheme and the lowest under the SBES scheme, achieving savings of 13.02% to 17.76% compared to NES. Although slightly higher than SBES, the MBES scheme also provides substantial energy savings of 6.17% to 13.62% compared with NES.

Furthermore, with the uplink rate fixed at 2.5 Gbit/s, the energy-saving effect tends to diminish when the number of SFUs increases to five, rather than improving continuously. This occurs because adding more SFUs reduces the available bandwidth per SFU. Once an SFU reaches its buffer threshold and triggers a report, it must activate its optical transmitter more frequently and use additional transmission cycles to send the buffered data, thereby lowering the overall energy-saving efficiency.

In the FTTR system, the SFU optical transmitter can operate at a higher uplink rate of 10 Gbit/s in addition to 2.5 Gbit/s. Therefore, the simulation also evaluates the performance of the proposed energy-saving strategy at this higher rate. For the 10 Gbit/s system, the SFU optical transmitter power consump-

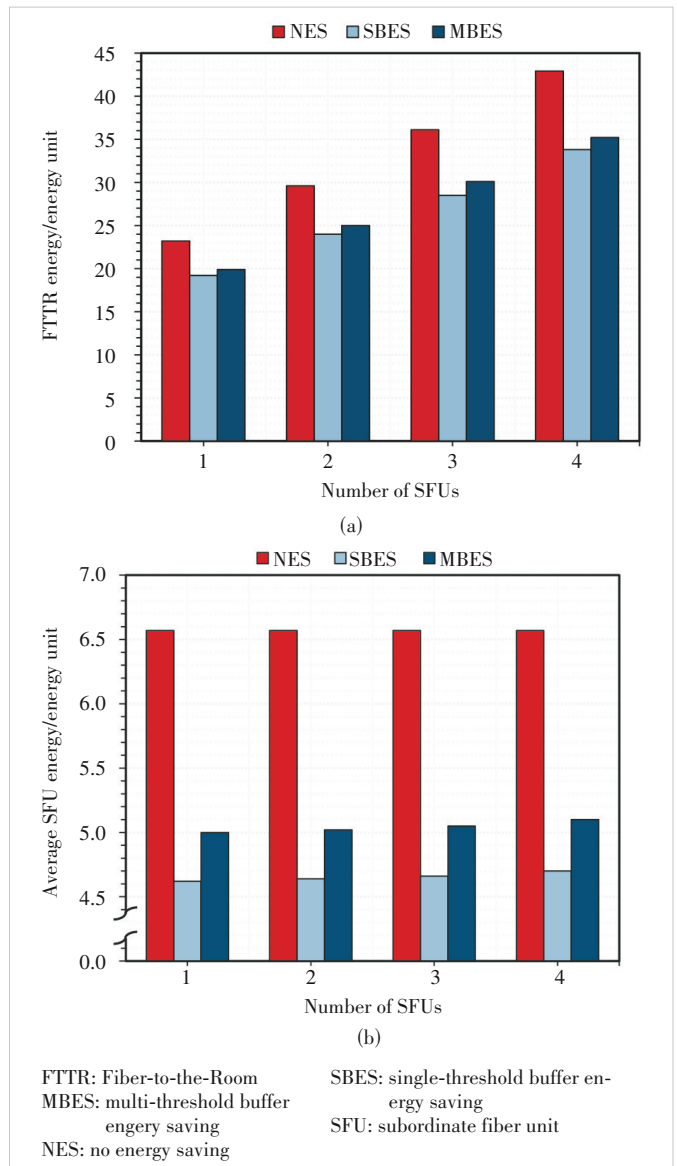


Figure 7. System energy consumption with a 10 Gbit/s uplink rate under the NES, SBES, and MBES schemes (STA=2): (a) FTTR system energy consumption and (b) average SFU energy consumption

tion is assumed to be 2 units, with a maximum of 6.8 units. All other power parameters remain consistent with those listed in Table 3. Fig. 7 presents the simulation results at 10 Gbit/s. With the substantial increase in transmitter power, the energy-saving effects of both MBES and SBES are more pronounced than at 2.5 Gbit/s. Furthermore, the advantages of both schemes increase as the number of SFUs grows. Specifically, for the total system energy consumption, the maximum savings achieved by SBES and MBES relative to NES reach 21.41% and 17.75%, respectively, compared with 10.43% and 7.28% at 2.5 Gbit/s. For average SFU energy consumption, the maximum savings are 29.11% (SBES) and 23.30% (MBES), substantially higher

than the 17.76% and 13.62% observed at 2.5 Gbit/s. These results demonstrate that the energy-saving potential of buffer-based schemes increases significantly with higher SFU optical transmitter power. This finding also highlights the effectiveness of the proposed strategy for next-generation FTTR systems, which are expected to adopt PONs with even higher data rates, such as 50G PON.

To evaluate the impact of the number of STAs on energy-saving performance and service latency, the number of SFUs is fixed at 2, while the number of STAs per SFU is increased from 1 to 4. Traffic and scheduling parameters are listed in Tables 1 and 2, respectively. Fig. 8 shows the change in average FTTR

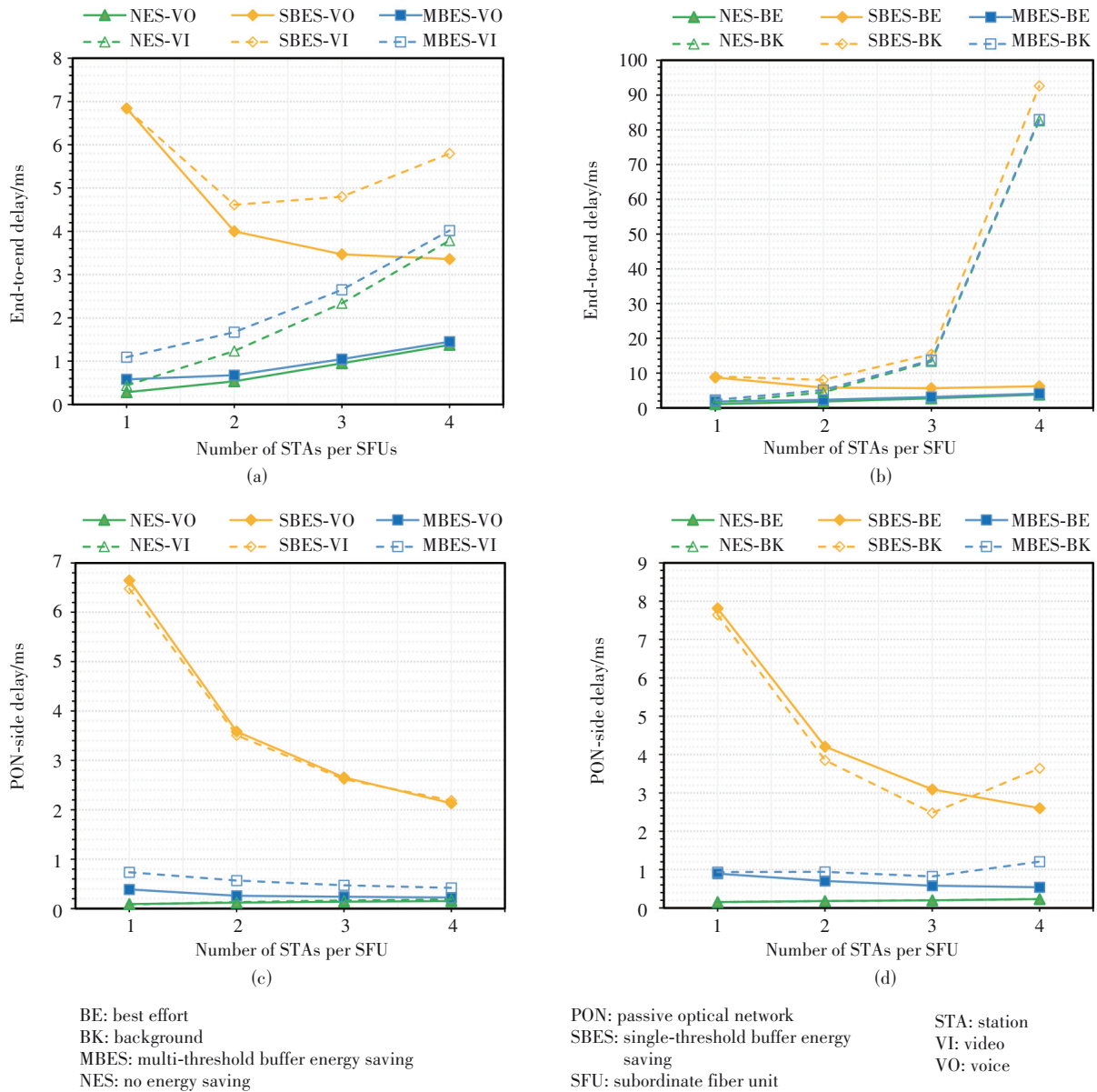


Figure 8. System latency under the NES, SBES, and MBES schemes (SFU=2): (a) end-to-end latency of high-priority services, (b) end-to-end latency of low-priority services, (c) optical-side latency of high-priority services, and (d) optical-side latency of low-priority services

system latency as the number of STAs increases. When two SFUs are deployed, end-to-end latency for all priority services rises with the number of STAs under both MBES and NES schemes. This is primarily due to increased contention on the Wi-Fi side, leading to higher EDCA-based access latency. Under NES, optical-side latency fluctuates only slightly because SFUs report every cycle and system bandwidth is sufficient. In MBES, more STAs increase the traffic arrival rate at the SFU, allowing buffers to reach thresholds faster, reducing SFU sleep time and optical-side latency. However, sleep duration is largely governed by the low thresholds of high-priority services, so SFUs do not enter long sleep periods even under low traffic loads. Consequently, optical-side latency fluctuations under MBES remain minor, and end-to-end latency is dominated by EDCA contention. As shown in Figs. 8c and 8d, although optical-side latency for BK services increases when the number of STAs reaches 4, end-to-end latency is still primarily determined by Wi-Fi access contention. Under SBES, with a small number of STAs, end-to-end latency is mainly affected by the queuing delay needed to fill the buffer. As the number of STAs increases, higher traffic volumes allow thresholds to be reached faster, reducing this delay. However, beyond a certain number of STAs, EDCA contention dominates, causing latency to rise again. For BK services, optical-side latency increases significantly when the number of STAs is 4, because BK traffic lacks guaranteed bandwidth; even after the buffer threshold is met, its scheduling is often postponed by high-priority traffic, leading to accumulated latency.

Fig. 9 compares the total FTTR system energy consumption and the average SFU energy consumption under the NES, MBES, and SBES schemes as the number of STAs per SFU increases from 1 to 4, with the number of SFUs fixed at 2. The results show that as the number of STAs grows, higher traffic volumes cause the buffer thresholds in the energy-saving schemes to be reached more quickly. This prompts SFUs to activate their optical transmitters more frequently for data reporting and transmission, reducing sleep time and gradually diminishing the energy-saving effects of all three schemes. Specifically, for total system energy consumption, savings from MBES decrease from 9.12% to 4.65%, while SBES savings decrease from 10.33% to 7.39%. For average SFU energy consumption, MBES savings decline from 17.05% to 8.70%, and SBES savings decrease from 19.32% to 13.83%.

To evaluate the impact of buffer threshold settings on the energy-saving strategy, we compare the performance of MBES under dynamic and fixed thresholds. The scenario uses 2 SFUs, each connected to 2 STAs. Three MBES configurations are tested: dynamic thresholds (starting from low fixed values), fixed low thresholds (VO threshold set to 1 500 B, with other priorities scaled proportionally), and fixed high thresholds (VO threshold set to 20 KB, with other priorities unchanged). Traffic parameters, dynamic threshold adjustment parameters, and latency requirements are provided in Tables 1 and 4. As shown

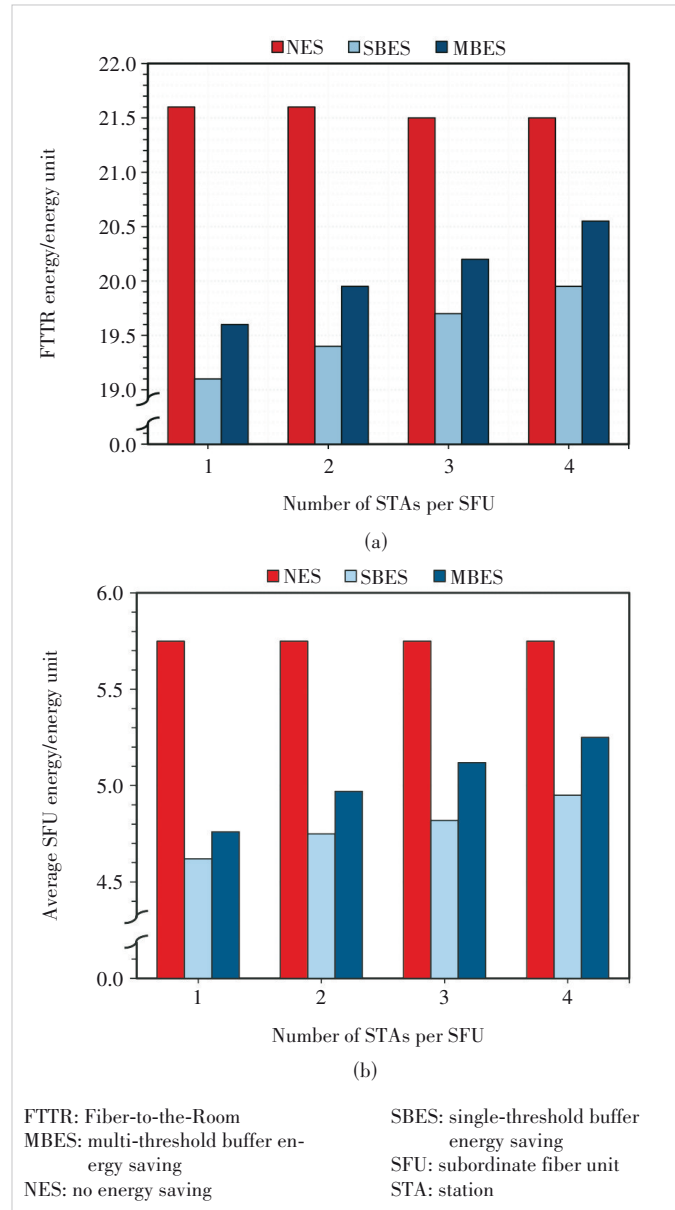


Figure 9. System energy consumption under the NES, SBES, and MBES schemes (SFU=2): (a) FTTR system energy consumption and (b) average SFU energy consumption

in Fig. 10a, the fixed-high-threshold scheme achieves the best energy savings, reducing average SFU consumption by 14% and total system consumption by 7.4%. However, the higher buffer threshold increases the buffer filling time, resulting in a significant reporting delay. Fig. 10b shows that VO service latency rises to 2.36 ms—an increase of 1.82 ms compared to the NES scheme—failing to meet the QoS requirement. In contrast, the fixed-low-threshold scheme ensures latency performance, with VO latency increasing by only 0.05 ms, but its energy savings are limited, achieving only 4.8% reduction in average SFU consumption and 2.5% reduction in total system consumption. Finally, the dynamic threshold scheme provides a better trade-

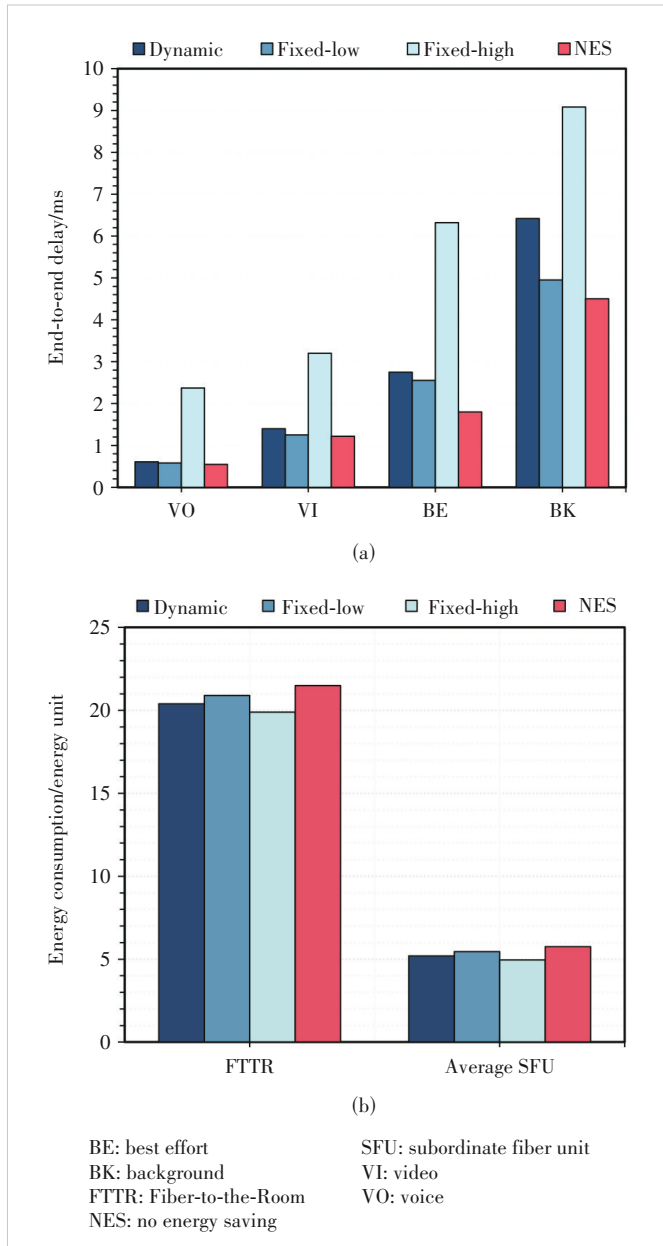


Figure 10. System latency and energy consumption under dynamic and fixed buffer thresholds: (a) latency comparison and (b) energy consumption comparison

off between energy and latency. By starting with low thresholds and gradually increasing them for services that consistently meet latency requirements, this approach reduces transmitter activation frequency and improves energy efficiency. Simulation results indicate that dynamic thresholds increase VO latency by only 0.083 ms while reducing average SFU consumption by 9.7% and total system consumption by 5.1%. Thus, the dynamic threshold scheme, through its adaptive adjustment mechanism, overcomes the limitations of fixed-threshold approaches and achieves co-optimization of latency assurance and energy savings.

6 Conclusions

This paper proposes an MBES scheme to address the increasing energy consumption in FTTR networks. By leveraging the architectural features of FTTR, the scheme enables efficient coordination between the MFU's centralized scheduling and the SFU's differentiated buffering mechanism, reducing system energy consumption while maintaining multi-level QoS. Specifically, at the SFU side, independent and dynamically adjustable buffer thresholds are assigned to different priority services, providing differentiated handling for latency-sensitive traffic. At the MFU side, the algorithm incorporates cooperative scheduling and dynamic threshold adjustment, taking Wi-Fi access latency into account during bandwidth allocation and using real-time performance feedback to form a closed-loop, cross-domain adaptive energy-saving mechanism.

Simulation results validate the effectiveness of the proposed MBES scheme. Compared with NES and SBES, MBES has achieved superior performance in meeting the strict latency requirements of high-priority services, such as VO and VI, while significantly reducing system energy consumption. It also provides better control over average system latency than SBES and maintains stable QoS guarantees under high-load, multi-user scenarios. In terms of energy savings, MBES has demonstrated considerable potential, particularly in future 10 Gbit/s scenarios with higher-power optical modules, achieving savings of up to 17.75%. Furthermore, for the MBES scheme, comparisons between dynamic and fixed threshold configurations show that the dynamic adjustment mechanism delivers a superior balance between latency and energy consumption, adaptively optimizing the performance-efficiency trade-off as traffic conditions vary.

References

- [1] MAROTTA A, VALCARENGHI L, KONDEPU K, et al. Fiber to the room challenges and opportunities [C]//The 16th International Conference on Communication Systems & NETWORKS (COMSNETS). IEEE, 2024: 1139 - 1142. DOI: 10.1109/COMSNETS59351.2024.10427248
- [2] ETSI. Fifth Generation Fixed Network (F5G); F5G generation definition release #1 [S/OL]. (2020-12-10) [2025-09-10]. https://www.etsi.org/deliver/etsi_gr/F5G/001_099/001/01.01_01_60/gr_F5G001v010101p.pdf
- [3] ZHANG D C, ZHU J L, LIU X, et al. Fiber-to-the-room: a key technology for F5G and beyond [J]. Journal of optical communications and networking, 2023, 15(9): D1 - D9. DOI: 10.1364/JOCN.485070
- [4] Broadband Development Alliance. White paper on fiber to the room (FTTR) [R/OL]. [2025-09-10]. <http://www.chinabda.cn/article/252789>
- [5] CAI J H, SHEN G X, LI J, et al. Is fiber-to-the-room (FTTR) green? Modeling and analysis of power and energy consumption [J]. IEEE transactions on green communications and networking, 2025, 9(2): 522 - 535. DOI: 10.1109/TGCN.2024.3447874
- [6] ITU-T. G.984.3: gigabit-capable passive optical networks (G-PON): transmission convergence layer specification [S/OL]. [2025-09-10]. <https://www.itu.int/rec/T-REC-G.984.3/en>
- [7] ITU-T. G.987.3: 10-Gigabit-capable passive optical networks (XG-PON): transmission convergence (TC) layer specification [S/OL]. [2025-09-10]. <https://www.itu.int/rec/T-REC-G.987.3/en>
- [8] ITU-T. G.9804.2: higher speed passive optical networks—common transmis-

- sion convergence layer specification [S/OL]. [2025-09-10]. <https://www.itu.int/rec/T-REC-G.9804.2/en>
- [9] ITU-T. High speed fibre-based in-premises transceivers-data link layer [S/OL]. (2024-07-05) [2025-09-10]. <https://www.itu.int/rec/T-REC-G.9942-202407-1/en>
- [10] YAN Y, WONG S W, VALCARENGHI L, et al. Energy management mechanism for Ethernet passive optical networks (EPONs) [C]/IEEE International Conference on Communications. IEEE, 2010: 1 – 5. DOI: 10.1109/ICC.2010.5502659
- [11] SHI L, MUKHERJEE B, LEE S S. Energy-efficient PON with sleep-mode ONU: progress, challenges, and solutions [J]. IEEE network, 2012, 26(2): 36 – 41. DOI: 10.1109/MNET.2012.6172273
- [12] ZHANG J J, ANSARI N. Toward energy-efficient 1G-EPON and 10G-EPON with sleep-aware MAC control and scheduling [J]. IEEE communications magazine, 2011, 49(2): s33 – s38. DOI: 10.1109/MCOM.2011.5706311
- [13] BUTT R A, AKHUNZADA A, FAHEEM M, et al. Enhanced energy savings with adaptive watchful sleep mode for next generation passive optical network [J]. Energies, 2022, 15(5): 1639. DOI: 10.3390/en15051639
- [14] ZIN A M, IDRUS S M, RAMLI A, et al. Performance evaluation of XG-PON with DBA based-watchful sleep mode [C]/The 7th International Conference on Photonics (ICP). IEEE, 2018: 1 – 3. DOI: 10.1109/ICP.2018.8533168
- [15] LAMBERT S, LANNOO B, DIXIT A, et al. Energy efficiency analysis of high speed triple-play services in next-generation PON deployments [J]. Computer networks, 2015, 78: 68 – 82. DOI: 10.1016/j.comnet.2014.10.037
- [16] DHAINI A R, HO P H, SHEN G X, et al. Energy efficiency in TDMA-based next-generation passive optical access networks [J]. IEEE/ACM transactions on networking, 2014, 22(3): 850 – 863. DOI: 10.1109/TNET.2013.2259596
- [17] SHAH NEWAZ S H, CUEVAS A, LEE G M, et al. Evaluating energy efficiency of ONUs having multiple power levels in TDM-PONs [J]. IEEE communications letters, 2013, 17(6): 1248 – 1251. DOI: 10.1109/LCOMM.2013.043013.122648
- [18] ZIN A M, IDRUS S M, ISMAIL N A, et al. Determination of optimized sleep interval for 10 gigabit-passive optical network using learning intelligence [J]. International journal of electrical and computer engineering (IJECE), 2022, 12(3): 2663. DOI: 10.11591/ijece.v12i3.pp2663-2671
- [19] IEEE. IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements part 11: wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN (802.11ax-2021) [S/OL]. [2025-09-10]. <https://ieeexplore.ieee.org/document/9442429>
- [20] LORINCZ J, CAPONE A, BEGUŠIĆ D. Heuristic algorithms for optimization of energy consumption in wireless access networks [J]. KSII transactions on Internet and information systems, 2011: 626 – 648. DOI: 10.3837/tiis.2011.04.001
- [21] PACK S, CHOI Y. An adaptive power saving mechanism in IEEE 802.11 wireless IP networks [J]. Journal of communications and networks, 2005, 7(2): 126 – 134. DOI: 10.1109/JCN.2005.6387860
- [22] GARROPPO R G, NENCIONI G, PROCISSI G, et al. The impact of the access point power model on the energy-efficient management of infrastructure wireless LANs [J]. Computer networks, 2016, 94: 99 – 111. DOI: 10.1016/j.comnet.2015.11.018
- [23] SILVA P, ALMEIDA N T, CAMPOS R. A comprehensive study on enterprise Wi-Fi access points power consumption [J]. IEEE access, 2019, 7: 96841 – 96867. DOI: 10.1109/ACCESS.2019.2928754
- [24] DEMBÉLÉ H, TOHME E, COLIN R. Assessing and modeling the energy consumption of PoE-powered WiFi access point [J]. IEEE access, 2023, 11: 74796 – 74804. DOI: 10.1109/ACCESS.2023.3295689
- [25] NISHIYAMA H, TOGASHI K, KAWAMOTO Y, et al. A cooperative ONU sleep method for reducing latency and energy consumption of STA in smart-FiWi networks [J]. IEEE transactions on parallel and distributed systems, 2015, 26(10): 2621 – 2629. DOI: 10.1109/TPDS.2014.2360405
- [26] CHOWDHURY P, TORNATORE M, SARKAR S, et al. Building a green wireless-optical broadband access network (WOBAN) [J]. Journal of lightwave technology, 2010, 28(16): 2219 – 2229. DOI: 10.1109/JLT.2010.2044369
- [27] HAN P C, GUO L, LIU Y J, et al. Joint wireless and optical power states scheduling for green multi-radio fiber-wireless access network [J]. Journal of lightwave technology, 2016, 34(11): 2610 – 2623. DOI: 10.1109/JLT.2016.2529644
- [28] SCHÜTZ G, CORREIA N. Design of QoS-aware energy-efficient fiber-wireless access networks [J]. Journal of optical communications and networking, 2012, 4(8): 586 – 594. DOI: 10.1364/JOCN.4.000586
- [29] ISLAM M M, FUNABIKI N, SAHA M, et al. An improvement of throughput measurement minimization method for access-point transmission power minimization in wireless local-area network [C]/International Conference on Consumer Electronics (ICCE-TW). IEEE, 2019: 1 – 2. DOI: 10.1109/ICCE-TW46550.2019.8991715
- [30] LELIGOU H C, LINARDAKIS C, KANONAKIS K, et al. Efficient medium arbitration of FSAN-compliant GPONs [J]. International journal of communication systems, 2006, 19(5): 603 – 617. DOI: 10.1002/dac.761
- [31] Huawei Technologies Co., Ltd. Huawei OptiXstar V173 Datasheet 01 [DB/OL]. (2023-05-15) [2025-09-10]. <https://e.huawei.com/cn/products/optical-terminal/optixstar-v173>
- [32] Huawei Technologies Co., Ltd. Huawei OptiXstar K153 Datasheet 01 [DB/OL]. [2025-09-10]. <https://e.huawei.com/cn/products/optical-terminal/optixstar-k153>
- [33] DHAINI A R, HO P H, SHEN G X. Toward green next-generation passive optical networks [J]. IEEE communications magazine, 2011, 49(11): 94 – 101. DOI: 10.1109/MCOM.2011.6069715
- [34] ZHANG J, LIU J, XIANG L, et al. Full-link AoI analysis of uplink transmission in next-generation FTTR WLANs [C]/Proceedings of IEEE 97th Vehicular Technology Conference (VTC2023-Spring). IEEE, 2023: 1 – 7. DOI: 10.1109/VTC2023-Spring57618.2023.10200633

Biographies

CAI Jinhan received his BE degree from Nanjing University of Science and Technology Zijin College, China in 2021 and MS degree in information and communication engineering from Soochow University, China in 2024, where he is currently pursuing his PhD degree with the School of Electronic and Information Engineering. His research interests include passive optical networks and fiber-wireless access networks.

ZAN Mingyuan received his BE degree from Nanjing University of Information Science and Technology, China in 2022 and MS degree in information and communication engineering from Soochow University, China in 2025. His research interests include passive optical networks and fiber-wireless access networks.

SHEN Gangxiang (shengx@suda.edu.cn) received his BE degree from Zhejiang University, China, MS degree from Nanyang Technological University, Singapore, and PhD degree from The University of Alberta, Canada in 2006. He is a Distinguished Professor and the Dean of the School of Electronic and Information Engineering at Soochow University, China. He previously worked as a Lead Engineer with Ciena, USA, and an ARC Postdoctoral Fellow with The University of Melbourne, Australia. His research interests include spectrum-efficient, green optical networks and integrated optical-wireless systems. He has authored over 300 peer-reviewed papers and received the Izaak Walton Killam Memorial Scholarship and the NSERC Fellowship. He was recognized as a Highly Cited Chinese Researcher by Elsevier (2014 – 2024) and is an Associate Editor of the *IEEE/OPTICA Journal of Lightwave Technology*. He is an OSA Fellow and was an IEEE Distinguished Lecturer of ComSoc.



C-WAN for FTTR: Enabling Low-Overhead Joint Transmission with Deep Learning

ZHANG Yang¹, CEN Zihan¹, ZHAN Wen², CHEN Xiang¹

(1. School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China;

2. School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China)

DOI: 10.12142/ZTECOM.202504008

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251218.1001.002.html>,
published online December 18, 2025

Manuscript received: 2025-09-20

Abstract: Fiber-to-the-Room (FTTR) networks with multi-access point (AP) coordination face significant challenges in implementing Joint Transmission (JT), particularly the high overhead of Channel State Information (CSI) acquisition. While the centralized wireless access network (C-WAN) architecture inherently provides high-precision synchronization through fiber-based clock distribution and centralized scheduling, efficient JT still requires accurate CSI with low signaling cost. In this paper, we propose a deep learning-based hybrid model that synergistically integrates temporal prediction and spatial reconstruction to exploit spatiotemporal correlations in indoor channels. By leveraging the centralized data and computational capability of the C-WAN architecture, the model reduces sounding frequency and the number of antennas required per sounding instance. Experimental results on a real-world synchronized channel dataset show that the proposed method lowers over-the-air resource consumption while maintaining JT performance close to that achieved with ideal CSI, offering a practical low-overhead solution for high-performance FTTR systems.

Keywords: Fiber-to-the-Room (FTTR); Joint Transmission (JT); centralized wireless access network (C-WAN); deep learning; Channel State Information (CSI)

Citation (Format 1): ZHANG Y, CEN Z H, ZHAN W, et al. C-WAN for FTTR: enabling low-overhead joint transmission with deep learning [J]. *ZTE Communications*, 2025, 23(4): 65 – 76. DOI: 10.12142/ZTECOM.202504008

Citation (Format 2): Y. Zhang, Z. H. Cen, W. Zhan, “C-WAN for FTTR: enabling low-overhead joint transmission with deep learning,” *ZTE Communications*, vol. 23, no. 4, pp. 65 – 76, Dec. 2025. doi: 10.12142/ZTECOM.202504008.

1 Introduction

In recent years, the extensive deployment of Fiber-to-the-Home (FTTH) has enabled gigabit and even multi-gigabit bandwidth capabilities for household access networks^[1].

However, the in-home wireless segment, constrained by issues such as uncontrolled medium competition over the air interface, has increasingly become a performance bottleneck affecting user experience. It often fails to deliver the high-speed, low-latency, and reliable connectivity promised by FTTH, thereby limiting the end-to-end network performance perceived by users^[2]. To fundamentally address the wireless coverage and capacity challenges in this last segment of the in-home network, Fiber-to-the-Room (FTTR) technology has emerged and is being widely deployed. This trend marks a new era of high-density, multi-access point (AP) coordinated deployment for home Wi-Fi systems^[3].

In a typical FTTR system, a Main FTTR Unit (MFU) is connected via fiber or hybrid fiber to Subordinate FTTR Units (SFUs) deployed in individual rooms, forming a star topology. This architecture, by independently deploying SFUs in each room, significantly mitigates signal attenuation caused by obstacles like walls, achieves seamless high-speed coverage

throughout the entire household, and greatly enhances network coverage capability and achievable data rates. However, the deployment of high-density APs also introduces a series of new challenges, the most critical being how to effectively achieve coordination among multiple SFUs to avoid co-channel interference and exploit potential cooperative gain.

Multi-AP coordinated transmission, particularly Joint Transmission (JT) technology, is regarded as the key to unlocking the performance potential of FTTR networks. JT allows multiple SFUs to cooperatively serve the same Station (STA) on the same time-frequency resources^[4]. Through coherent superposition of signals over the air, the Signal-to-Noise Ratio (SNR) at the receiver is significantly improved, thereby enhancing throughput and spectral efficiency^[5]. However, the practical deployment of JT technology faces two core challenges^[6]. First is the PHY layer synchronization challenge. JT requires the participating transmitters to maintain high consistency in frequency, time and phase. Any Carrier Frequency Offset (CFO) or transmission timing deviation can disrupt the coherence of the signals, leading to a loss of combining gain, or even triggering a negative cooperation effect, resulting in performance degradation^[7]. Second is the Chan-

nel State Information (CSI) acquisition overhead challenge. The performance of JT highly depends on accurate and timely downlink CSI. Traditional methods require frequent over-the-air sounding to acquire CSI, which consumes substantial wireless resources and generates significant signaling overhead in high-density AP scenarios, consequently reducing the effective throughput of the system.

For the first challenge, when multiple Analog-to-Digital Converters (ADCs) on a single AP use the same source clock and are managed uniformly by a local processor, there is no time-frequency synchronization problem among the antennas. However, when extended to multiple APs, each AP employs an independent local oscillator, introducing CFOs between antennas of different APs, which severely affects JT performance. To address the synchronization problem in distributed multi-AP systems, Ref. [8] proposed that subordinate APs could perform frequency calibration and time synchronization based on the Null Data Packet Announcement (NDPA) frame sent by the master AP. However, this imposes high computational requirements on the subordinate APs, and non-ideal factors over the air make it difficult to guarantee the calibration effectiveness. Ref. [9] proposed that subordinate APs calibrate their local oscillators by snooping the directional synchronization signals sent by the master AP to achieve time-frequency synchronization. However, this method introduces additional delay and over-the-air overhead. Meanwhile, Ref. [10] proposed a pre-calibration matrix to compensate for random phase offsets among distributed arrays, yet this solution requires manual calibration efforts. Although Ref. [11] introduced a master controller (MC)-based architecture that improves AP scheduling for JT, it fails to address the fundamental synchronization requirement for simultaneous multi-AP transmissions, consequently forcing sequential channel sounding across APs and resulting in even greater over-the-air overhead. These synchronization approaches share a common limitation: their reliance solely on over-the-air signaling mechanisms without leveraging the underlying FTTR infrastructure inevitably leads to substantial overhead. This fundamental constraint underscores the necessity of an architecture-level solution.

For the second challenge, as noted by Ref. [12], the inherent correlations of wireless channels across spatial, frequency, and temporal domains allow CNN-based techniques (originally successful in high-resolution image inpainting) to be adapted for pilot reduction and CSI interpolation. Similarly, DL-based schemes have been proven effective in compressing CSI feedback payload by exploiting these physical correlations^[13]. For instance, Ref. [14] introduced a one-sided deep learning framework that leverages plug-and-play priors to recover CSI from linear projections, effectively capturing spatial and frequency-domain characteristics without requiring joint encoder-decoder training. Meanwhile, in high-mobility scenarios, the method proposed in Ref. [15] esti-

mates full downlink CSI from partial uplink measurements by exploiting both spatial correlation and temporal dependency, demonstrating strong performance even with incomplete CSI. While these data-driven approaches have shown promise, they are predominantly designed and evaluated within the context of independent APs. A critical limitation of such frameworks is their inherent lack of architectural support for system-wide time-frequency synchronization and unified triggering. Consequently, they fail to coherently align and exploit the rich spatial correlation across multiple, distributed APs, which is precisely the key to unlocking the full performance potential of coordinated systems like JT^[4]. This architectural gap underscores the necessity of a deeply integrated solution that co-designs the network architecture and the learning algorithm, which is the core contribution of our proposed C-WAN framework.

This paper aims to address the aforementioned challenges by proposing a low-overhead JT solution coupled with the C-WAN architecture of FTTR. The C-WAN architecture uses the MFU as the centralized control center^[16], enabling high-precision clock distribution and joint triggering to each SFU through the fiber fronthaul network. This solves the multi-node synchronization problem at the physical layer, laying a solid foundation for the application of JT technology. Furthermore, this paper focuses on overcoming the CSI acquisition overhead challenge under the C-WAN architecture. We fully utilize the characteristic of centralized computing power at the MFU and the inherent spatiotemporal correlation of indoor channels to propose a DL-based CSI acquisition framework. The core of this framework is a hybrid deep learning model. It employs a temporal prediction branch and a spatial reconstruction branch to leverage historical CSI sequences for predicting future channel states and utilize observations from a subset of antennas to reconstruct full-dimensional CSI, respectively. The temporal prediction branch serves as the core component that delivers accurate estimates of future channel states, while the spatial reconstruction branch plays a complementary yet crucial role in both accelerating loss function convergence and providing performance improvements. This approach can substantially reduce the frequency of sounding and the number of antennas required per sounding instance, thereby significantly lowering over-the-air resource overhead while guaranteeing JT performance.

The remainder of this paper is organized as follows. Section 2 introduces the system model and formally defines the key challenges in JT implementation. Section 3 presents the C-WAN architecture and elaborates on its inherent advantages for overcoming synchronization limitations. Section 4 details the proposed deep learning-based CSI acquisition framework, including the hybrid model design and specialized loss function. Section 5 provides comprehensive experimental results and performance analysis. Finally, Section 6 concludes the paper with summary remarks and future research directions.

2 System Model and Challenges in JT Implementation

2.1 FTTR System Architecture and JT Transmission Model

A typical FTTR system, as shown in Fig. 1, consists of one MFU and K SFUs distributed in different rooms, connected via a fiber network in a star topology. The MFU acts as the control center of the system, responsible for network management, resource scheduling, and data exchange; while the SFUs serve as distributed wireless access points, responsible for signal transmission and reception.

Within this architecture, JT is employed as a key technique to enhance the quality of service for STAs. Its core idea is that multiple SFUs cooperatively transmit the same data stream to the STA on the same time-frequency resource block. Under ideal conditions, these signals combine coherently at the STA's receiver, significantly enhancing the SNR.

For quantitative analysis, we establish the following JT system model for the general case. Consider K SFUs participating in JT transmission to a single STA. The received signal at the STA can be expressed as:

$$y = \sum_{k=1}^K \sqrt{P_k} h_k e^{j\theta_k} s_k + n \quad (1),$$

where P_k , h_k , and θ_k represent the transmit power, complex channel gain, and phase offset (relative to a common reference) of the k -th SFU, respectively. Here, s_k denotes the transmitted symbol of k -th SFU satisfying $\mathbb{E}[|s|^2] = 1$, and n is the additive white Gaussian noise (AWGN) with power σ_n^2 . Under ideal synchronization conditions (i.e., perfect frequency, time, and phase alignment among all SFUs^[17–18]) and narrowband channel assumptions neglecting Inter-Symbol Interference (ISI), the expression simplifies as the phase terms become coherent. In this case, the power of the useful signal component is given by:

$$P_{rx} = \left| \sum_{k=1}^K \sqrt{P_k} h_k \right|^2 \quad (2).$$

For comparison, in non-cooperative transmission (where each SFU independently transmits non-coherent signals to the STA), the received signal is given by $y_{\text{non-coop}} = \sum_{k=1}^K \sqrt{P_k} h_k s_k + n$, and the total received power is simply the sum of the individual received power from each independent link: $P_{rx, \text{non-coop}} = \sum_{k=1}^K P_k |h_k|^2$. In contrast, JT achieves additional coherent combining gain, expressed as $P_{rx}/P_{rx, \text{non-coop}}$, which is typically greater than 1. This gain significantly enhances both spectral efficiency and transmission reliability compared to non-cooperative transmission.

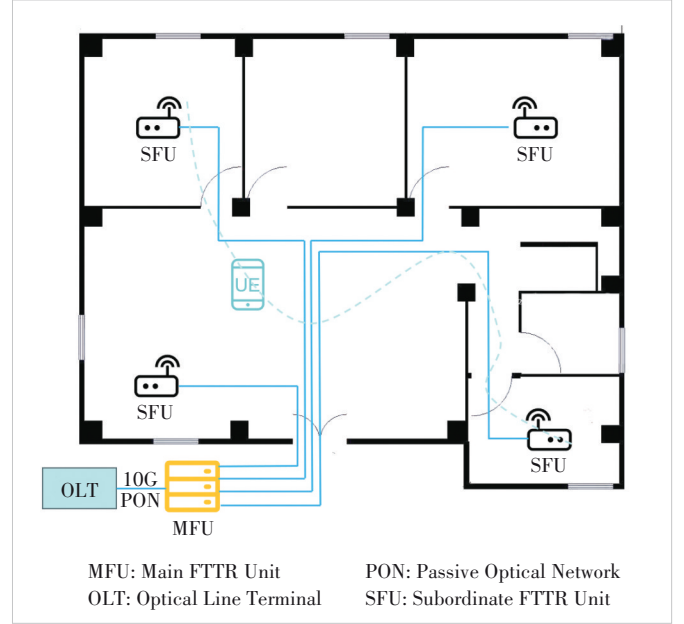


Figure 1. Typical Fiber-to-the-Room (FTTR) system

2.2 Key Challenges in JT Implementation

Although JT offers significant theoretical gains, its implementation in traditional distributed FTTR architectures (Fig. 2a), where SFUs operate independently, faces two major challenges: the PHY layer synchronization challenge and the CSI acquisition overhead challenge.

2.2.1 PHY Layer Synchronization Challenge

Each SFU uses an independent local oscillator, whose inherent crystal oscillator drift introduces CFOs (Δf_k). Simultaneously, the lack of a unified transmission trigger mechanism among distributed nodes introduces random transmission delay deviations ($\Delta \tau_k$), which subsequently translate into phase offsets ($\Delta \phi_k = 2\pi f_c \Delta \tau_k$). In this case, the received signal becomes time-varying:

$$y(t) = \sum_{k=1}^K \sqrt{P_k} h_k s(t - \tau_k) e^{j(2\pi f_c t + \phi_k)} \quad (3),$$

where $s(t)$ is the transmitted symbol waveform, τ_k is the propagation delay, f_c is the carrier frequency, and ϕ_k is the initial phase. Considering the two-SFU ($K = 2$) scenario, the received power can be expanded as:

$$P_{rx}(t) = P_1 |h_1|^2 + P_2 |h_2|^2 + 2\sqrt{P_1 P_2} |h_1 h_2| \cos(2\pi \Delta f \cdot t + \Delta \phi) \quad (4),$$

where $\Delta f = \Delta f_2 - \Delta f_1$ and $\Delta \phi = \Delta \phi_2 - \Delta \phi_1$. The cosine time-varying term in this expression indicates that the received power fluctuates periodically at the difference frequency Δf , causing unstable SNR and increased Bit Error Rate (BER),

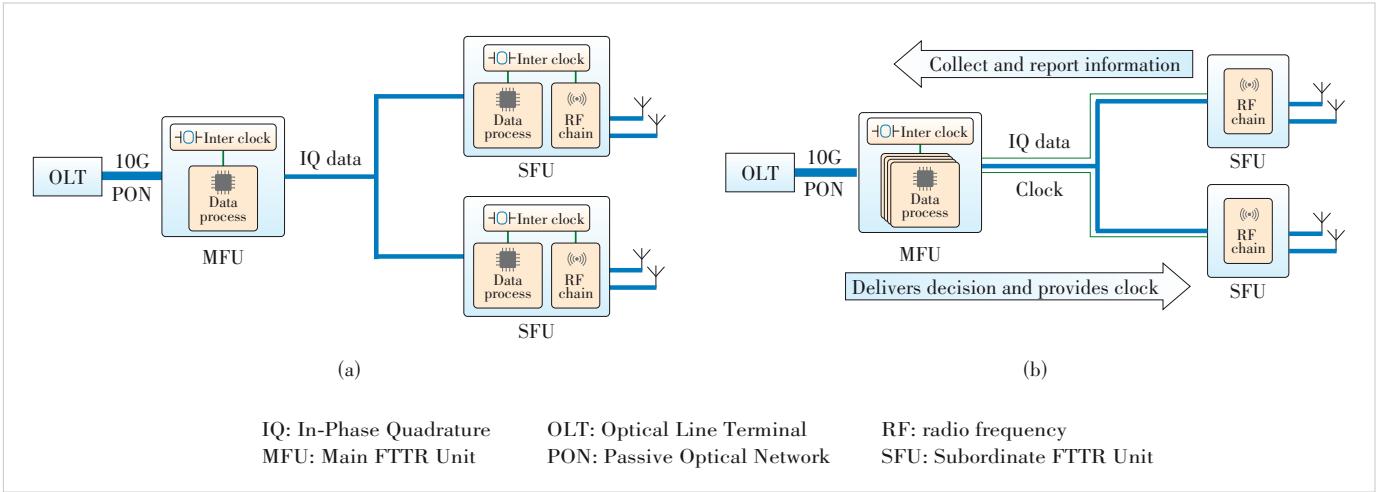


Figure 2. (a) Traditional distributed Fiber-to-the-Room (FTTR) architecture; (b) FTTR/centralized wireless access network (C-WAN) architecture

which severely disrupts the coherent combining effect of JT.

2.2.2 CSI Acquisition Overhead Challenge

The performance of JT also highly depends on accurate and timely downlink CSI. Each SFU needs to perform precoding based on CSI to adjust its beam direction precisely towards the target STA^[19]. However, acquiring the joint CSI for multiple APs requires frequent over-the-air sounding procedures.

In the traditional distributed architecture, to perform multi-AP channel sounding, a leading AP is needed to coordinate other APs to send Null Data Packet (NDP) frames in a roughly synchronized manner to the STA. The STA then needs to feed back the measured CSI. This process introduces substantial over-the-air signaling overhead and feedback delay, consuming significant air interface resources and severely constraining the system's effective throughput. Furthermore, it is difficult to guarantee the simultaneity of CSI measurements across distributed nodes, and CSI aging occurs during backhaul and processing, leading to precoding distortion that also significantly reduces JT gain.

3 C-WAN Architecture and its Enabling Role for JT

The aforementioned limitations indicate that addressing the implementation challenges of JT necessitates innovation at the network architecture level. A novel C-WAN architecture presents a viable solution to these challenges^[20].

3.1 C-WAN Architecture Overview

The innovation of C-WAN over the traditional FTTR architecture lies in its master-slave control and functional reconfiguration. As shown in Fig. 2b, this architecture comprises two core functional layers. The first is MFU (the central processing unit), which acts as the role of the traditional home gateway while also integrating baseband processing capabilities

and a centralized controller. All core signal processing (e.g., channel coding, precoding, scheduling) and coordination algorithms are completed within the MFU. The second layer is formed by SFUs (distributed radio units): The functionality of the SFUs deployed in various rooms is simplified, primarily handling radio frequency transmission and reception and simple PHY layer functions. They are connected to the MFU via a fiber fronthaul network.

This architecture transforms the traditional problem of distributed AP coordination into a centralized computing problem within a single central processing unit, thereby making high-precision JT feasible at the system level.

3.2 Solution of C-WAN to JT Synchronization Challenge

The C-WAN architecture addresses the issues of CFO and transmission timing synchronization at their root through the following mechanisms.

3.2.1 High-Precision Clock Distribution and Frequency Offset Elimination

C-WAN leverages the low-loss, high-stability physical characteristics of fiber links to distribute a unified reference clock from the MFU to all SFUs. Protocols like White Rabbit, which utilize precise hardware timestamps and delay compensation algorithms, can be employed to achieve sub-nanosecond clock synchronization accuracy^[21]. Measurements indicate that this scheme can achieve clock synchronization with a standard deviation below 100 ps in FTTR scenarios^[22].

Taking the 160 MHz channel bandwidth in the IEEE 802.11ax standard as an example, its sampling period is 6.25 ns. The synchronization error of C-WAN (100 ps) constitutes only about 1.6% of the sampling period, far below the typical symbol synchronization tolerance required for JT (usually <10%). This means the local oscillators of all SFUs can be effectively locked to the same frequency reference, eliminating CFO and laying a solid foundation for phase coherence^[23].

3.2.2 Centralized Scheduling and Joint Triggering

In the C-WAN architecture, the MFU acts as the sole scheduling center, possessing the capability to accurately measure and compensate for the fiber transmission delay of signals to each SFU. When JT transmission is required, the MFU issues joint trigger commands containing precise transmission timestamps to each SFU. Each SFU adjusts its transmission timing according to these instructions, thereby achieving symbol-level transmission synchronization and eliminating the phase deviation ($\Delta\phi$) caused by trigger randomness^[23].

3.3 C-WAN Support for Data and Computing Centralization

The value of C-WAN extends beyond solving the synchronization problem; it creates uniquely favorable conditions for deploying advanced algorithms such as deep learning to overcome the CSI overhead challenge. This advantage stems from three facets of centralization: 1) Data centralization: The MFU centralizes the raw In-Phase/Quadrature (IQ) or CSI data from all SFUs, naturally forming a global, consistent joint channel dataset. This provides the essential and unique data foundation for training deep learning models that require multi-AP data as input, which is difficult to obtain in a distributed architecture. 2) Computing power centralization: The MFU can integrate substantial computational resources, which are necessary for handling the computational load for deep learning model training and inference. 3) Control centralization: The centralized scheduler in the MFU may directly and rapidly issue unified precoding matrices and transmission instructions to all SFUs based on the prediction results of the deep learning model. This avoids the delays and inconsistencies associated with distributed decision-making, ensuring that the predicted CSI is applied timely and accurately.

On this basis, the next section will elaborate in detail on how deep learning can be leveraged within the C-WAN architecture to significantly reduce CSI acquisition overhead.

4 Deep Learning-Based Low-Overhead CSI Acquisition Method

Although the C-WAN architecture solves the JT synchronization challenge and aggregates computing and data resources at the MFU, adopting the traditional full-dimensional, periodic channel sounding method would still impose immense over-the-air overhead, severely constraining the effective throughput of the FTTR system.

Fortunately, the indoor FTTR C-WAN scenario provides three advantageous conditions for optimizing CSI acquisition. First, temporal correlation arises from the low mobility of STAs, leading to slow channel variation and small Doppler shifts. Historical CSI sequences contain the potential for predicting future CSI^[24]. Second, spatial correlation is inherent because the positions of SFUs are fixed after deployment, and the channel responses between their antenna arrays exhibit structural properties in the spatial domain.

This structure exists not only within individual SFUs but also between different SFUs. Therefore, the MFU may infer the channel state of all antennas using observations from a subset of antennas per SFU. Third, protocol adaptability allows the MFU to select clusters of antennas that contribute more significantly to channel reconstruction to participate in the sounding process, based on learned iterative weights. Since the C-WAN architecture virtualizes all SFUs as one large AP, this adjustment is transparent to the STA side, requires no additional signaling overhead, and offers excellent compatibility.

Leveraging the above advantages, this section proposes a deep learning-based low-overhead CSI acquisition framework that fully integrates with the FTTR C-WAN architecture^[25]. Its goal is to significantly extend the sounding period by utilizing historical CSI to predict future CSI, while maintaining JT performance. Furthermore, during sounding, it uses observations from a small number of antennas to reconstruct the CSI for all antennas, reducing the resource occupation per sounding instance.

4.1 Problem Definition

The traditional scheme requires full-dimensional sounding every T_{frame} , meaning all antennas across all SFUs participate in probing. Its overhead is:

$$\text{OH}_{\text{full}} \propto \frac{N_t \times K \times L_{\text{LTF}}}{T_{\text{frame}}} \quad (5),$$

where N_t denotes the number of sounding antennas per SFU, K is the total number of SFUs, and L_{LTF} represents the number of Long Training Field (LTF) symbols required per antenna. The objective of this paper is to reduce the overhead to:

$$\text{OH}_{\text{proposed}} = \frac{1}{T} \times \frac{1}{R} \times \text{OH}_{\text{full}} \quad (6),$$

where $T > 1$ is the sounding period extension factor, and $R = N_t/N_s > 1$ is the antenna compression ratio (N_s is the number of antennas actually participating in sounding), corresponding to the proportion of required HE-LTF symbols^[26]. Specifically, this paper addresses the following problem: Given the historical L_{frame} CSI sequence of all antennas $\hat{\mathbf{H}}(t-L+1), \dots, \hat{\mathbf{H}}(t)$ and the partial antenna observations $\mathbf{H}_{N_s}(t+1)$ at time instant $t+1$, the objective is to learn a mapping function $f(\cdot; \Theta)$ that accurately estimates the full-dimensional CSI at time $t+1$:

$$\begin{aligned} \hat{\mathbf{H}}(t+1) = f\left(\left\{\hat{\mathbf{H}}(t-L+1), \dots, \hat{\mathbf{H}}(t)\right\}, \right. \\ \left. \mathbf{H}_{N_s}(t+1); \Theta\right) \end{aligned} \quad (7).$$

4.2 Hybrid Deep Learning Model Design

To fully exploit the advantage of the C-WAN architecture in acquiring global CSI, this paper proposes a dual-branch hybrid deep learning model. It achieves high-precision estimation and reconstruction of full-dimensional CSI by synergistically utilizing the temporal and spatial correlation of the channel. This model enables accurate CSI estimation with reduced sounding overhead, thereby improving the overall throughput of JT transmission.

The structure of this hybrid model is shown in Fig. 3. It adopts a dual-branch parallel processing architecture, comprising three core components: a spatial reconstruction branch, a temporal prediction branch, and an adaptive fusion module. The outputs of the spatial reconstruction branch and the temporal prediction branch are fused using adaptive weighting, collaboratively leveraging the temporal and spatial correlations of the channel. Finally, the complete CSI matrix is output through convolutional layers, dimension reshaping, and complex number reconstruction, achieving high-precision

estimation of the full-dimensional CSI.

1) Spatial Reconstruction Branch: The purpose of this branch is to recover missing spatial information from the partial antenna observations. Leveraging the global channel information characteristic provided by the C-WAN architecture, the model selects the half of the antennas per SFU that contribute most to the reconstruction (based on the loss function) for sounding. This effectively reduces the dimensionality of the original CSI matrix. The original matrix, which encompasses all antennas, has dimensions $[N, \text{ANT}_x, \text{ANT}_y, \text{NUM}_{\text{tx}}, f_{\text{subcarrier}}]$. After reduction, the matrix contains only partial antenna observations, with dimensions $[N, \text{ANT}_x/2, \text{ANT}_y, \text{NUM}_{\text{tx}}, f_{\text{subcarrier}}]$. This reduction lowers the overhead of the sounding process. Here, N denotes the batch size; ANT_x and ANT_y are the number of antennas in the x - and y -directions of the array, respectively; NUM_{tx} is the number of transmitters, and $f_{\text{subcarrier}}$ is the number of OFDM subcarriers. After preprocessing steps such as complex number separation and spatial dimension flattening, the

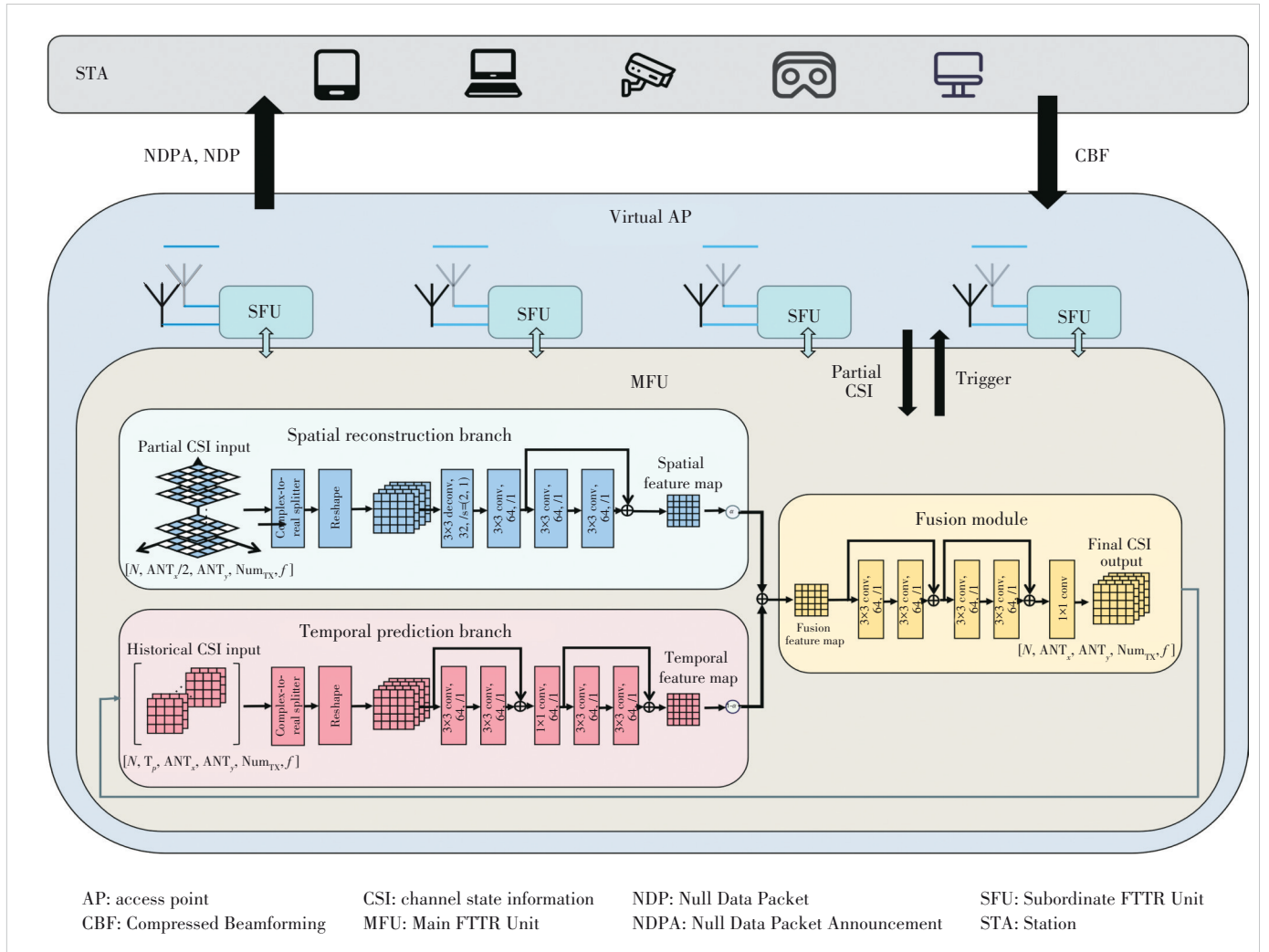


Figure 3. Hybrid model architecture that fuses CNN and ResNet

core of the perception branch lies in multi-level feature learning. The model employs a combination of deconvolutional layers, convolutional layers, and residual blocks for feature extraction, ultimately generating the Spatial Feature Map.

2) Temporal Prediction Branch: Targeting the characteristics of indoor FTTR scenarios where STA movement is slow, channel variation is continuous, and Doppler shift is low, this branch aims to predict the channel state by leveraging the temporal correlation of historical CSI sequences^[27]. The input to this branch is the historical complete CSI matrix spanning T_p time steps, with dimensions $[N \cdot T_p, \text{ANT}_x, \text{ANT}_y, \text{NUM}_{\text{tx}}, f_{\text{subcarrier}}]$. After data preprocessing including complex number separation and spatial dimension flattening, a feature representation is obtained. The core of the temporal prediction branch lies in multi-scale temporal modeling. The model employs two successive dilated residual blocks and a 1×1 convolutional block. This design allows the model to expand its temporal receptive field, enabling it to capture time dependencies at various scales. The output dimension of each dilated residual block remains consistent, ensuring uniformity in feature dimensions and ultimately generating the Temporal Feature Map. The introduction of dilated convolutions allows the model to gain richer temporal context information without a significant increase in the number of parameters, thus improving the accuracy of temporal prediction^[28].

3) Adaptive Fusion Module: To exploit the complementarity of spatial and temporal information, the model integrates an adaptive weighted fusion module. This module combines the outputs of the spatial reconstruction branch and the temporal prediction branch for more precise CSI estimation. The fusion formula is as follows:

$$\mathbf{F}_{\text{reconstructed}} = \alpha \cdot \mathbf{F}_{\text{spatial}} + (1 - \alpha) \cdot \mathbf{F}_{\text{temporal}} \quad (8),$$

where $\mathbf{F}_{\text{spatial}}$ represents the output features of the spatial reconstruction branch, $\mathbf{F}_{\text{temporal}}$ represents the output features of the temporal prediction branch, and α is a trainable weight parameter initialized to 0.5. Through the end-to-end training process, the model can adaptively learn the optimal weight ratio between spatial and temporal information. The fused features undergo further deep feature optimization through two additional residual blocks. Finally, the complete CSI matrix with dimensions $[N, \text{ANT}_x, \text{ANT}_y, \text{NUM}_{\text{tx}}, f]$ is output via 1×1 convolutional layers, dimension reshaping, and complex number reconstruction.

4.3 Loss Function Design

The optimization objective of this paper is to maximize the JT transmission performance, rather than directly minimize the Mean Squared Error (MSE) between the predicted CSI $\hat{\mathbf{H}}$ and the true CSI \mathbf{H} . Therefore, we design a Signal-to-Interference-plus-Noise Ratio (SINR)-oriented loss function based on the precoding matrix, constructed as follows.

Our design employs Zero-Forcing (ZF) precoding, a choice motivated by the generally high-SNR conditions inherent in FTTR deployments due to dense AP placement and short transmission distances. Under such propagation environments, ZF not only achieves near-optimal performance but also benefits model training through its sensitivity to channel inaccuracies.

The loss function construction proceeds as follows. First, based on the true CSI matrix \mathbf{H} and the predicted CSI matrix $\hat{\mathbf{H}}$, the precoding matrices are calculated using ZF precoding:

$$\mathbf{W} = \mathbf{H}^H (\mathbf{H} \mathbf{H}^H)^{-1} \quad (9),$$

$$\hat{\mathbf{W}} = \hat{\mathbf{H}}^H (\hat{\mathbf{H}} \hat{\mathbf{H}}^H)^{-1} \quad (10).$$

Subsequently, the predicted precoding matrix $\hat{\mathbf{W}}$ is used for signal transmission, and the received signal power and interference components are calculated. The useful received signal power is:

$$P_{\text{sig}} = \left\| \mathbf{H} \hat{\mathbf{W}} \right\|_F^2 \quad (11).$$

Due to the mismatch between the precoding matrix and the true channel, additional interference is introduced. The equivalent noise power P_{noise} includes the thermal noise power P_n (which can be estimated via Received Signal Strength Indicator (RSSI)) and the aforementioned interference power.

Finally, the loss function is defined as the negative value of the sum of achievable spectral efficiency across all subcarriers, aiming to optimize the overall throughput of the JT transmission. The SINR on the i -th subcarrier is:

$$\text{SINR}_i = \frac{P_{\text{sig}, i}}{P_{\text{noise}, i}} = \frac{\left\| \mathbf{H} \hat{\mathbf{W}} \right\|_{i,j}^2}{\sum_{j \neq i} \left\| \mathbf{H} \hat{\mathbf{W}} \right\|_{i,j}^2 + \sigma_n^2} \quad (12),$$

where σ_n^2 represents the thermal noise power, obtainable by measuring background noise power during packet intervals. The overall loss function is:

$$\mathcal{L} = -\frac{1}{N_{\text{sc}}} \sum_{i=1}^{N_{\text{sc}}} \log_2 (1 + \text{SINR}_i) \quad (13).$$

This loss function directly penalizes the SINR degradation caused by precoding mismatch due to CSI prediction errors, reflecting the fundamental requirements of the communication system for high throughput and reliable transmission.

4.4 Synergistic Integration of Hybrid Model Within FTTR C-WAN Architecture

As illustrated in Fig. 3, the proposed low-overhead sound-

ing process operates as follows: leveraging the C-WAN's logical abstraction of all SFUs as a single virtual AP, the MFU initiates a coordinated sounding process. It transmits an NDPA frame declaring a reduced number of antennas (a subset determined by the model's spatial reconstruction capability). Subsequently, the MFU uses its synchronized triggering mechanism to instruct the selected antennas across different SFUs to transmit NDP frames simultaneously. From the perspective of the STA, this process is indistinguishable from a standard sounding procedure; it responds with a Compressed Beamforming (CBF) report based on the declared antenna count, remaining unaware of the underlying antenna selection strategy.

Upon receiving the STA's feedback, the MFU, housing the centralized computational resources, executes the hybrid model. The model fuses the limited real-time observations from the partial antenna sounding with the rich historical channel information available in the centralized dataset. Through its temporal prediction and spatial reconstruction branches, it reconstructs the full-dimensional CSI for all antennas involved in the JT transmission. This end-to-end process, from coordinated partial sounding to centralized CSI reconstruction, achieves a significant reduction in over-the-air overhead while maintaining full backward compatibility with existing standards, as no changes are required at the STA side.

5 Experimental Results and Analysis

This section presents a comprehensive evaluation of the proposed C-WAN architecture and hybrid deep learning model for low-overhead JT in FTTR networks. The experiments are designed to validate the effectiveness of our solution in addressing both synchronization challenges and CSI acquisition overhead. Through systematic testing on a real-world synchronized channel dataset, we demonstrate the performance of our approach in terms of CSI estimation accuracy, spectral efficiency, and overall system throughput compared to baseline methods.

5.1 Channel Model and Dataset

In our experimental setup, which is based on the C-WAN architecture where the MFU acts as the central controller for acquiring downlink JT precoding CSI, we leverage the channel reciprocity of TDD systems^[16]. Since direct downlink CSI acquisition through STA feedback is complex and introduces significant overhead, we use uplink channel sounding measurements as a practical substitute to generate the training data for the proposed model.

This study utilizes a real-world channel dataset (espargos-0005^[5, 28]) for model development and validation. This dataset was collected using 4 distributed ESPARGOS antenna arrays (each with 8 antennas) deployed around the perimeter of a room, acting as receivers (simulating 4 SFUs). A moving transmit antenna (simulating a STA) transmitted Wi-Fi signals

within a 7×7×3 meter laboratory. Using a common external clock and synchronous triggering mechanism, all arrays synchronously collected CSI at a sampling rate of 100 Hz, simulating the synchronous multi-AP channel measurement environment of the C-WAN architecture. Thus, the channel of the k -th SFU at time t can be represented as a matrix $\mathbf{M} \in \mathbb{C}^{N_t \times N_{sc}}$, where the number of antennas per SFU $N_t = 8$ and the number of subcarriers $N_{sc} = 117$. The dataset contains CSI samples for 186 879 time instances. To strictly simulate the actual online prediction scenario, the data is split chronologically into a training set (70%), a validation set (15%), and a test set (15%).

Model performance is evaluated using two key metrics: the Normalized Mean Squared Error (NMSE) for assessing CSI estimation accuracy, and the spectral efficiency loss function (defined in Section 4) for quantifying the impact of predicted CSI on JT transmission performance. The NMSE is calculated between the predicted channel matrix $\hat{\mathbf{H}}$ and the true channel matrix \mathbf{H} , serving as a key metric to quantify the accuracy of

our CSI estimation, with the expression $\text{NMSE} = \frac{\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2}$.

In simulations, we select the following three baseline methods for comparison: 1) the Full Sounding (Oracle) Method, which assumes perfect and real-time full-dimensional CSI is available and thus represents the theoretical upper bound of JT transmission performance, used to measure the gap between the proposed method and optimal performance; 2) the Temporal Prediction Method, an ablated version of the fusion model that disables the spatial reconstruction branch (i.e., the $\mathbf{H}_{N_t}(t+1)$ input), and can only rely on historical information up to time $t-1$ to predict the channel at time t ; 3) the Spatial Reconstruction Branch, a counterpart ablation that retains only the spatial reconstruction branch while disabling the temporal prediction component, thereby estimating the full CSI solely from the partial antenna observations $\mathbf{H}_{N_t}(t+1)$ at the current time instant without leveraging any historical channel information.

5.2 Model Training Results Analysis

To validate the effectiveness of the proposed hybrid deep learning model, we trained the model using the Adam optimizer with a batch size of 64. Fig. 4a shows the change in NMSE on the training and test sets versus training iterations, while Fig. 4b shows the convergence curve of the spectral efficiency loss function \mathcal{L} . It can be observed that both loss functions decrease steadily with iterations, and the validation loss does not show a significant increase, indicating good generalization capability of the model. The spectral efficiency loss on the training and test sets eventually converges to similar values, further proving the strong adaptability and stability of the model.

As shown in Fig. 5a, we conducted experiments with varying numbers of sounding antennas ($N_s = 1, 2, 4, 6$) to evaluate

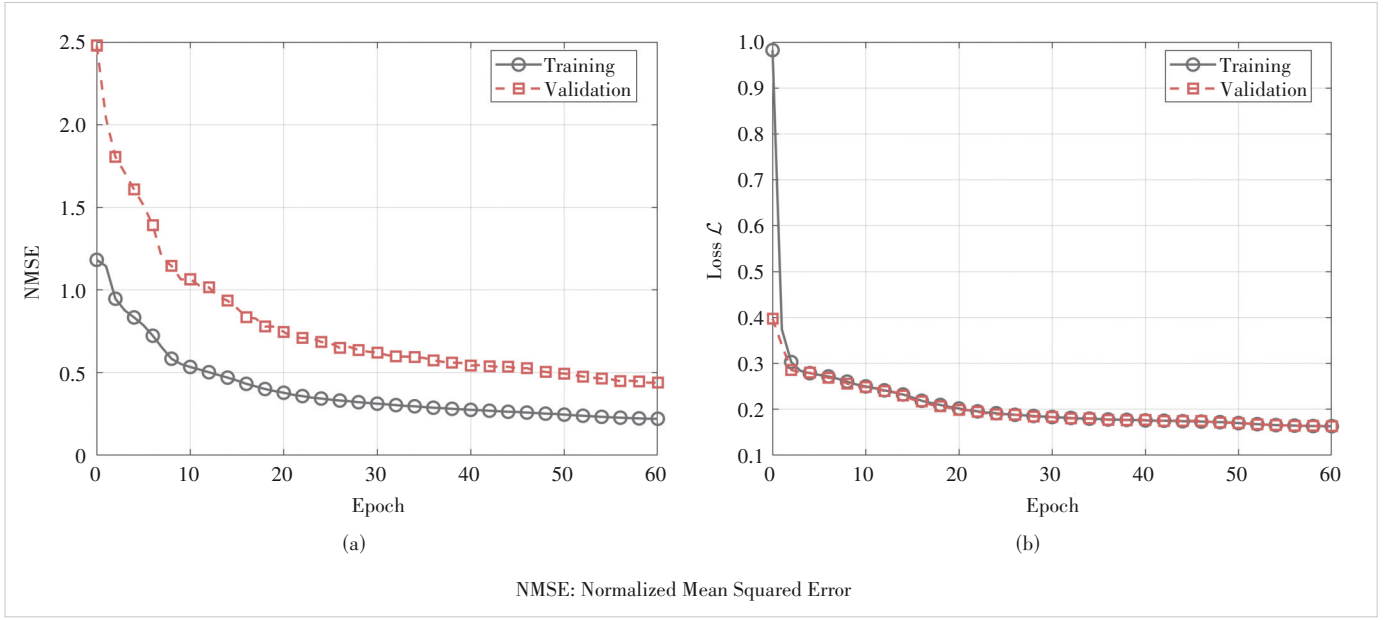


Figure 4. Loss convergence curves: (a) NMSE; (b) the spectral efficiency loss function \mathcal{L}

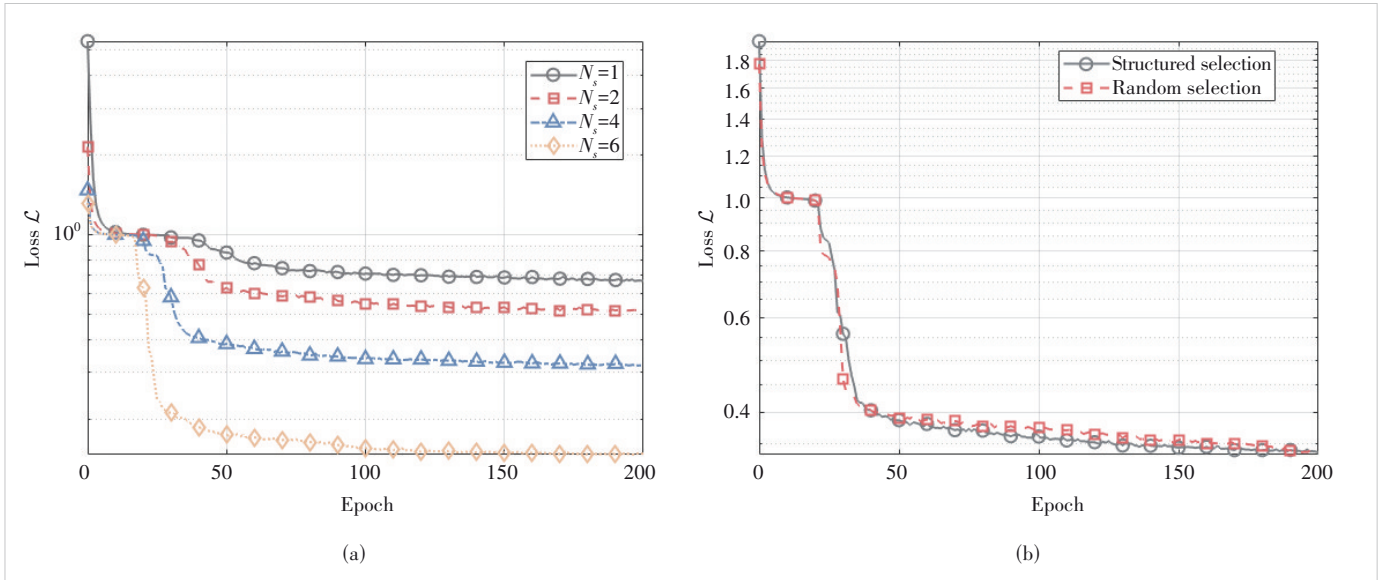


Figure 5. (a) Effect of N_s on the loss function; (b) Effect of N_s selection strategy on the loss function

their impact on CSI reconstruction performance. The results demonstrate that under the C-WAN architecture, the channel CSI exhibits significant spatial correlation, with larger N_s values yielding better prediction accuracy. In Fig. 5b, we further compare two antenna selection strategies: random selection and structured selection. Due to the close proximity of antenna arrays in the dataset, the difference in loss function convergence between these strategies was minimal. Considering the FTTR deployment scenario where APs are distributed across different rooms and antennas within the same AP possess stronger spatial correlation, we adopted $N_s = 4$ with a structured selection approach, which used one antenna from

each of the two antenna arrays per AP, for subsequent experiments. This configuration maintains a balance between reconstruction accuracy and practical deployment constraints.

Fig. 6a illustrates the evolution of trainable weights in the model's adaptive fusion module across training iterations, where the temporal prediction branch ultimately achieves an average weight of 0.98 after multiple iterations. This result indicates that in indoor low-mobility scenarios, channel variations exhibit strong temporal correlation. Relying solely on the historical CSI sequence is sufficient to make highly accurate estimates of the future channel state, while the corrective effect of the observations from the partial antennas at the cur-

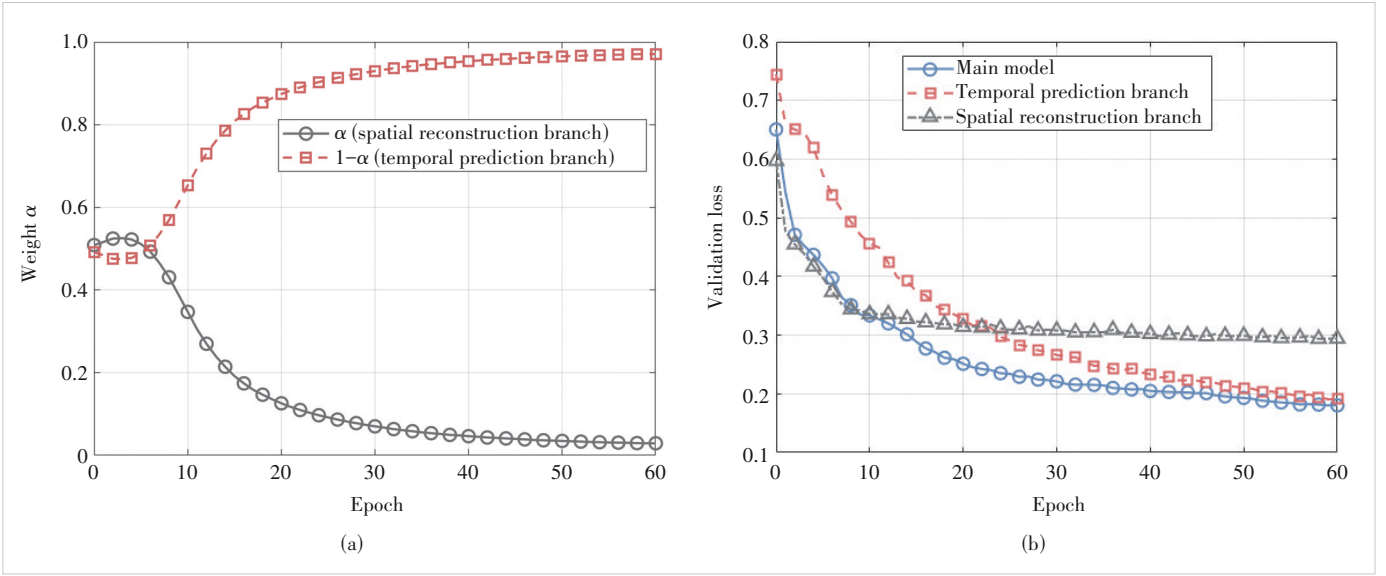


Figure 6. (a) Adaptive fusion weight α for the temporal branch and (b) performance comparison of model variants

rent moment is relatively limited. This finding empirically supports a new paradigm for channel acquisition dominated by prediction and assisted by sounding, suggesting that the sounding period can be significantly extended, thereby offering the potential for substantially reducing system overhead.

Although the temporal prediction branch dominates the model, the spatial reconstruction branch still provides indispensable value. To quantitatively evaluate its contribution, we conduct an ablation study comparing three model variants: the temporal-prediction-only model, the spatial-reconstruction-only model, and the full hybrid model with adaptive α . As illustrated in Fig. 6b, the inclusion of the spatial branch accelerates the convergence of the loss function and yields measurable performance gains. These results confirm the unique role of the spatial reconstruction branch in feature enhancement and training stabilization.

5.3 System Performance Evaluation

To further evaluate the practical effectiveness of the proposed method in the JT transmission scenario, we compared the Spectral Efficiency (SE) achievable by the system after employing the predicted CSI for precoding. Fig. 7 shows the Cumulative Distribution Function (CDF) curves of the SE on the test set for the proposed hybrid model and the two baseline methods.

The results show that the Oracle scheme with complete real-time CSI achieves the optimal spectral efficiency, representing the theoretical upper bound for JT transmission performance. The SE curve of the proposed hybrid model is the closest to the Oracle performance among all practical (non-oracle) methods, with a median SE gap of less than 5%, indicating that the hybrid model can effectively maintain JT gain even with low sounding overhead.

Notably, the simplified model with only the temporal prediction branch (i.e., ablated spatial reconstruction branch) performs the worst. This further confirms the important role of the spatial reconstruction branch in compensating for instantaneous channel information and improving the completeness of CSI estimation.

6 Conclusions

This paper addresses the critical challenges associated with implementing efficient JT in FTTR networks. To overcome physical layer synchronization issues, the proposed C-WAN architecture leverages a fiber fronthaul network to provide

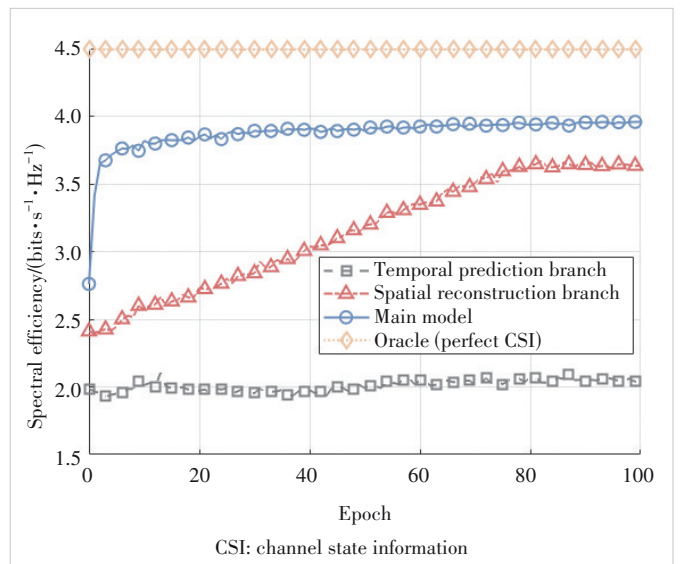


Figure 7. Cumulative Distribution Function (CDF) of achievable spectral efficiency for different CSI acquisition methods

high-precision clock distribution and coordinated transmission triggering. This approach effectively eliminates carrier frequency offsets and timing deviations, establishing a robust PHY-layer foundation for JT. To mitigate the overhead of CSI acquisition, we develop a hybrid deep learning model that utilizes the centralized data, computation, and control capabilities of the C-WAN architecture, along with inherent spatiotemporal correlations in indoor channels. By integrating temporal prediction and spatial reconstruction branches, the model achieves highly accurate CSI estimates while drastically reducing the need for frequent sounding. Experimental results based on a real-world synchronized channel dataset demonstrate that the proposed solution significantly reduces over-the-air resource consumption while maintaining JT performance close to that achieved with ideal CSI. This work offers a practical and efficient framework for high-performance coordinated transmission in next-generation FTTR systems. Future work will explore transfer learning and domain adaptation techniques to fine-tune or train the model on test sets of varying scenario scales, thereby enhancing its generalization capability across diverse deployment environments. Additionally, assessing the feasibility and effectiveness of deploying our algorithm on hardware platforms will also be considered as part of future work.

References

- [1] ZHANG D Z, ZENG T, YANG Z Y, et al. FTTR-B technology exploration and practice [C]//Proc. IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2024: 37 – 41. DOI: 10.1109/ICC-Workshops59551.2024.10615937
- [2] NUNEZ D, SMITH M, BELLALTA B. Multi-AP coordinated spatial reuse for Wi-Fi 8: group creation and scheduling [C]//Proc. 21st Mediterranean Communication and Computer Networking Conference (MedComNet). IEEE, 2023: 203 – 208. DOI: 10.1109/MedComNet58619.2023.10168857
- [3] SHEN G X, LI J, CAI J H, et al. Enhancing fiber-to-the-room (FTTR) technologies: addressing key challenges and solutions (invited tutorial) [C]//Proc. Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR). IEEE, 2024: 1 – 2. DOI: 10.1109/CLEO-PR60912.2024.10676562
- [4] SUNDARAVARADHAN S P, PORAT R, TOUSSI K N. Increasing spatial multiplexing gain in future multi-AP Wi-Fi systems via joint transmission [J]. IEEE communications standards magazine, 2022, 6(2): 20 – 26. DOI: 10.1109/MCOMSTD.0001.2100085
- [5] CHEN B H, JIAO X J, LIU W, et al. An experimental study of Wi-Fi joint transmission with multiple openwifi access points [C]//IEEE Wireless Communications Networking Conference (WCNC). IEEE, 2025: 1 – 67. DOI: 10.1109/WCNC61545.2025.10978612
- [6] HAMED E, RAHUL H, ABDELGHANY M A, et al. Real-time distributed MIMO systems [C]//Proc. 2016 ACM SIGCOMM Conference (SIGCOMM '16). Association for Computing Machinery, 2016: 412 – 425. DOI: 10.1145/2934872.2934905
- [7] TITUS A, BANSAL R, SREEJITH T V, et al. Decision problems for joint transmission in multi-AP coordination framework of IEEE 802.11be [C]//Proc. International Conference on Communication Systems & Networks (COMSNETS). IEEE, 2021: 326 – 333. DOI: 10.1109/COMSNETS51098.2021.9352818
- [8] LEVINBOOK Y, EZRI D, MELZER E. AP cooperation in Wi-Fi: joint transmission with a novel precoding scheme, resilient to phase offsets between transmitters [J]. Signal processing, 2024, 220: 109432. DOI: 10.1016/j.sigpro.2024.109432
- [9] KUNNATH GANESAN U, SARVENDRANATH R, LARSSON E G. Beam-Sync: over-the-air synchronization for distributed massive MIMO systems [J]. IEEE transactions on wireless communications, 2024, 23(7): 6824 – 6837. DOI: 10.1109/TWC.2023.3335089
- [10] HE J L, YAO H, ZHAO D S, et al. Multi-channel phase synchronization method for independent asynchronous local oscillators based on pre-calibration matrix [C]//Proc. IEEE International Symposium on Antennas and Propagation and INC/USNC – URSI Radio Science Meeting (AP-S/INC-USNC-URSI). IEEE, 2024: 449 – 450. DOI: 10.1109/AP-S/INC-USNC-URSI52054.2024.10687303
- [11] VERMA S, RODRIGUES T K, KAWAMOTO Y, et al. A survey on multi-AP coordination approaches over emerging WLANs: future directions and open challenges [J]. IEEE communications surveys & tutorials, 2024, 26(2): 858 – 889. DOI: 10.1109/COMST.2023.3344167
- [12] LIU Z Y, ZHANG L, DING Z. Overcoming the channel estimation barrier in massive MIMO communication via deep learning [J]. IEEE wireless communications, 2020, 27(5): 104 – 111. DOI: 10.1109/MWC.001.1900413
- [13] JIANG W, SCHOTTEN H D. Deep learning for fading channel prediction [J]. IEEE open journal of the communications society, 2020, 1: 320 – 332. DOI: 10.1109/OJCOMS.2020.2982513
- [14] CHEN W, WAN W X, WANG S Y, et al. CSI-PPNet: a one-sided one-for-all deep learning framework for massive MIMO CSI feedback [J]. IEEE transactions on wireless communications, 2024, 23(7): 7599 – 7611. DOI: 10.1109/TWC.2023.3342735
- [15] BANERJEE B, ELLIOTT R C, KRZYMIEŃ W A, et al. Machine learning assisted DL CSI estimation for high-mobility multi-antenna users with partial UL CSI availability in TDD massive MIMO systems [C]//Proc. IEEE Globecom Workshops (GC Wkshps). IEEE, 2022: 1579 – 1585. DOI: 10.1109/GCWkshps56602.2022.10008644
- [16] WANG J Z, LIU W C, MENG J, et al. Fiber to the radio/C-WAN architecture and its performance analysis [C]//Proc. IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2024: 709 – 713. DOI: 10.1109/ICCWorkshops59551.2024.10615771
- [17] ZHENG J K, ZHANG J Y, BJÖRNSON E, et al. Impact of channel aging on cell-free massive MIMO over spatially correlated channels [J]. IEEE transactions on wireless communications, 2021, 20(10): 6451 – 6466. DOI: 10.1109/TWC.2021.3074421
- [18] EUCHNER F, KELLNER D, STEPHAN P, et al. CSI dataset espargos-0005: four phase- and time-synchronous ESPARGOS antenna arrays in a lab room [EB/OL]. [2020-05-01]. <https://darus.uni-stuttgart.de/dataset.xhtml?persistentId=doi:10.18419/DARUS-4754>
- [19] VIEIRA J, RUSEK F, TUFVESSON F. Reciprocity calibration methods for massive MIMO based on antenna coupling [C]//2014 IEEE Global Communications Conference. IEEE, 2014: 3708 – 3712. DOI: 10.1109/GLOCOM.2014.7037384.
- [20] WU X M, ZENG Y, SI X S, et al. Fiber-to-the-room (FTTR): standards and deployments [C]//Proc. Optical Fiber Communication Conference (OFC) 2023. Optica Publishing Group, 2023: 1 – 3. DOI: 10.1364/ofc.2023.tu3f.6
- [21] JIMÉNEZ LÓPEZ M, GUTIÉRREZ RIVAS J L, DÍAZ ALONSO J. A white-rabbit network interface card for synchronized sensor networks [C]//Proc. SENSORS, 2014 IEEE. IEEE, 2014: 2000 – 2003. DOI: 10.1109/ICSENS.2014.6985426
- [22] XU Y L, RAJAGOPALA A D, FRUITWALA N, et al. Multi-FPGA synchronization and data communication for quantum control and measurement [EB/OL]. [2025-06-11]. <https://arxiv.org/abs/2506.09856>
- [23] BALAN H V, ROGALIN R, MICHALOLIAKOS A, et al. AirSync: enabling distributed multiuser MIMO with full spatial multiplexing [J]. IEEE/ACM transactions on networking, 2013, 21(6): 1681 – 1695. DOI:

- 10.1109/TNET.2012.2230449
- [24] LIU X, LI J W, WU X M, et al. Fiber-to-the-room (FTTR) technologies for the 5th generation fixed network (F5G) and beyond [C]//Proc. IEEE Future Networks World Forum (FNWF). IEEE, 2022: 351 – 354. DOI: 10.1109/FNWF55208.2022.00068
- [25] LI J L, ZHANG Q, XIN X J, et al. Deep learning-based massive MIMO CSI feedback [C]//18th International Conference on Optical Communications and Networks (ICOCN). IEEE, 2019: 1 – 3. DOI: 10.1109/ICOCN.2019.8934725
- [26] NAKAMURA T, BOUAZIZI M, YAMAMOTO K, et al. Wi-Fi-based fall detection using spectrogram image of channel state information [J]. IEEE Internet of Things journal, 2022, 9(18): 17220 – 17234. DOI: 10.1109/IJOT.2022.3152315
- [27] SEGUEL F, SALIHU D, HAEGELE S, et al. Complex-valued deep learning for WiFi-based indoor positioning: method and performance [C]//European Wireless 2024; 29th European Wireless Conference. VDE, 2024: 59 – 65
- [28] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. [2018-04-19]. <https://arxiv.org/abs/1803.01271>

Biographies

ZHANG Yang received his BE degree in communication engineering from Sun Yat-sen University, China in 2023. He is currently pursuing the MS degree in integrated circuit engineering at Sun Yat-sen University. His research inter-

ests include multi-AP coordination and distributed MIMO technologies.

CEN Zihan received his BE degree in electronic information engineering from Central South University, China in 2024. He is currently pursuing the MS degree in communication engineering at Sun Yat-sen University. His research interests include artificial intelligence and AI-RAN.

ZHAN Wen (zhanw6@mail.sysu.edu.cn) received his BS and MS degrees from the University of Electronic Science and Technology of China in 2012 and 2015, respectively. He obtained his PhD from the City University of Hong Kong, China in 2019, where he later worked as a research assistant and a post-doctoral fellow. Since 2020, he has been with the School of Electronics and Communication Engineering, Sun Yat-sen University, China, where he is currently an Associate Professor. His research interests include Internet of Things, modeling, and performance optimization of next-generation mobile communication systems.

CHEN Xiang received his BE and PhD degrees from the Department of Electronic Engineering, Tsinghua University, China in 2002 and 2008, respectively. From July 2008 to December 2014, he was with the Wireless and Mobile Communication Technology Research and Development Center (Wireless Center) and the Aerospace Center, Tsinghua University. In July 2005 and from September 2006 to April 2007, he visited NTT DoCoMo Research and Development (YRP), and the Wireless Communications and Signal Processing (WC-SP) Laboratory, Taiwan Tsing Hua University, China. Since January 2015, he has been with the School of Electronics and Information Technology, Sun Yat-sen University, where he is currently a Full Professor. His research interests include 5G/6G wireless and mobile communication networks and the Internet of Things (IoT).



Empowering Grounding DINO with MoE: An End-to-End Framework for Cross-Domain Few-Shot Object Detection

DONG Xiugang, ZHANG Kaijin, NONG Qingpeng,
JU Minhan, TU Yaofeng

(Nanjing R&D Center, ZTE Corporation, Nanjing 210012, China)

DOI: 10.12142/ZTECOM.202504009

<https://kns.cnki.net/kcms/detail/34.1294.TN.20250915.1501.002.html>,
published online September 15, 2025

Manuscript received: 2025-08-15

Abstract: Open-set object detectors, as exemplified by Grounding DINO, have attracted significant attention due to their remarkable performance on in-domain datasets like Common Objects in Context (COCO) after only few-shot fine-tuning. However, their generalization capabilities in cross-domain scenarios remain substantially inferior to their in-domain few-shot performance. Prior work on fine-tuning Grounding DINO for cross-domain few-shot object detection has primarily focused on data augmentation, leaving broader systemic optimizations unexplored. To bridge this gap, we propose a comprehensive end-to-end fine-tuning framework specifically designed to optimize Grounding DINO for cross-domain few-shot scenarios. In addition, we propose Mixture-of-Experts (MoE)-Grounding DINO, a novel architecture that integrates the MoE architecture to enhance adaptability in cross-domain settings. Our approach demonstrates a significant 15.4 Mean Average Precision (mAP) improvement over the Grounding DINO baseline on the Roboflow20-VL benchmark, establishing a new state of the art for cross-domain few-shot object detection (CD-FSOD). The source code and models will be made available upon publication.

Keywords: cross-domain few-shot object detection; Grounding DINO; Mixture-of-Experts; open-set object detection; pseudo-labeling

Citation (Format 1): DONG X G, ZHANG K J, NONG Q P, et al. Empowering Grounding DINO with MoE: an end-to-end framework for cross-domain few-shot object detection [J]. *ZTE Communications*, 2025, 23(4): 77 – 85. DOI: 10.12142/ZTECOM.202504009

Citation (Format 2): X. G. Dong, K. J. Zhang, Q. P. Nong, et al., “Empowering Grounding DINO with MoE: an end-to-end framework for cross-domain few-shot object detection,” *ZTE Communications*, vol. 23, no. 4, pp. 77 – 85, Sept. 2025. doi: 10.12142/ZTECOM.202504009.

1 Introduction

Open-set object detectors, such as Grounding DINO^[1] and Grounding DINO 1.5^[2], have attracted significant attention due to their ability to detect arbitrary objects described by natural language, even those unseen during pre-training. Leveraging vast internet-scale pre-training data, these models exhibit compelling performance on numerous in-domain datasets, such as Common Objects in Context (COCO)^[3], often requiring only minimal few-shot fine-tuning to adapt to specific tasks. This capability makes them highly promising for real-world applications where exhaustive annotation is impractical.

Despite their impressive performance in in-domain scenarios, the generalization capabilities of these detectors in cross-domain settings remain a substantial challenge^[4–5]. Cross-domain few-shot object detection (CD-FSOD) aims to recognize and localize novel objects in target domains that exhibit significant discrepancies from the source domain (i.e., the internet-scale pre-training data), using only a limited number of annotated examples. The inherent domain shift,

coupled with the severe data scarcity, leads to considerable and often prohibitive performance degradation for models like Grounding DINO. This substantial drop in accuracy and reliability critically undermines their practical utility in real-world applications where target data diverges from pre-training data and the available annotations are limited, such as autonomous systems, medical diagnosis, or remote sensing. Prior work^[4,6] on adapting Grounding DINO for CD-FSOD has mainly focused on isolated techniques such as data augmentation, leaving broader systemic optimizations largely unexplored.

To address this critical gap, our work introduces the first comprehensive end-to-end fine-tuning framework specifically designed to optimize Grounding DINO for CD-FSOD. This novel framework covers a holistic pipeline, spanning from dataset construction optimization to multi-stage model training, ensuring maximal data utilization and robust adaptation.

A central and pioneering contribution of this framework is the Mixture-of-Experts (MoE)-Grounding DINO, a novel ar-

chitecture that integrates the MoE^[7] module. While MoE has been widely adopted in large language models (LLMs) such as Mistral^[8] and DeepSeek-V2^[9], its application to open-set object detection, particularly for CD-FSOD, remains unexplored. MoE allows a model to selectively activate different “expert” sub-networks based on the input, effectively increasing the model capacity^[10]. We argue that this architecture is particularly well-suited for CD-FSOD since it provides a flexible mechanism to learn more diverse representations, allowing different experts to specialize in varied visual characteristics from the limited cross-domain training data.

Experimental results demonstrate the efficacy of our proposed framework. Notably, our approach achieves a significant 15.4 Mean Average Precision (mAP) improvement over the standard Grounding DINO baseline on the challenging Roboflow20-VL^[11] benchmark, establishing a new state-of-the-art model for CD-FSOD. Our contributions can be summarized as follows:

- 1) We propose the first end-to-end framework specifically designed for the task of CD-FSOD;
- 2) We are the first to integrate the MoE architecture into the domain of open-set object detection;
- 3) Our framework establishes a new state-of-the-art model on the Roboflow20-VL benchmark by fine-tuning Grounding DINO for CD-FSOD.

2 Related Work

2.1 Open-Set Object Detection

Compared to traditional object detection models that can only detect objects from a pre-defined set of categories, open-set object detection detects objects based on arbitrary user-provided textual category descriptions. Among influential approaches, Grounded Language-Image Pre-Training (GLIP)^[12] demonstrates particularly effective performance by formulating object detection as a visual grounding task, aligning region embeddings with text embeddings. RegionCLIP^[13] enhances open-set object detection by first generating region proposals via a Region Proposal Network (RPN) and then comparing similarities between region and text embeddings encoded by Contrastive Language-Image Pre-Training (CLIP)^[14].

Grounding DINO^[11] is an open-set object detector built on the DINO^[15] architecture. It consists of an image encoder and a text encoder, followed by a transformer-based feature enhancer and cross-modality decoder to effectively align visual and textual modalities. Grounding DINO is pre-trained on large-scale detection and grounding datasets, including Objects365^[16], Graphical Question Answering (GQA)^[17], etc. It achieves a zero-shot performance of over 53.0 mAP on the COCO benchmark. Building upon Grounding DINO, Grounding DINO 1.5^[2] further advances the model and becomes the state-of-the-art model by leveraging larger image backbones and over 20 million pre-training data. However, since

Grounding DINO 1.5 is a closed-source model, we use Grounding DINO as an alternative in our experiments.

2.2 Few-Shot Object Detection

Few-shot object detection methods can generally be categorized into meta-learning-based approaches and transfer-learning-based approaches.

Meta-learning-based approaches learn class prototype representations for each base category and infer novel categories by aligning regions of interest (RoI) with prototype representations. For example, Meta Region-Based Convolutional Neural Networks (R-CNN)^[18] conducted meta-learning over RoI features, using a support branch to create a category attention vector, which is then fused with RoI features for object detection. Ref. [19] leveraged LLMs to perform few-shot adaptation, relying on prototypes generated from DINO v2^[20] and region proposals extracted using Deformable Detection Transformer (DETR)^[21]. The two-stage fine-tuning approach (TFA)^[22] proposed a two-phase fine-tuning method based on Faster R-CNN^[23], surpassing prior meta-learning-based approaches by freezing the trained base class parameters and fine-tuning only the detection heads.

On the other hand, transfer-learning-based approaches focus on fine-tuning pre-trained object detectors using few-shot data. For example, Ref. [24] fine-tuned Grounding DINO to the agricultural domain in a few-shot setting by replacing the text encoder with randomly initialized trainable text embeddings.

CD-FSOD was initially proposed by Refs. [4] and [5]. They highlighted the importance of evaluating and enhancing the efficacy of pre-trained object detectors across varied domains that are rarely encountered in conventional internet-scale object detection pre-training datasets. Ref. [4] additionally outlined a couple of enhancements for adapting open-set object detectors (such as Grounding DINO and Detic^[25]) for CD-FSOD, including prompt engineering, federated fine-tuning, and multi-modal prompting. Concurrently, Ref. [6] focused on optimizing data augmentation techniques when fine-tuning Grounding DINO for CD-FSOD.

Despite these efforts, to the best of our knowledge, no prior research has focused on developing a comprehensive end-to-end framework tailored for fine-tuning Grounding DINO specifically within CD-FSOD scenarios.

2.3 Mixture-of-Experts

The core concept of the MoE architecture is to enable different components (experts) of a model to specialize in different aspects of the data^[10]. In recent years, MoE has gained significant popularity in LLMs, including Mistral^[8] and DeepSeek-V2^[9]. A common application of MoE in such models is to replace the traditional feed-forward network (FFN) with an MoE-based variant, as seen in Switch Transformer^[26] and Open-MoE^[27]. Each MoE layer consists of multiple ex-

perts, with only a subset activated for a given input. This selection is governed by a gating function (router), which dynamically routes different inputs to different experts. While MoE has been extensively explored in LLMs, to the best of our knowledge, we are the first to incorporate it into open-set object detection.

3 Method

3.1 End-to-End Framework for Cross-Domain Few-Shot Object Detection

In this work, we present a novel end-to-end framework that leverages the pre-trained Grounding DINO to address the challenge of data scarcity specifically in CD-FSOD. To the best of our knowledge, this is the first end-to-end framework for CD-FSOD. Our framework consists of three key stages, as conceptually illustrated in Fig. 1.

Firstly, an LLM-based text prompt optimization module is used. Given the raw category descriptions from the data, this module utilizes an LLM to generate more representative and effective text prompts, which are crucial for guiding Grounding DINO to better recognize novel concepts.

The second stage aims to adapt Grounding DINO to the target domain. In the CD-FSOD setting, since few bounding box annotations are available per category, not all target objects within an image are labeled. Therefore, we use a pseudo-labeling strategy to maximize the data utilization. This process generates additional training signals to enhance the fine-tuning performance of Grounding DINO.

In the last stage, we fine-tune our proposed MoE-Grounding DINO. This novel architecture is initialized from the Grounding DINO adapted in the second stage, and is further trained to substantially boost the overall model performance in the challenging CD-FSOD setting.

The detailed methodologies for each stage of this framework will be elaborated in subsequent sections.

3.2 LLM-Based Text Prompt Optimization

Grounding DINO leverages and aligns text and image modalities for object detection tasks. Its robust performance is attributed to pre-training on extensive internet-scale datasets,

leading to a well-established alignment between text and image modalities. Therefore, when fine-tuning it on a downstream dataset, the quality of the text prompts (i.e., category descriptions here) becomes crucial. This is especially critical in CD-FSOD scenarios, where data scarcity limits the model's adaptation. In addition, an optimal alignment between the text prompts and Grounding DINO's pre-training data is also essential for maximizing the fine-tuning performance.

However, the quality of text prompts in downstream datasets is often a significant hurdle. For instance, in the actions-zzid2-zb1hq-fsod-amih dataset from the Roboflow20-VL^[11] benchmark, the class "Attack" is described as "Players hit the ball over the net", and the class "Set" is described as "Players push the ball upwards with their fingertips". Such vague or overly simplistic prompts can make it difficult for Grounding DINO to learn robust representations, potentially leading to slow convergence or suboptimal performance during fine-tuning. Consequently, optimizing these text prompts before they are fed into Grounding DINO for fine-tuning is essential for achieving desirable results.

To address this, as illustrated in Fig. 2, we propose a three-step methodology that leverages an LLM, specifically Qwen2.5-VL^[28], to optimize category descriptions. In the first step, Qwen2.5-VL is prompted with images from the dataset to generate concise scenario descriptions that capture the essence of the dataset effectively. In the second step, Qwen2.5-VL is tasked with generating detailed category descriptions for each class. In detail, given an image with a target object enclosed within a red bounding box, Qwen2.5-VL is prompted to produce multiple informative, representative and distinctive descriptions tailored to the target object based on the original category description and the dataset description generated in Step 1. In Step 3, we craft the ideal text prompt for each class. For each category, we fine-tune Grounding DINO with randomly combined textual descriptions generated in Stage 2, aiming to identify the most effective combination that serves as the optimal text prompt for that particular class.

3.3 MoE-Grounding DINO

In few-shot object detection, maximizing the object detector's ability to extract supervision from scarce data is crucial.

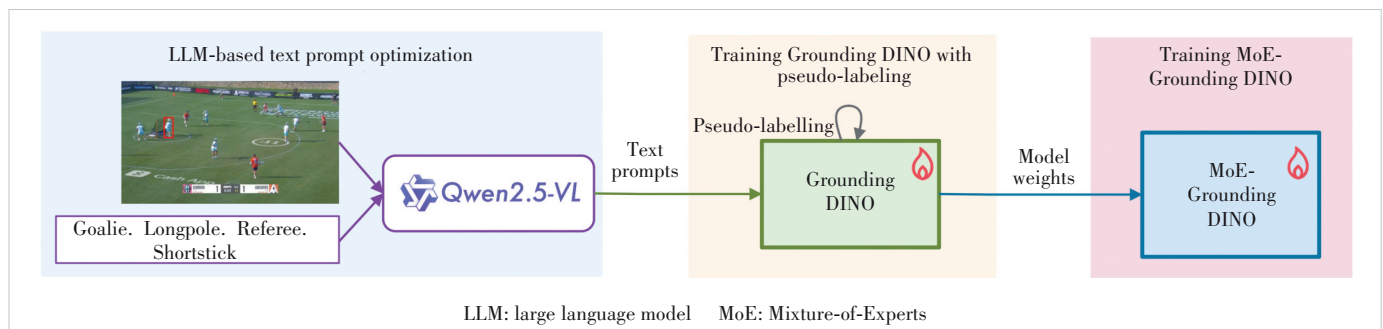


Figure 1. An illustration of the proposed end-to-end framework for cross-domain few-shot object detection (CD-FSOD)

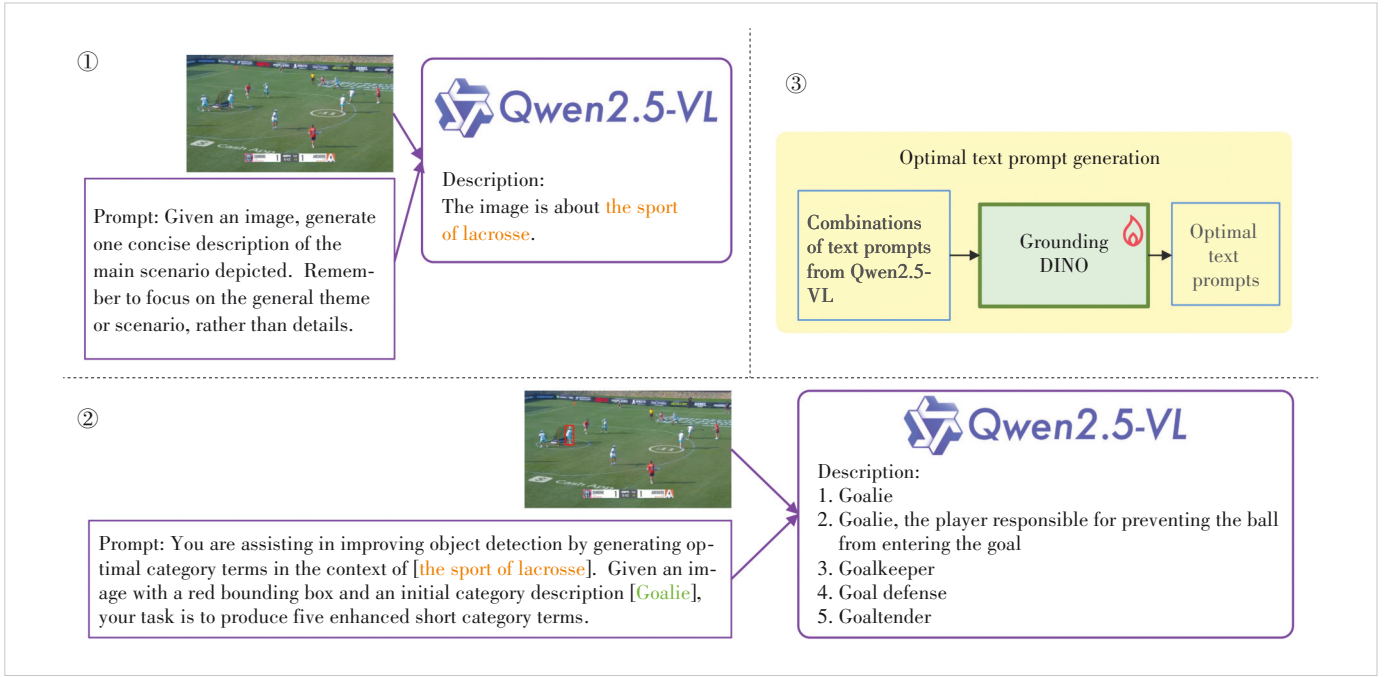


Figure 2. An example of the LLM-based text prompt optimization stage

In CD-FSOD, beyond the few-shot constraint, the cross-domain setting further increases the difficulty, as detectors usually require additional supervision to effectively adapt to a new domain. Therefore, success in CD-FSOD hinges on maximizing supervision and enabling the model to learn as much as possible from the limited training data.

The MoE architecture has demonstrated significant efficacy across LLMs. MoE functions by replicating individual experts multiple times and employs a dynamic routing mechanism. This mechanism enables different experts to specialize in distinct aspects of the input data, thereby empowering the model to capture more diverse and informative representations. It intuitively provides models with a broader parameter search space for learning, making it particularly well-suited for the CD-FSOD scenario.

We introduce MoE-Grounding DINO, a novel architecture built upon Grounding DINO. As depicted in Fig. 3, it incorporates the MoE module into Grounding DINO by substituting the FFN layers of the cross-modality decoder with MoE-FFN layers. The MoE-FFN layer, as detailed in Fig. 4, is designed to consist of a shared FFN and N routed FFNs (i.e., experts). For

a given input, the router dynamically determines K routed FFNs to activate. The outputs of the shared FFN and K activated routed FFNs are then aggregated to form the output of the MoE-FFN layer. This process can be expressed as:

$$\text{MoE_FFN}(x; \theta, \{\text{FFN}_{-r}\}_{i=1}^N, \text{FFN}_{-s}) = \text{FFN}_{-s}(x) + \sum_{j=1}^K \vartheta(x; \theta)_j f_j(x; \text{FFN}_{-r_j}) \quad (1),$$

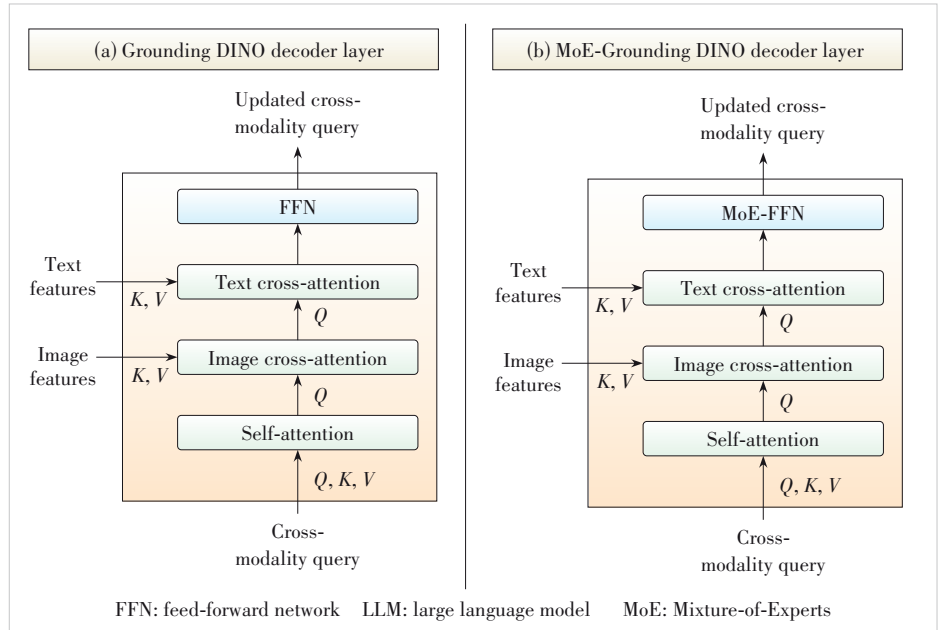


Figure 3. An illustration of the MoE-Grounding DINO decoder layer

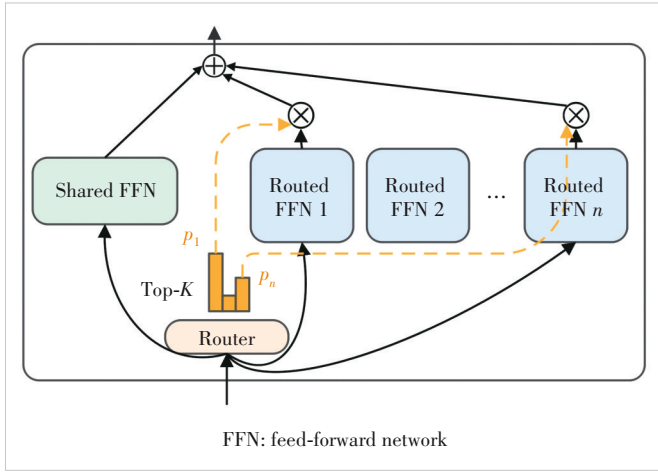


Figure 4. An illustration of the MoE-FFN layer

$$\vartheta(x; \theta)_j = \text{softmax}(\text{TopK}(g(x; \theta) + R_{\text{noise}}, K))_j \quad (2)$$

$$f_j(x; \text{FFN}_r)_j = \text{FFN}_r(\text{Dispatcher}(x)_j) \quad (3)$$

where FFN_r denotes individual routed FFNs, FFN_s refers to the shared FFN, $g(x; \theta)$ represents the routing function, R_{noise} is the noise term, and Dispatcher refers to the operation that dispatches the input to each expert.

Following Ref. [10], an additional auxiliary load balancing loss is applied during training. This loss is crucial for encouraging a more balanced utilization of the experts, preventing certain experts from dominating the model's learning.

The MoE-FFN module is seamlessly integrated into Grounding DINO without the need for retraining, as it can be initialized from the weights of Grounding DINO. This initialization process will be detailed later.

3.4 Training Strategy

Our proposed end-to-end framework (Fig. 3), is structured into three distinct stages: LLM-based text prompt optimization, Grounding DINO fine-tuning with pseudo-labeling, and MoE-Grounding DINO fine-tuning. This section elaborates on the methodologies and procedures employed for model training in Stages 2 and 3.

3.4.1 Grounding DINO Fine-Tuning with Pseudo-Labeling

The second stage is dedicated to fine-tuning the pre-trained Grounding DINO, augmented by a pseudo-labeling strategy. In CD-FSOD, only a limited number of bounding box annotations are typically provided per category. This often means that not all target objects within an image are initially labeled. To maximize the utilization of available data and compensate for annotation sparsity, we apply a simple pseudo-labeling approach to enhance Grounding DINO's performance.

As illustrated in Fig. 5, we first fine-tune Grounding DINO on the provided limited labeled training data. Subsequently, the fine-tuned model is employed to infer predictions on training instances. We only retain high-confidence predictions to ensure the quality of these inferred labels. This is followed by rigorous post-processing steps, such as Non-Maximum Suppression (NMS), to refine the generated pseudo-labels. The original ground-truth annotations are then combined with these high-confidence pseudo-labels to construct a refined training dataset. Finally, the fine-tuned Grounding DINO undergoes further training on this newly constructed refined dataset, leveraging the increased data volume to enhance its robustness and generalization.

3.4.2 MoE-Grounding DINO Fine-Tuning

The third stage involves fine-tuning our proposed MoE-Grounding DINO. For initialization, all components of MoE-Grounding DINO inherit the parameters from the Grounding

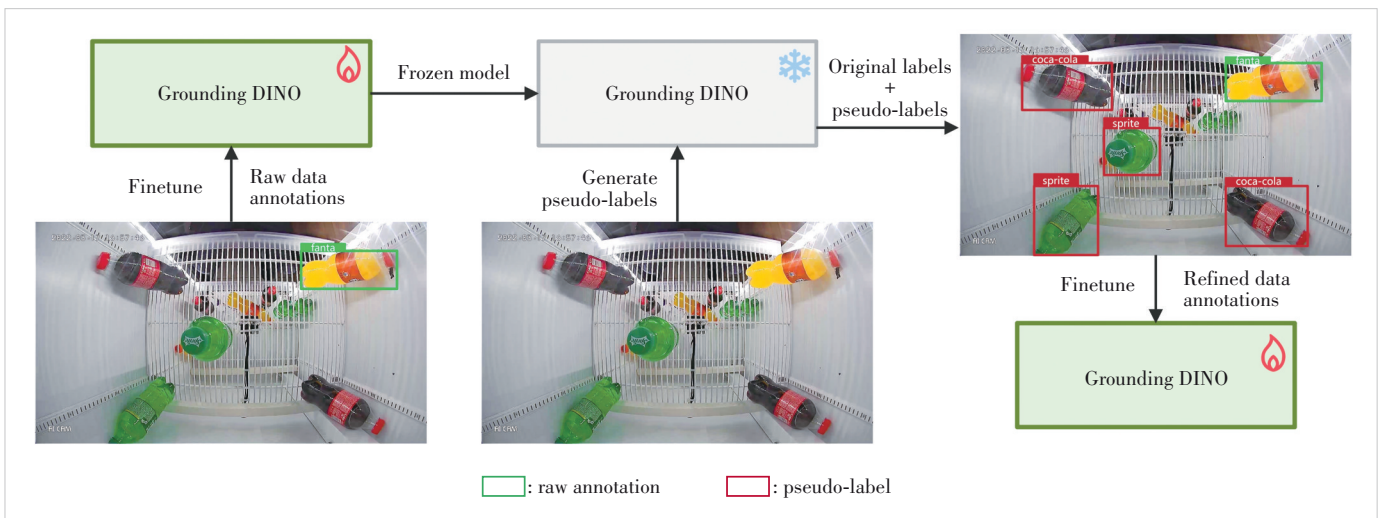


Figure 5. Illustration of the stage of fine-tuning Grounding DINO with pseudo-labeling refinement

DINO model fine-tuned in Stage 2, except for the newly introduced MoE-FFN layers. For these MoE-FFN layers, the shared FFN is initialized with the weights of the fine-tuned Grounding DINO to provide a solid foundation, while the routed FFNs are initialized with the weights from the pre-trained Grounding DINO model to encourage diversity among experts and facilitate robust training. The router, implemented as a simple FFN layer, is initialized randomly.

When fine-tuning MoE-Grounding DINO, we only allow the MoE-FFN layers, the bounding box regression head, and the classification head to be trainable to ensure stable training. As before, MoE-Grounding DINO is trained on the refined dataset constructed in Stage 2.

3.5 Data Augmentation

The success of object detectors in CD-FSOD hinges on learning robust and domain-invariant features from limited data. Therefore, we apply strong data augmentation techniques during fine-tuning to maximize the information extracted from constrained training data and to enhance feature diversity. Specifically, we combine RandomFlip, RandomResize, RandomCrop, YOLOXHSVRandomAug, and Copy-Paste to mitigate overfitting and enhance the variability of the training data.

4 Experiments

4.1 Datasets

We evaluate our approach on the Roboflow20-VL dataset^[11], a widely adopted benchmark for CD-FSOD. This dataset is specifically designed to assess the generalization ability of object detectors on out-of-domain data, presenting a significant challenge as it contains concepts rarely encountered in internet-scale pre-training datasets. Roboflow20-VL comprises 20 datasets spanning seven diverse domains: Aerial, Document, Flora & Fauna, Industrial, Medical, Sports, and Others. To rigorously evaluate cross-domain adaptability, we adopt the provided 10-shot learning setup (i.e., only 10 bounding box annotations per class). Models are evaluated using the mAP metric independently for each class. We follow Ref. [29] for efficient dataset management.

4.2 Implementation Details

All experiments were conducted within the PyTorch framework on eight NVIDIA V100 GPUs. We used the Swin-L version of Grounding DINO with pre-trained weights from the mmdetection implementation (MM-GDINO-L*)^[30], which represents the best-performing open-source model available. This model was pre-trained on large-scale datasets, including Objects365-V2^[16], OpenImageV6^[31], GoldG^[32], COCO, and Ref-COCO^[33]. For LLM-based text prompt optimization, we used Qwen2.5-VL 7B. Our MoE-Grounding DINO was configured with $N=3$ experts and a top- $K=2$ routing mechanism.

During fine-tuning, we set the learning rate to 2^{-5} for Grounding DINO and 1^{-5} for MoE-Grounding DINO. Both models were fine-tuned for 24 epochs. All other training parameters (e.g., loss weights, number of queries) followed the default settings of Grounding DINO. For pseudo-label generation, we applied a confidence threshold of 0.5 and a Non-Maximum Suppression (NMS) Intersection over Union (IoU) threshold of 0.5. The relatively high confidence threshold was chosen empirically to prioritize label quality, as missing annotations only reduce data utilization, whereas incorrect annotations can significantly degrade model performance.

4.3 Main Results on Roboflow20-VL

Table 1 summarizes the CD-FSOD performance on the Roboflow20-VL benchmark, comparing our framework with TFA^[22], Federated Detic^[4], and Grounding DINO^[1]. While the representative baseline TFA achieves competitive performance on in-domain few-shot benchmarks such as COCO, it struggles to generalize under the CD-FSOD setting (yielding only 9.8 mAP overall). Federated Detic, which employs federated fine-tuning to mitigate the impact of negative categories, performs better than TFA (20.3 mAP overall) but is still outperformed by directly fine-tuned Grounding DINO. Notably, directly fine-tuning Grounding DINO delivers a substantial improvement of 16.2 mAP over its zero-shot baseline (33.3 mAP vs. 17.1 mAP). However, this result remains far below its fine-tuning performance on the in-domain COCO dataset, which reaches 60.3 mAP as reported in Ref. [26]. This contrast underscores a crucial point: while fine-tuning Grounding DINO exhibits high performance on conventional object detection da-

Table 1. Comparison of performance of zero-shot, fine-tuned Grounding DINO with our proposed framework on Roboflow20-VL

Framework	Aerial	Document	Flora & Fauna	Industrial	Medical	Sports	Other	All
TFA ^[22]	9.4	3.8	16.8	14.4	2.7	1.3	10.2	9.8
Federated Detic ^[4]	11.6	14.3	30.8	24.7	8.9	17.4	21.0	20.3
Grounding DINO (zero-shot)	30.6	5.0	33.9	13.0	0.4	5.5	16.8	17.1
Grounding DINO (sft)	39.8	34.5	45.6	37.8	23.3	26.3	24.7	33.3
ETS ^[6]	41.6	27.4	48.1	49.2	27.4	30.9	33.7	36.9
Our framework	49.6	46.3	55.0	61.3	42.5	41.2	45.1	48.7

ETS: enhance then search TFA: two-stage fine-tuning approach

tasets like COCO, its performance under the CD-FSOD setting lags significantly behind expectations.

Table 1 clearly demonstrates that our proposed CD-FSOD framework notably outperforms directly fine-tuning Grounding DINO across all domains in the Roboflow20-VL benchmark. Specifically, our framework achieves an impressive overall enhancement of 15.4 mAP. In detail, our framework surpasses the baseline by 9.8 mAP, 11.8 mAP, 9.4 mAP, 23.5 mAP, 19.2 mAP, 14.9 mAP and 20.4 mAP in the Aerial, Document, Flora & Fauna, Industrial, Medical, Sports, and Others domains, respectively.

Furthermore, our work achieved first place in the Roboflow-20VL Few-Shot Object Detection Challenge^[11] at the Workshop on Visual Perception and Learning in an Open World at CVPR 2025, where it was also prominently featured.

These results not only underscore the efficacy of our proposed framework, but also highlight its versatility and robustness across diverse domains under the CD-FSOD setting. The significant performance gains achieved through our methodology demonstrate its potential in enhancing object detection tasks in varied real-world scenarios.

4.4 Ablation Experiments

We conducted a series of ablation experiments to systematically evaluate the effectiveness of each component of our framework in the overall performance. Additionally, we explored the impact of the number of routed experts N in our proposed MoE-Grounding DINO architecture.

1) LLM-based text prompt optimization. As shown in Table 2, fine-tuning Grounding DINO with our optimized text prompts yields a significant improvement of 7.0 mAP compared to fine-tuning with the original dataset. This substantial gain underscores the critical role of high-quality text prompts in the CD-FSOD setting and validates the effectiveness of our proposed LLM-based text prompt optimization module in generating more descriptive and discriminative prompts.

2) Data augmentation strategies. Employing the data augmentation

techniques we detailed before provides an additional benefit of 4.3 mAP (Table 2). This improvement highlights the importance of enhancing the diversity and variability of the training data, particularly in few-shot scenarios where data scarcity is a major challenge. By increasing data diversity, our augmentation strategies enable the model to generalize better to unseen instances.

3) Pseudo-labeling. After continuing to train Grounding DINO on a dataset augmented with pseudo-labels generated from an initial fine-tuned model, we observe a further performance improvement of 1.9 mAP (Table 2). This demonstrates the effectiveness of our pseudo-labeling approach in maximizing data utilization under the CD-FSOD setting, allowing the model to learn from a larger and more comprehensive set of examples, thereby mitigating the limitations of limited labeled data.

4) Fine-tuning MoE-Grounding-DINO. We evaluate the impact of fine-tuning our proposed MoE-Grounding DINO. Building upon the previously fine-tuned Grounding DINO, incorporating the MoE modules further boosts the overall detection performance by 2.2 mAP (Table 2). This significant improvement highlights the effectiveness of employing the MoE module to seamlessly expand the parameter space of Grounding DINO, thereby handling the data scarcity challenge present in the CD-FSOD context.

5) Effect of the number of routed experts. Using three routed experts yields the optimal performance, as shown in Fig. 6. Using a larger number of experts provides no addi-

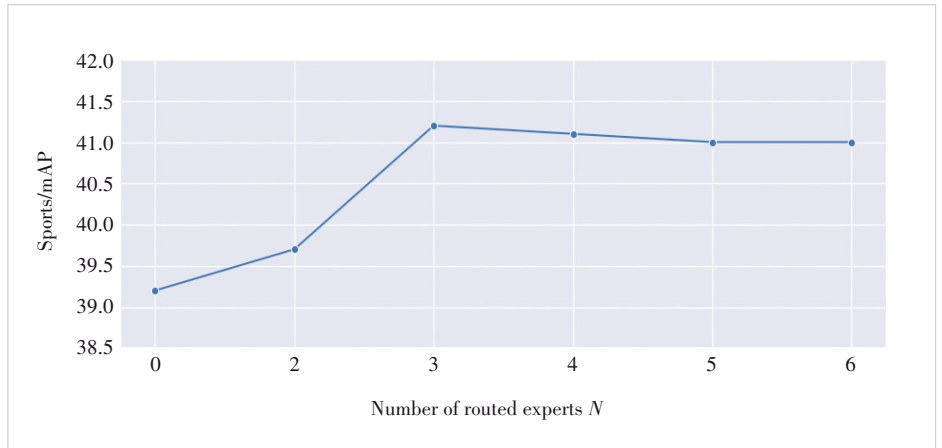


Figure 6. Effect of the total number of routed experts (N) in MoE-Grounding DINO, evaluated on the Sports domain, with K fixed at 2. An N value of 0 indicates the baseline model without the MoE module

Table 2. Effectiveness of different components in our end-to-end framework on Roboflow20-VL

Component	Aerial	Document	Flora & Fauna	Industrial	Medical	Sports	Other	All
Grounding DINO (sft)	39.8	34.5	45.6	37.8	23.3	26.3	24.7	33.3
+ text prompt optimization	39.9	40.2	49.7	47.4	36.2	33.8	34.8	40.3 (+7.0)
+ data augmentation	48.2	42.5	53.6	52.0	37.8	38.1	40.2	44.6 (+4.3)
+ pseudo-labeling	49.6	44.0	54.6	56.4	40.2	39.2	41.6	46.5 (+1.9)
+ MoE-Grounding DINO	49.6	46.3	55.0	61.3	42.5	41.2	45.1	48.7 (+2.2)

tional benefit. We attribute this to the few-shot learning scenario, where the limited number of training samples may be insufficient to effectively fine-tune a larger number of experts.

5 Conclusions

In this paper, we introduce the first end-to-end framework for cross-domain few-shot object detection, which significantly advances the state of the art on the Roboflow20-VL benchmark. Our framework uniquely integrates LLM-based text prompt optimization and a multi-stage training pipeline with pseudo-labeling. Another core contribution of our work is the novel MoE-Grounding DINO, which marks the pioneering application of the mixture-of-experts in open-set object detection.

Although representing a significant advancement, our framework still faces certain limitations. Firstly, the pseudo-labeling effectiveness can be sensitive to the initial model's performance and the severity of the domain shift. Noise in pseudo-labels can potentially propagate errors. Secondly, the substitution of MoE modules introduces computational and memory overhead, challenging real-time deployment and resource-constrained devices.

These limitations require further attention and improvement in future research to enhance the robustness and applicability of our framework. In addition, we will integrate the MoE architecture into other open-set object detectors to validate its efficacy and generalizability beyond Grounding DINO.

References

- [1] LIU S L, ZENG Z Y, REN T H, et al. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection [C]//European Conference on Computer Vision. ECCV, 2024: 38 – 55. DOI: 10.1007/978-3-031-72970-6_3
- [2] REN T H, JIANG Q, LIU S L, et al. Grounding DINO 1.5: advance the “edge” of open-set object detection [EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2405.10300>
- [3] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//European Conference on Computer Vision. ECCV, 2014: 740 – 755. DOI: 10.1007/978-3-319-10602-1_48
- [4] MADAN A, PERI N, KONG S, et al. Revisiting few-shot object detection with vision-language models [C]//Proc. Advances in Neural Information Processing Systems 37. NeurIPS, 2024: 19547 – 19560. DOI: 10.52202/079017-0617
- [5] FU Y Q, WANG Y, PAN Y X, et al. Cross-domain few-shot object detection via enhanced open-set object detector [C]//European Conference on Computer Vision. ECCV, 2024: 247 – 264. DOI: 10.1007/978-3-031-73636-0_15
- [6] PAN J C, LIU Y X, HE X, et al. Enhance then search: an augmentation-search strategy with foundation models for cross-domain few-shot object detection [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2025: 1539 – 1547. DOI: 10.1109/CVPRW67362.2025.00143
- [7] JACOBS R A, JORDAN M I, NOWLAN S J, et al. Adaptive mixtures of local experts [J]. Neural computation, 1991, 3(1): 79 – 87
- [8] JIANG A Q, SABLAYROLLES A, ROUX A, et al. Mixtral of experts [R]. 2024
- [9] LIU A, FENG B, WANG B, et al. Deepseek-v2: a strong, economical, and efficient mixture-of-experts language model [R]. 2024
- [10] CAI W, JIANG J, WANG F, et al. A survey on mixture of experts [R]. 2024
- [11] ROBICHEAUX P, POPOV M, MADAN A, et al. Roboflow100-vl: a multi-domain object detection benchmark for vision-language models [R]. 2025
- [12] LI L H, ZHANG P C, ZHANG H T, et al. Grounded language-image pre-training [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 10955 – 10965. DOI: 10.1109/CVPR52688.2022.01069
- [13] ZHONG Y W, YANG J W, ZHANG P C, et al. RegionCLIP: region-based language-image pretraining [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 16772 – 16782. DOI: 10.1109/CVPR52688.2022.01629
- [14] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision [C]//International Conference on Machine Learning. PMLR, 2021: 8748 – 8763
- [15] ZHANG H, LI F, LIU S, et al. Dino: DETR with improved denoising anchor boxes for end-to-end object detection [R]. 2022
- [16] SHAO S, LI Z M, ZHANG T Y, et al. Objects365: a large-scale, high-quality dataset for object detection [C]//Proc. IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 8430 – 8439. DOI: 10.1109/iccv.2019.00852
- [17] HUDSON D A, MANNING C D. GQA: a new dataset for real-world visual reasoning and compositional question answering [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 6693 – 6702. DOI: 10.1109/CVPR.2019.00686
- [18] YAN X P, CHEN Z L, XU A N, et al. Meta R-CNN: towards general solver for instance-level low-shot learning [C]//Proc. IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 9577 – 9586. DOI: 10.1109/iccv.2019.00967
- [19] HAN G X, LIM S N. Few-shot object detection with foundation models [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2024: 28608 – 28618. DOI: 10.1109/CVPR52733.2024.02703
- [20] OQUAB M, DARCET T, MOUTAKANNI T, et al. Dinov2: learning robust visual features without supervision [R]. 2023
- [21] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection [R]. 2021
- [22] WANG X, HUANG T, GONZALEZ J, et al. Frustratingly simple few-shot object detection [C]//International Conference on Machine Learning. PMLR, 2020: 9919 – 9928
- [23] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]//IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137 – 1149. DOI: 10.1109/TPAMI.2016.2577031
- [24] SINGH R, PUHL R B, DHAKAL K, et al. Few-shot adaptation of grounding DINO for agricultural domain [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2025: 5332 – 5342. DOI: 10.1109/CVPRW67362.2025.00530
- [25] ZHOU X, GIRDHAR R, JOULIN A, et al. Detecting twenty-thousand classes using image-level supervision [C]//European Conference on Computer Vision. ECCV, 2022: 350 – 368. DOI: 10.1007/978-3-031-20077-9_21
- [26] FEDUS W, ZOPH B, SHAZEER N. Switch transformers: scaling to trillion parameter models with simple and efficient sparsity [J]. Journal of machine learning research, 2022, 23(120): 1 – 39
- [27] XUE F Z, ZHENG Z A, FU Y, et al. OpenMoE: an early effort on open mixture-of-experts language models [R]. 2024
- [28] BAI S, CHEN K Q, LIU X J, et al. Qwen2.5-VL technical report [R]. 2025
- [29] HAN Y J, NIU J H, TU Y F. Development trends and challenges of data management systems [J]. ZTE technology journal, 2023, 27(2): 64 – 71. DOI: 10.12142/ZTETJ.202304012
- [30] ZHAO X Y, CHEN Y C, XU S L, et al. An open and comprehensive pipeline for unified object grounding and detection [R]. 2024

- [31] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The open images dataset v4: unified image classification, object detection, and visual relationship detection at scale [J]. International journal of computer vision, 2020, 128 (7): 1956 – 1981. DOI: 10.1007/s11263-020-01316-z
- [32] JENKINS P, SACHDEVA R, KEBE G Y, et al. Presentation and analysis of a multimodal dataset for grounded language learning [R]. 2020
- [33] KAZEMZADEH S, ORDONEZ V, MATTEN M, et al. ReferItGame: referring to objects in photographs of natural scenes [C]//Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). USAACL, 2014: 787 – 798. DOI: 10.3115/v1/d14-1086

Biographies

DONG Xiugang is an AI engineer at ZTE Corporation. His work primarily focuses on the research and development of large computer vision and video understanding models. His research interests span multiple areas, including open-set object detection, semantic segmentation, video spatio-temporal localization, and general video understanding.

ZHANG Kaijin is an AI engineer at ZTE Corporation, specializing in the research and development of large computer vision models. His research interests include open-set object detection, semantic segmentation, and keypoint detection.

NONG Qingpeng is an AI engineer at ZTE Corporation, specializing in the research and development of large computer vision models. His research interests include open-set object detection, semantic segmentation, and keypoint detection.

JU Minhan received his Bachelor’s degree in data science and big data technology from Xi’an Jiaotong-Liverpool University, China. He is currently pursuing a master’s degree in data science at the University of Sydney, Australia. His research interests include machine learning, data mining, and big data system modeling. He is now interning at ZTE Corporation.

TU Yaofeng (tu.yaofeng@zte.com.cn) is the Deputy Dean of the Central Research Institute of ZTE Corporation. As a PhD and senior researcher, he focuses his research on big data, databases, AI, large models, and cloud computing.



Shortened PAC Codes and List Decoding

LIU Aolin^{1,2}, FENG Bowen¹, LIANG Chulong³, XU Jin³,
ZHANG Qinyu^{1,2}

(1. Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China;
2. Pengcheng Laboratory, Shenzhen 518055, China;
3. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202504010

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251205.1832.002.html>,
published online December 08, 2025

Manuscript received: 2024-08-14

Abstract: Shortening is a standard rate-matching method for polar codes in wireless communications. Since polarization-adjusted convolutional (PAC) codes also have a block length limited to the integer powers of two, they also require rate-matching. To this end, we first analyze the limitations of existing shortening patterns for PAC codes and explore their feasibility. Subsequently, we propose a novel shortening scheme for PAC codes based on list decoding, where the receiver is allowed to treat the values of the deleted bits as undetermined. This approach uses a specialized PAC codeword and activates multiple decoding paths during the initialization of list decoding, enabling it to achieve the desired reliability.

Keywords: polarization-adjusted convolutional codes; rate-matching; shortened codes; list decoding

Citation (Format 1): LIU A L, FENG B W, LIANG C L, et al. Shortened PAC codes and list decoding [J]. ZTE Communications, 2025, 23(4): 86 – 96. DOI: 10.12142/ZTECOM.202504010

Citation (Format 2): A. L. Liu, B. W. Feng, C. L. Liang, et al., “Shortened PAC codes and list decoding,” *ZTE Communications*, vol. 23, no. 4, pp. 86 – 96, Dec. 2025. doi: 10.12142/ZTECOM.202504010.

1 Introduction

Polar coding is a high-reliability channel coding technique^[1], which has been widely used for control information in 5G communication systems. In wireless communications, rate-matching is a key step for polar codes to adjust the block length and code rate to fit actual requirements, since the block length of polar codes follows $N = 2^n$. In 3GPP TS 38.212^[2], puncturing or shortening is applied to polar codes when the required block length E satisfies $N/2 < E < N$. The standard specifies the puncturing when the actual required code rate $K/E \leq 7/16$. In this case, $(N - E)$ bits are deleted from the mother codeword. These punctured bits are unknown to the receiver and can be considered as bits transmitted over zero-capacity channels. In contrast, the shortening of polar codes is applied when the actual required code rate $K/E > 7/16$. When constructing a shortened polar codeword, $(N - E)$ bits are deleted and their values are fixed by freezing the correlated source bits or by applying dynamic frozen bits (DFBs). Therefore, the corresponding $(N - E)$ sub-channels can be considered as one-capacity channels, and the

shortening is also known as capacity-one puncturing, as described in Ref. [3].

Polarization-adjusted convolutional (PAC) coding is a representative pre-transformed polar coding scheme^[4]. As shown in Ref. [5], with sequential or list decoding, PAC codes at short block lengths can achieve better block error rate (BLER) performance than their underlying polar codes. The construction of PAC codes involves rate-profiling and convolutional pre-transform prior to the code-rate-1 polar coding. PAC codes can be viewed as a concatenated coding scheme with polar coding as the inner code. Hence, the block length of PAC codewords is also $N = 2^n$. Consequently, rate-matching is also required for PAC codes. However, the pre-transform enables the information bits of PAC codes to utilize more sub-channels^[6], thereby providing a concatenated gain. This makes it difficult to obtain capacity-one sub-channels for PAC codes via the conventional one-to-one bit freezing used in shortened polar codes. To implement shortened PAC codes without using DFBs, a sufficient number of one-capacity sub-channels must be reserved and cannot be allocated to information bits. This reservation reduces the channel utilization of the resulting PAC codes, thus diminishing the gain provided by the pre-transform and causing a performance loss.

In this paper, we focus on the construction of shortened

This work was supported in part by the National Natural Sciences Foundation of China (NSFC) under Grant Nos. 62301185 and 62027802, in part by ZTE Industry-University-Institute Cooperation Funds, and in part by The Major Key Project of PCL under Grant No. PCL2024A01.

PAC codes and the corresponding list decoding.

First, we discuss the limitations in constructing shortened PAC codes. We present three conventional approaches for constructing shortened PAC codes, by referencing the shortening schemes of polar codes. The first approach allocates more well-polarized locations to frozen bits, thereby obtaining enough one-capacity sub-channels. The second approach partially relaxes the pre-transform constraint to free sufficient one-capacity sub-channels. The third approach employs DFBs to ensure the deleted bits remain fixed.

Second, we propose a novel scheme for constructing shortened PAC codes by exploiting the multi-path nature of list decoding. Unlike conventional schemes where the deleted bits are restricted to fixed values, the values of the deleted bits are considered valid in at least one path during the initial stage of list decoding. In this case, the corresponding sub-channels can be regarded as conditional one-capacity channels. To aid decoding, several bits are frozen and used to carry pre-designed bits that depend on the values of the deleted bits.

Third, the corresponding list decoding is introduced, where multiple decoding paths are activated to accommodate the results derived from various possible combinations of values for the deleted bits. The corresponding values of the pre-designed bits are then calculated and assigned to each path.

With the slightly modified list decoding, the proposed shortened PAC codes achieve better BLER performance than comparable shortened or punctured polar codes.

2 Preliminaries

2.1 Rate-Matching of Polar Codes

The polar codeword with block-length $N = 2^n$ is generated by $\mathbf{x} = \mathbf{u}\mathbf{G}_N^{[1]}$, where the generator matrix $\mathbf{G}_N = \mathbf{F}_2^{\otimes n}\mathbf{B}_N$, in which the involutory matrix \mathbf{B}_N refers to bit-reversal operation and $\mathbf{F}_2^{\otimes n}$ refers to the n -order Kronecker product of the kernel matrix $\mathbf{F}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

The puncturing of polar codes has been extensively studied. Among the notable works, Ref. [7] introduces the quasi-uniform puncturing (QUP) for polar codes. The indices in the QUP pattern $\mathcal{P}_{\text{QUP}} = \beta_N(\mathcal{U}_{N-E})$ are located quasi-uniformly in the mother codeword. The reversal function is $\beta_{2^n}(\zeta) = \mathcal{U}_{2^n}\mathbf{B}_{2^n}(\mathcal{U}_{2^n}, \zeta)$. In Ref. [8], a worst quality puncturing (WQP) pattern is formed by the indices of the least reliable sub-channels. In contrast, Ref. [9] selects puncturing positions based on columns with the minimum column weight (MCW). A low-complexity method in Ref. [10] provides a bit-reversed QUP (BRP) pattern, given by $\mathcal{P}_{\text{BRP}} = \mathcal{U}_{N-E}$. Moreover, the changes in sub-channels' reliability caused by puncturing are discussed. A mandatory frozen subset $\beta_N(\mathcal{U}_{N-E}) \subseteq \mathcal{A}^c$ is provided to avoid transmitting information bits over zero-capacity

sub-channels. Based on the changes in sub-channels' reliability, an information set approximation puncturing (ISAP) pattern is proposed for low-code-rate codewords^[11], where certain guard bits are set as unpuncturable. Furthermore, Ref. [12] defines a minimum reliability score (MMRS) as the objective and devises a puncturing pattern to maximize this score. The MMRS pattern performs particularly well for high-code-rate codewords.

Research on shortening of polar codes has also yielded several important schemes. In Ref. [13], a criterion is established to obtain an $E \times E$ reduced-dimensional generator matrix by searching for columns with weight 1. The shortening pattern \mathcal{S} consists of the indices of the deleted $(N - E)$ columns of the generator matrix. The index set of the deleted rows is called the mandatory frozen subset, denoted as $\mathcal{F}_{\text{polar}}$ with $\mathcal{F}_{\text{polar}} \subseteq \mathcal{A}^c$. This subset ensures the existence of $(N - E)$ one-capacity sub-channels. Since the unreversed generator matrix $\mathbf{F}_2^{\otimes n}$ is a lower-triangular matrix, Ref. [13] also provides a simple shortening pattern by setting the mandatory frozen subset as $\mathcal{F}_{\text{polar}} = \{E + 1, E + 2, \dots, N\}$. Ref. [14] adapts the method in Ref. [13] and provides fundamental insights into the shortening of PAC systematic codes. In Ref. [15], the shortening pattern and the frozen set are jointly optimized to achieve lower bit-error probability, and the DFBs are applied to part of the bits of $\mathbf{u}_{\mathcal{F}}$ to ensure that the values of the deleted bits are fixed. In Ref. [9], a low-complexity method is introduced to select both puncturing and shortening patterns, where a WANG-LIU bit-reversed shortening (BRS) pattern is used. The same pattern is also presented in Ref. [16]. In Refs. [17] and [18], a shortening pattern is introduced based on evaluating the reliability of sub-channels via Gaussian approximation (GA). Ref. [19] redesigns a subblock-wise interleaver to improve rate-matching performance, based on an analysis of the locations of zero-capacity and one-capacity sub-channels.

2.2 PAC Codes and List Decoding

In PAC codes, a convolutional transform is employed as a precoding stage before the polar transform, providing a significant concatenated gain^[20-21]. The construction proceeds in two steps. First, the information bit vector \mathbf{d} is mapped to a data carrier word $\mathbf{v} = \{\mathbf{v}_{\mathcal{A}}, \mathbf{v}_{\mathcal{A}^c}\}$ via rate-profiling, i.e., $\mathbf{v}_{\mathcal{A}} = \mathbf{d}$ and $\mathbf{v}_{\mathcal{A}^c} = \mathbf{0}$. Second, the interim codeword \mathbf{u} is obtained by the convolutional pre-transform $\mathbf{u} = \mathbf{v}\mathbf{T}_N$. $\mathbf{T}_N = (t_{a,b})_{a,b \in \mathcal{U}_N}$ is the N -dimension generator matrix of convolutional codes, expanded from $\mathbf{c} = (c_1, c_2, \dots, c_s)$, where

$$\begin{cases} t_{a,b} = c_{b-a+1}, & \text{if } a \leq b < a + s \\ t_{a,b} = 0, & \text{otherwise} \end{cases} \quad (1),$$

and where \mathcal{U}_N refers to the integer sequence $\mathcal{U}_N = \{1, 2, \dots, N\}$. Therefore, each interim bit can be obtained by

$$u_i = \sum_{j=1}^s c_j v_{i-j+1} \quad (2).$$

Finally, the PAC codeword \mathbf{x} is obtained by $\mathbf{x} = \mathbf{u}\mathbf{G}_N^{[4]}$.

In the list decoding of PAC codes, the Log-Likelihood Ratio (LLR) corresponding to bit u_i , $i \in \mathcal{U}_N$ on each decoding path is

$$\text{expressed as } L_i^n[\ell] = \ln \frac{\Pr(\mathbf{y}, \hat{\mathbf{u}}_{\mathcal{U}_{i-1}}[\ell] | \hat{u}_i[\ell] = 0)}{\Pr(\mathbf{y}, \hat{\mathbf{u}}_{\mathcal{U}_{i-1}}[\ell] | \hat{u}_i[\ell] = 1)}, \quad \ell \in \mathcal{L},$$

where \mathbf{y} refers to the received symbols of the PAC codeword, and $\hat{u}_i[\ell]$ is obtained from Eq. (2). This LLR is derived according to the recursive formulas in Ref. [1], given by

$$\begin{cases} L_{2^{p+1}q+p}^{m-1}[\ell] = L_{2^{p+1}q+2p-1}^{m-1}[\ell] \boxplus L_{2^{p+1}q+2p}^{m-1}[\ell] \\ L_{2^{p+1}q+p}^{m-1}[\ell] = \nabla[\ell] \cdot L_{2^{p+1}q+2p-1}^{m-1}[\ell] + L_{2^{p+1}q+2p}^{m-1}[\ell] \end{cases} \quad (3),$$

where the operation $a \boxplus b = 2 \tanh^{-1}(\tanh(a/2) \tanh(b/2))$, and the parameters are $m \in \mathcal{U}_n$, $p \in \mathcal{U}_{2^p}$, $q \in \mathcal{U}_{2^{p-1}-1}$, $2^p = 2^{n-m}$ and $\nabla[\ell] = (-1)^{\hat{u}_{q \cdot 2^{p+1} + \mathcal{U}_{2^p}}[\ell] G_{2^p}(\mathcal{U}_{2^p, p})}$. Path-metric (PM) is used to evaluate the reliability of each candidate path, which is expressed as:

$$PM_{\hat{v}_i[\ell]} = \sum_{j=1}^i BM_{\hat{v}_j[\ell]} = \sum_{j=1}^i \ln(1 + e^{-(1-2\hat{u}_j[\ell])L_j^n[\ell]}) \quad (4),$$

where $BM_{\hat{v}_i[\ell]}$ refers to the branch metric (BM) of the path ℓ at the i -th bit. Up to L paths with smaller PMs are retained.

3 Encoding and Decoding Schemes for Shortened PAC Codes

3.1 Limitations in Construction of Shortened PAC Codes

The unreversed generator matrix of PAC codes is $\mathbf{H}_N = \mathbf{T}_N \mathbf{F}_2^{\otimes n}$, which can be expressed as:

$$\mathbf{H}_N = \begin{bmatrix} \mathbf{A}_{(N-s+1) \times (s-1)} & \mathbf{C}_{(N-s+1) \times (N-s+1)} \\ \mathbf{D}_{(s-1) \times (s-1)} & \mathbf{E}_{(s-1) \times (N-s+1)} \end{bmatrix} \quad (5),$$

where the submatrix \mathbf{C} is lower-triangular. The reversed generator matrix of PAC codes is $\mathbf{J}_N = \mathbf{H}_N \mathbf{B}_N$. The minimum column-weight of \mathbf{J}_N is now $\sum_{i=1}^s c_i$, so there is no longer a column with weight 1. To make the existing shortening methods of polar codes applicable to PAC codes, we propose the following three straightforward modification schemes.

1) Expanding the mandatory frozen subset \mathcal{F}

According to Eq. (2), the value of u_i depends on the bits $v_{i+1-\text{supp}(c)}$, where $\text{supp}(\cdot)$ refers to the set-theoretic support function, i.e., the indices of the non-zero elements in the set. To obtain the same shortening pattern \mathcal{S} as the shortened polar codes, the indices of all dependent bits of $\mathbf{u}_{\mathcal{F}_{\text{polar}}}$ should be included in the mandatory frozen subset, i.e.,

$$\mathcal{F}_{\text{PAC}} \supseteq \{f_i + 1 - \text{supp}(c)\}, \quad \forall f_i \in \mathcal{F}_{\text{polar}} \quad (6),$$

where \mathcal{F}_{PAC} refers to the mandatory frozen subset of PAC codes with $\mathbf{v}_{\mathcal{F}_{\text{PAC}}} = \mathbf{0}$. Therefore, the information set \mathcal{A} is narrowed down to $\{\mathcal{U}_N \setminus \mathcal{F}_{\text{PAC}}\}$. The strengths and weaknesses of this scheme are as follows.

- Strengths: The decoding of the PAC code does not need to be modified.

- Weaknesses: Compared to shortened polar codes, more indices corresponding to unfavorably polarized sub-channels have to be selected to carry the information bits, which diminishes reliability. For the WANG-LIU pattern, the mandatory frozen subset is expanded to $\mathcal{F}_{\text{PAC}} = \{E - s + 1, \dots, E, E + 1, \dots, N\}$. For the other shortening patterns, more elements are included in \mathcal{F}_{PAC} , causing a significant performance loss. In particular, the BRS pattern is not directly applicable.

2) Reconstructing the pre-transform matrix \mathbf{T}_N

By reconstructing \mathbf{T}_N , we can avoid the expansion of the mandatory frozen subset caused by the pre-transform and make the original subset $\mathcal{F}_{\text{polar}}$ valid, i.e. $\mathcal{F}_{\text{PAC}} = \mathcal{F}_{\text{polar}}$. The reconstructed pre-transform matrix $\tilde{\mathbf{T}}_N = (\tilde{t}_{a,b})_{a,b \in \mathcal{U}_N}$ can be expressed as:

$$\begin{cases} \tilde{t}_{a,b} = c_{b-a+1}, & \text{if } a \leq b < a + s \text{ and } b \notin \mathcal{F}_{\text{polar}} \\ \tilde{t}_{a,b} = 1, & \text{else if } a = b \text{ and } b \in \mathcal{F}_{\text{polar}} \\ \tilde{t}_{a,b} = 0, & \text{otherwise} \end{cases} \quad (7).$$

In this case, the interim codeword generated from $\tilde{\mathbf{T}}_N$ satisfies $\mathbf{u}_{\mathcal{F}_{\text{PAC}}} = \mathbf{v}_{\mathcal{F}_{\text{PAC}}}$, and the corresponding deleted bits $\mathbf{x}_{\mathcal{S}}$ are actually the polar code bits. The characteristics of this scheme are summarized below.

- Strengths: It retains an identical shortening pattern and mandatory frozen subset as conventional shortened polar codes, ensuring compatibility.

- Weaknesses: It requires a modified decoding procedure: standard polar code decoding rules are used for bits $\mathbf{v}_{\mathcal{F}_{\text{PAC}}}$, while PAC code decoding rules for the remaining bits, thereby increasing receiver complexity.

3) Applying dynamic frozen bits to $\mathbf{v}_{\mathcal{F}}$

The expansion of the mandatory frozen subset can be prevented by applying DFBs, which constrain the deleted bits to zero, i.e., $\mathbf{x}_{\mathcal{S}} = \mathbf{0}$ and $\mathcal{F}_{\text{PAC}} = \mathcal{F}_{\text{polar}}$. The mandatory frozen bits $\mathbf{v}_{\mathcal{F}_{\text{PAC}}}$ are designated as DFBs. Their values are derived from a linear combination of information bits. This combination can be expressed as:

$$\mathbf{v}_{\mathcal{F}_{\text{PAC}}} = \mathbf{g}(\mathbf{v}_{\mathcal{A}}) = \mathbf{v}_{\mathcal{A}} \mathbf{K}(:, \mathcal{A})^T \quad (8),$$

where the matrix \mathbf{K} is the reduced row echelon form obtained by applying Gaussian elimination to $\mathbf{J}_N(:, \mathcal{S})^T$. The submatrix $\mathbf{K}(:, \mathcal{F}_{\text{PAC}})$ should be an identity matrix, while $\mathbf{J}_N(:, \mathcal{S})$ refers

to the submatrix consisting of the \mathcal{S} -th columns of \mathbf{J}_N .

For this scheme, the one-capacity sub-channels of shortening are conditionally present in the decoding graph. When the information bits $\mathbf{v}_{\mathcal{A}}$ are correctly decoded, the results of the DFBs are accurate, and the one-capacity sub-channels exist. Once the bits $\mathbf{v}_{\mathcal{A}}$ are decoded incorrectly, the DFB results become unreliable and the one-capacity sub-channels no longer exist.

The strengths and weaknesses of this scheme are discussed as follows.

- Strengths: The shortening pattern and the mandatory frozen subset are the same as those of shortened polar codes.
- Weaknesses: The decoding process needs to be modified. The linear function $g(\cdot)$ should be taken as the inputs for decoding, serving as an additional step to decode the DFBs. The function $g(\cdot)$ varies with different block-lengths and code-rates.

3.2 A Novel Shortening Scheme for PAC Codes

We provide a novel idea to allow $\mathbf{x}_{\mathcal{S}}$ to be undetermined at the receiver, aiming to reduce the size of \mathcal{F}_{PAC} . A subset \mathcal{I} is removed from \mathcal{F}_{PAC} and incorporated into the information set; thus, the mandatory frozen subset becomes $\{\mathcal{F}_{\text{PAC}} \setminus \mathcal{I}\}$. All possible values of $\mathbf{x}_{\mathcal{S}}$ are assigned to the paths during the initialization of list decoding, ensuring that at least one path contains the true value.

Similar to the DFB scheme, the one-capacity sub-channels of shortening are conditionally present in the list decoding process. When the paths assigned with the true value of $\mathbf{x}_{\mathcal{S}}$ can be retained in the list during decoding, the one-capacity sub-channels exist in these paths and contribute to reliable decoding performance. Therefore, despite $\mathbf{x}_{\mathcal{S}}$ being undetermined at the receiver, this scheme is essentially a shortening scheme rather than a puncturing one.

The number of possible combinations of the value of $\mathbf{x}_{\mathcal{S}}$ is up to $2^{|\mathcal{I}|}$, denoted as $\hat{\mathbf{x}}_{\mathcal{S}}[\ell] = \hat{\mathbf{v}}_{\mathcal{I}}[\ell] \mathbf{J}_N(\mathcal{I}, \mathcal{S})$, $\ell \in \mathcal{U}_{2^{\mathcal{I}}}$, where $\mathbf{J}_N(\mathcal{I}, \mathcal{S})$ refers to the submatrix consisting of the \mathcal{I} -th row and \mathcal{S} -th column of \mathbf{J}_N , and $\hat{\mathbf{v}}_{\mathcal{I}}[\ell] \in \{0, 1\}^{|\mathcal{I}|}$. Therefore, the main task of this scheme is to eliminate the incorrectly assigned paths early in the decoding process while retaining the correct one.

1) Findings from list decoding

We can activate $2^{|\mathcal{I}|}$ paths in list decoding to contain the different results derived from various value-combinations of $\mathbf{x}_{\mathcal{S}}$, when $L \geq 2^{|\mathcal{I}|}$. The initial LLR corresponding to the ℓ -th path is expressed as:

$$L_i^0[\ell] = \begin{cases} (1 - 2\hat{x}_i[\ell]) \cdot \infty, & \text{if } i \in \mathcal{S} \\ \ln \frac{\Pr(\mathbf{y}|\hat{\mathbf{x}}_i[\ell] = 0)}{\Pr(\mathbf{y}|\hat{\mathbf{x}}_i[\ell] = 1)}, & \text{otherwise} \end{cases} \quad (9)$$

Before starting path-killing, the PM derived from the correct value-combination $\hat{\mathbf{x}}_{\mathcal{S}}[\ell]$ needs to be lower than the

other PMs, to avoid killing the correct path $\tilde{\ell}$. The expected values of the PMs of the correct path $\tilde{\ell}$ and incorrect path $\bar{\ell} \in \{\mathcal{U}_{2^{\mathcal{I}}} \setminus \tilde{\ell}\}$ should have

$$\langle \Delta_{\tilde{\ell}} \rangle = \left\langle \text{PM}_{\hat{\mathbf{v}}_{\min(\mathcal{A})-1}[\tilde{\ell}]} - \text{PM}_{\hat{\mathbf{v}}_{\min(\mathcal{A})-1}[\bar{\ell}]} \right\rangle > 0 \quad (10),$$

where $\langle \cdot \rangle$ refers to the expected value function. To diminish the adverse impact of incorrect paths, the minimum value $\min(\langle \Delta_{\tilde{\ell}} \rangle)$ should be as large as possible.

According to the recursive formula (3), the LLR results $L_{\mathcal{U}_{\mathcal{S}}}^n[\ell]$ derived from the full codeword's initial LLRs $L_{\mathcal{U}_{\mathcal{N}}}^0[\ell]$ are equivalent to those derived from the sub-codeword's LLRs $L_{\mathcal{U}_{\mathcal{S}}}^m[\ell]$, where the sub-codeword is expressed as $\mathbf{x}_{\mathcal{U}_{\mathcal{S}}}^m$ and the corresponding LLRs are

$$L_i^m[\ell] = \bigoplus_{j=(i-1)M+1}^{iM} L_j^0[\ell], i \in \mathcal{U}_{2^{\mathcal{I}}}, \ell \in \mathcal{U}_{2^{\mathcal{I}}} \quad (11).$$

Example 1: Consider an example in which the codeword is with the block length $N = 32$, and the initial LLRs are $\{L_1^0[\ell], L_2^0[\ell], \dots, L_{32}^0[\ell]\}$. The LLRs of the sub-codeword $\mathbf{x}_{\mathcal{U}_{\mathcal{S}}}^2$ can be obtained by $\{(L_1^2[\ell] = L_1^0[\ell] \oplus L_2^0[\ell] \oplus L_3^0[\ell] \oplus L_4^0[\ell]), \dots, (L_8^2[\ell] = L_{29}^0[\ell] \oplus L_{30}^0[\ell] \oplus L_{31}^0[\ell] \oplus L_{32}^0[\ell])\}$.

For an integer Λ , where $2^{p-1} < \Lambda \leq 2^p$, we can obtain the reversal of \mathcal{U}_{Λ} , which is expressed as $\mathcal{R}_{\Lambda} = \beta_{2^p}(\mathcal{U}_{\Lambda})$, where the reversal function is $\beta_{2^p}(\xi) = \mathcal{U}_{2^p} \mathbf{B}_{2^p}(\mathcal{U}_{2^p}, \xi)$. We define the coefficients of the sub-codeword bits $\mathbf{x}_{\mathcal{R}_{\Lambda}}^m$ in each path ℓ as:

$$h_i^m[\ell] \triangleq \hat{x}_i^m[\ell] \oplus \hat{x}_i^m[\tilde{\ell}], i \in \mathcal{R}_{\Lambda} \quad (12).$$

The coefficients $h_{\mathcal{R}_{\Lambda}}^m[\ell]$ can be partitioned into two parts. One part, denoted as $\mathbf{h}_r^m[\ell] = \{h_s^m[\ell]\}$, $\mathbf{r} \subseteq \mathcal{R}_{\Lambda}$, is influenced by the initial infinite LLRs $L_{\mathcal{S}}^0[\ell]$. According to Eq. (11), $L_{\mathcal{S}}^0[\ell]$ can be mapped to the sub-codeword's LLRs as:

$$L_{\mathcal{S}}^m[\ell] = L_{\mathcal{S}}^0[\tilde{\ell}] \cdot \left(\prod_{s \in \mathcal{S}_{\delta}} \text{sgn}(L_s^0[\ell]) \text{sgn}(L_s^0[\tilde{\ell}]) \right) \quad (13),$$

where $\text{sgn}(\cdot)$ denotes the sign function. The subset $\mathcal{S}_{\delta} \subseteq \mathcal{S}$ is expressed as:

$$\mathcal{S}_{\delta} = \left\{ s \in \mathcal{S} : \delta = \left\lceil \frac{s}{2^m} \right\rceil \right\} \quad (14),$$

where $\lceil \cdot \rceil$ denotes the ceiling function. These coefficients satisfy:

$$h_{\delta}^m[\ell] = \bigoplus_{s \in \mathcal{S}_{\delta}} \left(\hat{x}_s[\ell] \oplus \hat{x}_s[\bar{\ell}] \right), \quad \forall \delta \in \mathcal{r} \quad (15).$$

The other part of the coefficients $\mathbf{h}_{\mathcal{R}_{\Lambda}\mathcal{V}}^m[\ell]$ are not affected by the initial infinite LLRs.

Example 2: Extend Example 1 where $\Lambda = 6$ and $E = 27$ by applying the WANG-LIU pattern $\mathcal{S} = \{32, 16, 24, 8, 28\}$. The coefficients $\mathbf{h}_{\mathcal{R}_{\delta}}^m[\ell]$, $\mathcal{R}_{\delta} = \{1, 5, 3, 7, 2, 6\}$ can be divided into two sets: set $\mathbf{h}_{\mathcal{r}}^m[\ell] = \{h_7^2[\ell], h_2^2[\ell], h_6^2[\ell]\}$, influenced by the infinite LLRs $\{L_{28}^0[\ell], L_8^0[\ell], L_{24}^0[\ell]\}$, and its complement, $\mathbf{h}_{\mathcal{R}_{\Lambda}\mathcal{V}}^m[\ell] = \{h_1^2[\ell], h_5^2[\ell], h_3^2[\ell]\}$, which is not affected by the infinite LLRs.

To better distinguish between correct and incorrect paths, we impose a structure on this part by freezing the bits $\mathbf{v}_{\mathcal{U}_{\Lambda}}$ to specific values.

The effect on this part is formulated as:

$$\mathbf{h}_{\mathcal{R}_{\Lambda}\mathcal{V}}^m[\ell] = \left(\hat{\mathbf{v}}_{\mathcal{U}_{\Lambda}}[\ell] \oplus \hat{\mathbf{v}}_{\mathcal{U}_{\Lambda}}[\bar{\ell}] \right) \mathbf{J}_{2^{\nu}}(\mathcal{U}_{\Lambda}, \mathcal{R}_{\Lambda}\mathcal{V}) \quad (16).$$

In this case, we have the following lemma.

Lemma 1: A lower bound exists for the expected PM difference $\langle \Delta_{\bar{\ell}} \rangle$ between the incorrect path $\bar{\ell}$ and the correct path $\bar{\ell}$ before path-killing, which is expressed as:

$$\langle \Delta_{\bar{\ell}} \rangle > \left| \langle L_{\bar{\ell}}^n[\bar{\ell}] \rangle \right| \quad (17),$$

where $\nu = \min \left(\text{supp} \left((\hat{\mathbf{x}}_{\mathcal{R}_{\Lambda}}^m[\bar{\ell}] \oplus \hat{\mathbf{x}}_{\mathcal{R}_{\Lambda}}^m[\bar{\ell}]) \mathbf{G}_{2^{\nu}}(\mathcal{R}_{\Lambda}, \mathcal{U}_{\Lambda}) \right) \right)$.

Proof: Since $\hat{u}_{\ell}[\bar{\ell}] = 1 - \hat{u}_{\ell}[\bar{\ell}]$ denotes the first difference between the two result paths of the interim codeword, the corresponding resulting LLR satisfies $L_{\ell}^n[\bar{\ell}] = L_{\ell}^n[\bar{\ell}]$. Therefore, we can obtain $\langle \text{BM}_{\hat{v}_{\ell}[\bar{\ell}]} - \text{BM}_{\hat{v}_{\ell}[\bar{\ell}]} \rangle > \left| \langle L_{\ell}^n[\bar{\ell}] \rangle \right|$ from Eq. (4). According to Lemma 1 in Ref. [22], a front erroneous decision reduces the magnitude of subsequent resulting LLRs, i.e., $\left| \langle L_{\ell}^n[\bar{\ell}] \rangle \right| > \left| \langle L_{\ell}^n[\bar{\ell}] \rangle \right|$, where $\ell \in \{\nu + 1, \dots, N\}$. Therefore,

the corresponding BMs satisfy $\langle \text{BM}_{\hat{v}_{\ell}[\bar{\ell}]} \rangle > \langle \text{BM}_{\hat{v}_{\ell}[\bar{\ell}]} \rangle$.

2) The selection of \mathcal{I} and the value of $\hat{\mathbf{v}}_{\mathcal{U}_{\Lambda}}[\ell]$

The unfrozen set \mathcal{I} can follow that $\mathcal{I} \subseteq \beta_N(\mathcal{S})$ to ensure the value-combinations $\hat{\mathbf{x}}_{\mathcal{S}}[\ell]$, $\ell \in \mathcal{U}_{2^m}$ are independent. According to Lemma 1, we design the coefficients $\mathbf{h}_{\mathcal{R}_{\Lambda}}^m[\ell]$ to maximize the lower bound $\min \left(\langle \Delta_{\bar{\ell}} \rangle \right)$. The sub-codeword is constructed as $\mathbf{x}_{\mathcal{R}_{\Lambda}}^m[\ell] = \mathbf{u}_{\mathcal{B}}[\ell] \mathbf{G}_{2^{\nu}}(\mathcal{B}, \mathcal{R}_{\Lambda})$, where $\mathbf{u}_{\mathcal{B}}[\ell] = \{0, 1\}^{|\mathcal{I}|}$. GA is an ideal way to approximate the LLR magni-

tude $\left| \langle L_{\ell}^n[\bar{\ell}] \rangle \right|$ and thereby identify a small set $\mathcal{B} \subseteq \mathcal{U}_{\Lambda}$ to raise the lower bound. This set is composed of indices $b \in \beta_{2^{\nu}}(\mathcal{r})$ with the largest LLR magnitudes $\left| \langle L_b^n[\bar{\ell}] \rangle \right|$, which can be expressed as:

$$\mathcal{B} = \arg \max_{\mathcal{B}' \subseteq \beta_{2^{\nu}}(\mathcal{r}), |\mathcal{B}'| = |\mathcal{I}|} \sum_{b \in \mathcal{B}'} \left| \langle L_b^n[\bar{\ell}] \rangle \right| \quad (18).$$

Consequently, an upper bound of the block error probability, derived from Lemma 1, is given by

$$\Pr \left(\text{PM}_{\hat{v}_{\ell}[\bar{\ell}]} < \text{PM}_{\hat{v}_{\ell}[\bar{\ell}]} \right) < \frac{1}{2} \text{erfc} \left(\frac{1}{2} \sqrt{\min \left(\left| \langle L_{\mathcal{B}}^n[\bar{\ell}] \rangle \right| \right)} \right) \quad (19).$$

Example 3: Building on Example 2, where $\mathcal{B} = \{4, 6\}$ is always selected from $\beta_8(\mathcal{r}) = \{4, 5, 6\}$, the generator matrix of the sub-codewords becomes:

$$\mathbf{G}_8(\{4, 6\}, \{1, 5, 3, 7, 2, 6\}) = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \quad (20).$$

The minimum distance of all the possible sub-codewords is 4.

An appropriate set Λ can be determined by lowering this upper bound while minimally impacting the selection of the information set \mathcal{A} . The unfrozen set \mathcal{I} is then selected via a one-to-one mapping from \mathcal{B} .

Lemma 2: For each element $b \in \mathcal{B}$, there is always a corresponding element $(N + b - 2^{\nu}) \in \mathcal{F}_{\text{PAC}}$.

Proof: Assume $\delta \in \beta_{2^{\nu}}(b)$. Then, the corresponding set \mathcal{S}_{δ} can be obtained from Eq. (14). From Lemma 1 in Ref. [23], the elements follows that $\mathbf{G}_N(N + b - 2^{\nu}, \mathcal{S}_{\delta}) = 1$. Therefore, $u_{N+b-2^{\nu}}$ should be frozen bits or DFBs.

According to Lemma 2, we provide a simple way to obtain \mathcal{I} from \mathcal{B} , which is expressed as:

$$\mathcal{I} = \mathcal{B} + N - 2^{\nu} \quad (21).$$

Once the unfrozen set \mathcal{I} is determined, we need to examine whether all the coefficients in $\mathbf{h}_{\mathcal{r}}^m[\ell]$ are affected by the value of $\mathbf{x}_{\mathcal{S}}$.

Actually, due to the Kronecker-product construction of the generator matrix, there will be a part of elements $\mathbf{q} = \{\mathbf{r} \mathbf{v}_{\text{sub}}\}$ whose corresponding coefficients $\mathbf{h}_{\mathcal{Q}}^m[\ell]$, $\mathcal{Q} \in \mathbf{q}$ are fixed to 0 for each path $\ell \in \mathcal{U}_{2^m}$, if $|\mathcal{S}_{\mathcal{Q}}|$ is even. In other words, only the coefficients $\mathbf{h}_{\mathcal{r}_{\text{sub}}}^m[\ell]$ are affected by the the initial infinite LLRs.

Proposition 1: For the WANG-LIU pattern, the indices of the actually affected part $\mathbf{h}_{\mathcal{r}_{\text{sub}}}^m[\ell]$ can be obtained as follows:

$$\text{supp}(\beta_{2^p}(\mathbf{r}_{\text{sub}})) = \begin{cases} \mathcal{U}_{E\%2^p} \cap \mathcal{U}_\Lambda, & \text{if } \lfloor E/2^p \rfloor \text{ is even} \\ \mathcal{U}_\Lambda \setminus \mathcal{U}_{E\%2^p}, & \text{otherwise} \end{cases} \quad (22).$$

Proof: From the coefficient derivation in Eq. (15), we have $\bigoplus_{s \in \mathcal{S}_\delta} \hat{x}_s[\ell] = \bigoplus_{s \in \mathcal{S}_\delta} \hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, s)$, where

$$\bigoplus_{s \in \mathcal{S}_\delta} \hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, s) = \begin{cases} 0, & \text{if } |\mathcal{S}_\delta| \text{ is even} \\ \hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, 2^m \cdot \delta), & \text{otherwise} \end{cases} \quad (23).$$

Eq. (23) is obtained because $\hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, \mathcal{S}_\delta)$ is an all-zero (or all-one) sequence when the WANG-LIU pattern is adopted. Therefore, we have $\mathbf{r}_{\text{sub}} = \{\delta : |\mathcal{S}_\delta| \text{ is odd}\}$. We assume that δ' is the bit-reversal result of δ , i.e., $\delta' = \beta_{2^p}(\delta)$. In this case, the event that $|\mathcal{S}_\delta|$ is odd is equivalent to the event that $|\mathcal{Q}_{\delta'}|$ is odd, where $\mathcal{Q}_{\delta'} = \{s \in \mathcal{F} : \delta' = s\%2^p\}$. Consequently, Eq. (22) is obtained.

Example 4: We examine a specific case of Example 2 where $E = 20$. Under this condition, the actually affected coefficients are $\mathbf{h}_{\mathbf{r}_{\text{sub}}}^m[\ell] = \{h_7^2[\ell], h_2^2[\ell], h_6^2[\ell]\}$. The indices of these coefficients are $\mathbf{r}_{\text{sub}} = \{1, 5, 3, 7\}$. Here, an additional coefficient $h_2^2[\ell]$ is affected by $\{L_6^0[\ell], L_8^0[\ell]\}$. Since $L_6^0[\ell] = L_8^0[\ell]$, the effects of $L_6^0[\ell]$ and $L_8^0[\ell]$ on $h_2^2[\ell]$ are offset.

We now introduce an effect on the unaffected part by assigning specific values to $\hat{v}_{\mathcal{U}_\Lambda}[\ell]$ for each path. Following Lemma 1, the corresponding sub-codeword bits of each path should be constructed to satisfy $\hat{\mathbf{x}}_{\mathcal{R}_\Lambda \setminus \mathbf{r}_{\text{sub}}}^m[\ell] = \hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, 2^m(\mathcal{R}_\Lambda \setminus \mathbf{r}_{\text{sub}}))$. According to Eq. (16), the specific bits $\hat{v}_{\mathcal{U}_\Lambda}[\ell]$ of each path $\ell \in \mathcal{U}_{2^m}$ can be calculated as:

$$\hat{v}_{\mathcal{U}_\Lambda}[\ell] = \hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, 2^m(\mathcal{R}_\Lambda \setminus \mathbf{r}_{\text{sub}})) \mathbf{J}_{2^p}^{-1}(\mathcal{R}_\Lambda \setminus \mathbf{r}_{\text{sub}}, \mathcal{U}_\Lambda) \quad (24),$$

where $\mathbf{J}_{2^p}^{-1} = \mathbf{G}_{2^p} \mathbf{T}_{2^p}^{-1}$. The inverse matrix $\mathbf{T}_{2^p}^{-1}$ exists, since \mathbf{T}_{2^p} is an upper triangular matrix.

Example 5: Revisit Example 2, where the sub-codewords should be constructed via Eq. (16). The part $\hat{\mathbf{x}}_{\mathbf{r}_{\text{sub}}}^m[\ell] = \{\hat{x}_7^2[\ell], \hat{x}_2^2[\ell], \hat{x}_6^2[\ell]\}$ has been obtained by the effects from initial infinite LLRs, which can be expressed as:

$$\begin{aligned} \{\hat{x}_7^2[\ell], \hat{x}_2^2[\ell], \hat{x}_6^2[\ell]\} &= \left\{ \frac{1}{2} (1 - \text{sgn}(L_{28}^0[\ell])), \right. \\ &\left. \frac{1}{2} (1 - \text{sgn}(L_8^0[\ell])), \frac{1}{2} (1 - \text{sgn}(L_{24}^0[\ell])) \right\} = \\ &\{\hat{x}_{28}^0[\ell], \hat{x}_8^0[\ell], \hat{x}_{24}^0[\ell]\} \end{aligned} \quad (25).$$

Consequently, the bits $\hat{\mathbf{x}}_{\mathcal{R}_\Lambda \setminus \mathbf{r}_{\text{sub}}}^m[\ell] = \{\hat{x}_1^2[\ell], \hat{x}_5^2[\ell], \hat{x}_3^2[\ell]\}$ should be constructed to satisfy:

$$\begin{aligned} \{\hat{x}_1^2[\ell], \hat{x}_5^2[\ell], \hat{x}_3^2[\ell]\} &= \\ \{\hat{v}_{28}[\ell], \hat{v}_{30}[\ell]\} \mathbf{J}_{32}(\{28, 30\}, \{4, 20, 12\}) \end{aligned} \quad (26).$$

Therefore, the specific bits $\hat{v}_{\mathcal{U}_\Lambda}[\ell]$ must be set to $\hat{v}_{\mathcal{U}_\Lambda}[\ell] = \{\hat{v}_{28}[\ell], \hat{v}_{30}[\ell]\} \mathbf{J}_{32}(\{28, 30\}, \{4, 20, 12\}) \mathbf{J}_8^{-1}(\{1, 5, 3\}, \mathcal{U}_6)$ to ensure that Eq. (26) can be obtained.

In summary, when constructing the shortened PAC codes, the values of the frozen bits $\mathbf{v}_{\mathcal{U}_\Lambda}$ need to be calculated from the bits \mathbf{v}_T using Eq. (24). The proposed construction of shortened PAC codes is detailed in Algorithm 1.

Algorithm 1: Construction of shortened PAC codes

Input: Shortening pattern \mathcal{S} , information bits \mathbf{d} , \mathbf{J}_N and Λ

Output: Shortened PAC codes $\mathbf{x}_{\mathcal{U}_\Lambda \setminus \mathcal{S}}$

- 1 Calculate the mandatory frozen subset \mathcal{F}_{PAC} .
 - 2 Select the small set $\mathcal{B} \subseteq \mathcal{U}_\Lambda$ based on Eq. (18).
 - 3 Calculate \mathcal{I} based on Eq. (21).
 - 4 Select the information set \mathcal{A} from $\{\{\mathcal{U}_N \setminus \mathcal{U}_\Lambda\} \setminus \{\mathcal{F}_{\text{PAC}} \setminus \mathcal{I}\}\}$ via rate-profiling.
 - 5 $\mathbf{v}_{\mathcal{A}} \leftarrow \mathbf{d}$.
 - 6 Calculate $\mathbf{v}_{\mathcal{U}_\Lambda}$ from the bits \mathbf{v}_T by Eq. (24).
 - 7 Construct the mother PAC codeword by $\mathbf{x} = \mathbf{v} \mathbf{J}_N$.
-

3.3 List Decoding for Proposed Shortened PAC Codes

Section 3.2 has provided a brief discussion on the list decoding of the proposed shortened PAC coding scheme. Here, we detail the list decoding of the proposed shortened PAC codes. The decoding is provided in Algorithm 2. In the initialization of list decoding, $2^{|\mathcal{I}|}$ paths are activated. For each activated path ℓ , a unique value combination is assigned to the bits $\hat{v}_T[\ell]$, i.e., $\hat{v}_T[\ell] = \{b_{|\mathcal{I}|}, b_{|\mathcal{I}|-1}, \dots, b_1\}_j = \text{d2b}(\ell - 1)$, where $\text{d2b}(\cdot)$ refers to a decimal-to-binary converter. Thereby, the initial LLRs $L_{\mathcal{U}_\Lambda}^0[\ell]$ can be computed from $\hat{v}_T[\ell]$. Simultaneously, the specific bits $\hat{v}_{\mathcal{U}_\Lambda}[\ell]$ are calculated using Eq. (24).

Algorithm 2: List decoding of shortened PAC codes

- 1 **for** $\ell \in \mathcal{U}_{2^m}$ **do**
- 2 Calculate the variable bits $\hat{\mathbf{x}}_{\mathcal{S}}[\ell] = \hat{v}_T[\ell] \mathbf{J}_N(\mathcal{I}, \mathcal{S})$, where $\hat{v}_T[\ell] = \{b_{|\mathcal{I}|}, b_{|\mathcal{I}|-1}, \dots, b_1\}_j = \text{d2b}(\ell - 1)$.
- 3 Calculate initial LLRs $L_{\mathcal{U}_\Lambda}^0[\ell]$ according to Eq. (9).
- 4 Calculate the specific bits $\hat{v}_{\mathcal{U}_\Lambda}[\ell]$ according to Eq. (24).
- 5 **end**
- 6 //Initialization differs from conventional one.
- 7 **for** $i = 1$ **to** N **do**
- 8 **for** $\ell \in \mathcal{p}$ **do**
- 9 Derive the resulting LLR $L_i^n[\ell]$ according to Eq. (3).
- 10 **if** $i \leq \Lambda$ **or** $i \in \mathcal{I}$ **then**
- 11 $\hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$.


```

12      //No need to make decisions.
13      else if  $i \notin \mathcal{A}$  then
14           $\hat{v}_i[\ell] \leftarrow 0, \hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$ .
15      else
16           $\hat{v}_i[\ell] \leftarrow \{0, 1\}, \hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$ .
17      end
18      Calculate  $PM_{\hat{v}_i[\ell]}$  according to Eq. (4).
19  end
20  //PM-updating process is slightly changed.
21  if  $i \in \mathcal{A}$  then
22      if the number of candidate paths exceeds  $L$  then
23          Retain  $L$  paths with smaller PMs  $PM_{\hat{v}_i[\ell]}$ .
24      else
25          Retain all candidate paths.
26      end
27  end
28  //Path-killing process is unchanged.
29 end
30  $\ell^* \leftarrow \arg \min_{\ell \in \mathcal{L}} PM_{\hat{v}_i[\ell]}$ .
31 return  $\hat{v}_{\mathcal{U}_A}[\ell^*]$ .

```

In the PM-updating process of the list decoding, decisions are unnecessary for bits $\hat{v}_{\mathcal{U}_A}[\ell]$ and $\hat{v}_{\mathcal{I}}[\ell]$, because they have been already assigned during initialization. For the remaining bits, PM-updating follows the conventional list decoding for PAC codes. The path-killing process is also unchanged.

According to Algorithm 2, the decoding complexity remains dominated by the derivation of resulting LLRs, although the computation of the bits $\hat{v}_{\mathcal{U}_A}[\ell]$ in the initialization introduces a small amount of addition computation. The complexity of the decoding can still be expressed as $O(LN \log_2 N)$.

3.4 Application of Dynamic Frozen Bits to Proposed Scheme

As described in Section 3.1, DFBs can be applied to enforce $\mathcal{F}_{\text{PAC}} = \mathcal{F}_{\text{polar}}$, which can also be applied in the proposed scheme. In this case, the bits $\mathbf{v}_{\mathcal{F}_{\text{PAC}}}$ serve as DFBs. Since the subset $\mathcal{I} \subseteq \mathcal{F}_{\text{PAC}}$ is also included in the information set \mathcal{A} , the actual information bits appearing in the data carrier word are $\mathbf{v}_{\mathcal{A} \setminus \mathcal{I}}$. The remaining $|\mathcal{I}|$ information bits are implied in the specific bits $\mathbf{v}_{\mathcal{U}_A}$.

The shortened codeword is constructed as follows. First, the data carrier word \mathbf{v} is still mapped from \mathbf{d} , comprising the information bits $\mathbf{v}_{\mathcal{A}} = \mathbf{d}$ and the frozen bits $\mathbf{v}_{\mathcal{A}^c} = 0$. Assuming that \mathcal{I} consists of the e -th index in \mathcal{A} where $e \in \mathcal{E}$, i.e., $\mathcal{I} = \mathcal{A}_{\mathcal{E}}$, we have $\mathbf{v}_{\mathcal{I}} = \mathbf{d}_{\mathcal{E}}$. Second, the values of the special bits $\mathbf{v}_{\mathcal{U}_A}$ are calculated from $\mathbf{v}_{\mathcal{I}}$ via Eq. (24). Then, the values of DFBs $\mathbf{v}_{\mathcal{F}_{\text{PAC}}}$ are updated as:

$$\mathbf{v}_{\mathcal{F}_{\text{PAC}}} = \hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}}} + \mathbf{v}_{\mathcal{U}_A \setminus \mathcal{F}_{\text{PAC}}} \mathbf{K}(:, \mathcal{U}_A \setminus \mathcal{F}_{\text{PAC}})^T \quad (27),$$

where the matrix \mathbf{K} is identical to that in Eq. (8). The vector $\hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}}}$ on the right-hand side of Eq. (27) is composed of three parts: $\{\hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}} \cup \mathcal{U}_A} = \mathbf{d}_{\mathcal{E}} \mathbf{J}_N(\mathcal{I}, 2^m(\mathcal{R}_A \setminus \mathcal{r}_{\text{sub}})) \mathbf{J}_{2^m}^{-1}(\mathcal{R}_A \setminus \mathcal{r}_{\text{sub}}, \mathcal{F}_{\text{PAC}} \cup \mathcal{U}_A), \hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}} \setminus (\mathcal{I} \cup \mathcal{U}_A)} = \mathbf{0}, \hat{\mathbf{v}}_{\mathcal{I}} = \mathbf{d}_{\mathcal{E}}\}$.

Therefore, the values of the bits $\mathbf{v}_{\mathcal{I}}$ are updated, and $\mathbf{v}_{\mathcal{I}}$ do not appear as information bits. Finally, the PAC codeword is generated from \mathbf{v} and then shortened using pattern \mathcal{S} . From Eq. (27), we obtain $\mathbf{v} \mathbf{K}^T = \hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}}}$. Therefore, the value of the deleted bits is given by:

$$\mathbf{x}_{\mathcal{S}} = \hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}}} \mathbf{R}^T \quad (28),$$

where \mathbf{R} is an elementary matrix with $\mathbf{J}_N(:, \mathcal{S})^T = \mathbf{R} \mathbf{K}$. Unlike conventional DFB schemes, $\mathbf{x}_{\mathcal{S}}$ are not constrained to zeros in this design. Overall, the construction of the proposed shortened PAC codes with DFBs is summarized in Algorithm 3.

Algorithm 3: Construction of shortened PAC codes with DFBs

Input: Shortening pattern \mathcal{S} , information bits \mathbf{d} , \mathbf{J}_N and Λ

Output: Shortened PAC codes $\mathbf{x}_{\mathcal{U}_A \setminus \mathcal{S}}$

1 Execute Lines 1 to 6 of Algorithm 1.

2 Update the DFBs $\mathbf{v}_{\mathcal{F}_{\text{PAC}}}$ using Eq. (27).

3 Execute Line 7 of Algorithm 1.

Accordingly, the decoding is modified. During initialization, $2^{|\mathcal{I}|}$ paths are activated. Different from conventional decoding, the length of the data carrier word result is $N + |\mathcal{I}|$ for each path, i.e., $\hat{\mathbf{v}}_{\mathcal{U}_A \cup \mathcal{I}}[\ell]$. For each path, the specific bits $\hat{\mathbf{v}}_{\mathcal{U}_A}[\ell]$ are calculated from $\hat{\mathbf{v}}_{\mathcal{I}}[\ell]$ according to Eq. (24), where $\hat{\mathbf{v}}_{\mathcal{I}}[\ell] = \text{d2b}(\ell - 1)$. Moreover, the values of $\hat{\mathbf{v}}_{\mathcal{I}}[\ell]$ need to be copied into $\hat{\mathbf{v}}_{\mathcal{N} + \mathcal{U}_{\mathcal{I}}}[\ell]$, i.e., $\hat{\mathbf{v}}_{\mathcal{N} + \mathcal{U}_{\mathcal{I}}}[\ell] = \hat{\mathbf{v}}_{\mathcal{I}}[\ell]$. The value combination of $\hat{\mathbf{x}}_{\mathcal{S}}[\ell]$ for each path is then determined, and the initial LLRs are calculated accordingly. According to Eq. (28), we have $\hat{\mathbf{x}}_{\mathcal{S}}[\ell] = \hat{\mathbf{v}}_{\mathcal{F}_{\text{PAC}}}[\ell] \mathbf{R}^T$ for each path $\ell \in \mathcal{U}_{2^m}$.

In the PM-updating process, no decision is required when reaching the specific bits. For each DFB $\hat{v}_f[\ell], f \in \mathcal{F}_{\text{PAC}}$, where f is the j -th element in \mathcal{F}_{PAC} , its value is calculated as:

$$\hat{v}_f[\ell] = \hat{\mathbf{v}}_{\mathcal{U}_f}[\ell] \mathbf{K}(j, \mathcal{U}_f)^T \quad (29).$$

The decision processes for the frozen bits $\hat{\mathbf{v}}_{\mathcal{A}^c \setminus \mathcal{F}_{\text{PAC}}}[\ell]$ and information bits $\hat{\mathbf{v}}_{\mathcal{A} \setminus \mathcal{I}}[\ell]$ remain unchanged.

Upon completing the PM-updating and path-killing processes, the values $\hat{\mathbf{v}}_{\mathcal{N} + \mathcal{U}_{\mathcal{I}}}[\ell]$ are copied back into $\hat{\mathbf{v}}_{\mathcal{I}}[\ell]$, because $\hat{\mathbf{v}}_{\mathcal{N} + \mathcal{U}_{\mathcal{I}}}[\ell]$ match with the specific bits $\hat{\mathbf{v}}_{\mathcal{U}_A}[\ell]$. The decoding procedure is formalized in Algorithm 4.

Algorithm 4: List decoding of shortened PAC codes with DFBs

1 for $\ell \in \mathcal{U}_{2^m}$ do

```

2   $\hat{v}_T[\ell] = \hat{v}_{N+U_{in}}[\ell] = \{b_{|I|}, b_{|I|-1}, \dots, b_1\}_j = d2b(\ell-1)$ .
3  Calculate the specific bits  $\hat{v}_{U_A}[\ell]$  according to Eq. (24).
4  Calculate the variable bits by  $\hat{x}_S[\ell] = \hat{v}_{F_{PAC}}[\ell] R^T$ .
5  Calculate initial LLRs  $L_{U_A}^0[\ell]$  according to Eq. (9).
6 end
7  $j = 1$ .
8 for  $i = 1$  to  $N$  do
9   for  $\ell \in p$  do
10    Derive the resulting LLR  $L_i^n[\ell]$  according to Eq. (3).
11    if  $i \in \{U_A \setminus F_{PAC}\}$  or  $i \in \mathcal{I}$  then
12       $\hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$ .
13      //No need to make decisions.
14    else if  $i \in F_{PAC}$  then
15       $\hat{v}_i[l] \leftarrow \hat{v}_{U_A}[l] K(j, U_i)^T$ ,
16       $j \leftarrow j + 1$ ,
17       $\hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$ .
18    else if  $i \in \{A^c \setminus F_{PAC}\}$  then
19       $\hat{v}_i[\ell] \leftarrow 0$ ,  $\hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$ .
20    else
21       $\hat{v}_i[\ell] \leftarrow \{0, 1\}$ ,
22       $\hat{u}_i[\ell] \leftarrow \sum_{k=1}^s c_k \hat{v}_{i-k+1}[\ell]$ .
23    end
24    Calculate  $PM_{\hat{v}_i[\ell]}$  according to Eq. (4).
25  end
26  if  $i \in \mathcal{A}$  then
27    if the number of candidate paths exceeds  $L$  then
28      Retain  $L$  paths with smaller PMs  $PM_{\hat{v}_i[\ell]}$ .
29    else
30      Retain all candidate paths.
31    end
32  end
33 for  $\ell = 1$  to  $2^{|I|}$  do
34    $\hat{v}_T[\ell] \leftarrow \hat{v}_{N+U_{in}}[\ell]$ .
35 end
36  $\ell^* \leftarrow \arg \min_{\ell \in \mathcal{L}} PM_{\hat{v}_T[\ell]}$ .
37 return  $\hat{v}_{U_A}[\ell^*]$ .

```

4 Simulation Results

4.1 Comparison of Shortened Schemes of PAC and Polar Codes

We simulate the performance of the proposed shortened PAC codes over a binary-

input additive white Gaussian noise (BI-AWGN) channel. The set $c = (1, 0, 1, 1, 0, 1, 1)$ is applied to construct T_N . List decoding is employed with a list size $L = 8$. Parameters are set as $|I| = 3$ and $\Lambda = 28$. Since the proposed scheme can be readily adapted for polar codes, its performance under this adaptation is also simulated for comparison. We evaluate the three schemes described in Section 3.1: expanding \mathcal{F} , reconstructing T_N , and applying DFB. All simulations utilize the WANG-LIU shortening pattern^[13]. For further comparison, we also include simulation results for the rate-matching of polar codes specified in the 3GPP TS 38.212 (Release 17) standard^[2], employing both cyclic redundancy check (CRC)-8 and the standard CRC.

Fig. 1 illustrates the BLER performance under the conditions of $E = 104$ and $R = 1/3$. For PAC codes, under WANG-LIU pattern and RM rate-profiling, the reconstruct T_N and DFB schemes achieve optimal performance, yielding a gain exceeding 0.2 dB over other schemes. Under the BRS pattern and RM rate-profiling, the schemes utilizing DFB outperform those without DFB. Additionally, all proposed schemes for PAC codes surpass the CRC-Polar scheme from the TS 38.212 standard. Fig. 2 presents the BLER performance under the conditions of $E = 104$ and $R = 1/2$. For PAC codes, under the WANG-LIU pattern and RM rate-profiling, the pro-

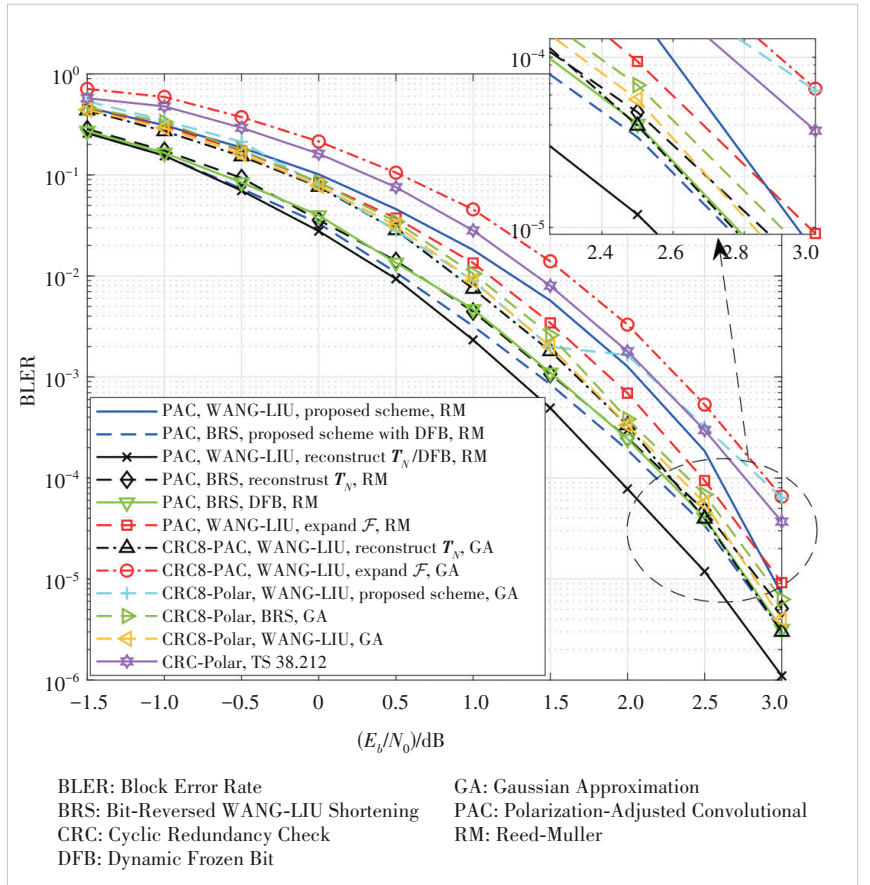


Figure 1. BLER performance of shortened PAC and polar codes with $E = 104$ and $K = 35$

posed scheme and the reconstruct T_N deliver optimal performance, offering an approximate 0.3 dB gain over the CRC-Polar scheme from the TS 38.212 standard at $\text{BLER} = 1 \times 10^{-5}$. Fig. 3 shows the BLER performance under the conditions of $E = 104$ and $R = 3/5$. For PAC codes, under the WANG-LIU pattern and RM rate-profiling, the proposed scheme and the expand \mathcal{F} achieve optimal performance, providing about a 0.25 dB gain compared to the CRC-Polar scheme from the TS 38.212 standard at $\text{BLER} = 1 \times 10^{-5}$.

Fig. 4 depicts the BLER performance under the conditions of $E \in \{84, 120, 180\}$ and $R = 1/3$. A total of six proposed shortening schemes are evaluated. Fig. 4a presents the schemes without DFB, while Fig. 4b shows those with DFB. As seen in Fig. 4a, the proposed scheme, reconstruct T_N and expand \mathcal{F} have the same property that the performance improves gradually as the block length increases at a fixed code rate. However, Fig. 4b reveals an inconsistency in the performance trend of the DFB-based schemes as the block length grows.

Based on the above analysis, a promising hybrid approach is to adopt the proposed scheme with DFB for low code rates and the proposed scheme without DFB for medium to high code rates. This combination is expected to enhance the overall performance of shortened PAC codes.

5 Conclusions

This paper explores the feasibility of applying existing shortening patterns of polar codes to PAC codes. A novel shortening scheme for PAC codes is proposed, which incorporates specially designed bits into the encoding process and introduces a minor modification to the list decoding algorithm. For short block lengths and over a portion of the low-to-medium rate range, the proposed shortened PAC coding scheme demonstrates superior BLER performance compared to existing shortened polar coding schemes, the standard rate-matching scheme, and conventional shortened PAC coding schemes.

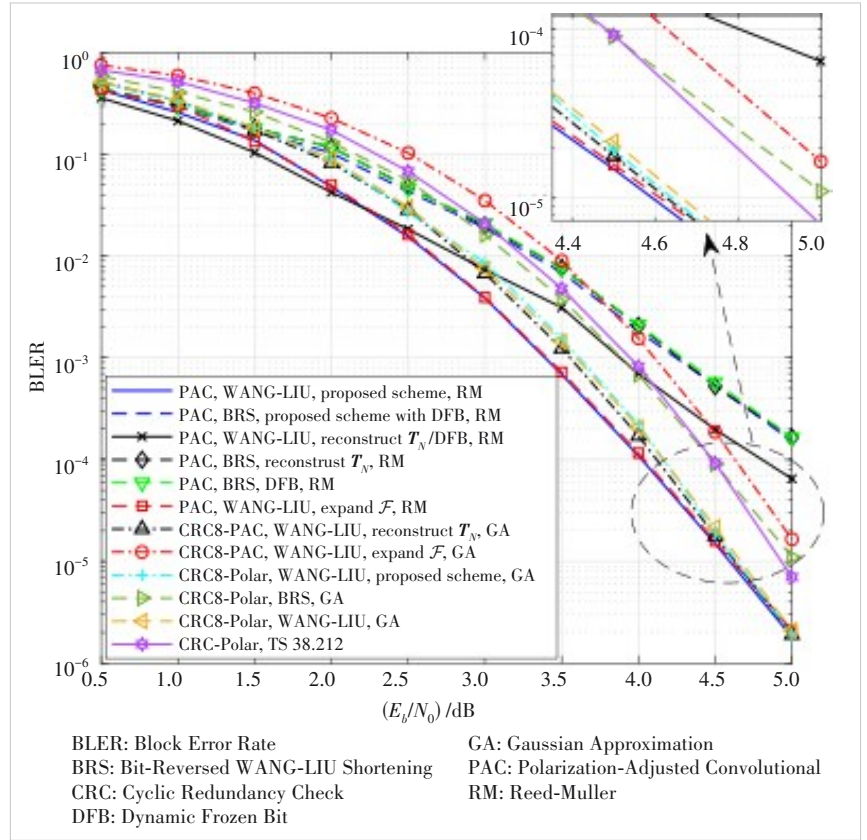


Figure 2. BLER performance of shortened PAC and polar codes with $E=104$ and $K=52$

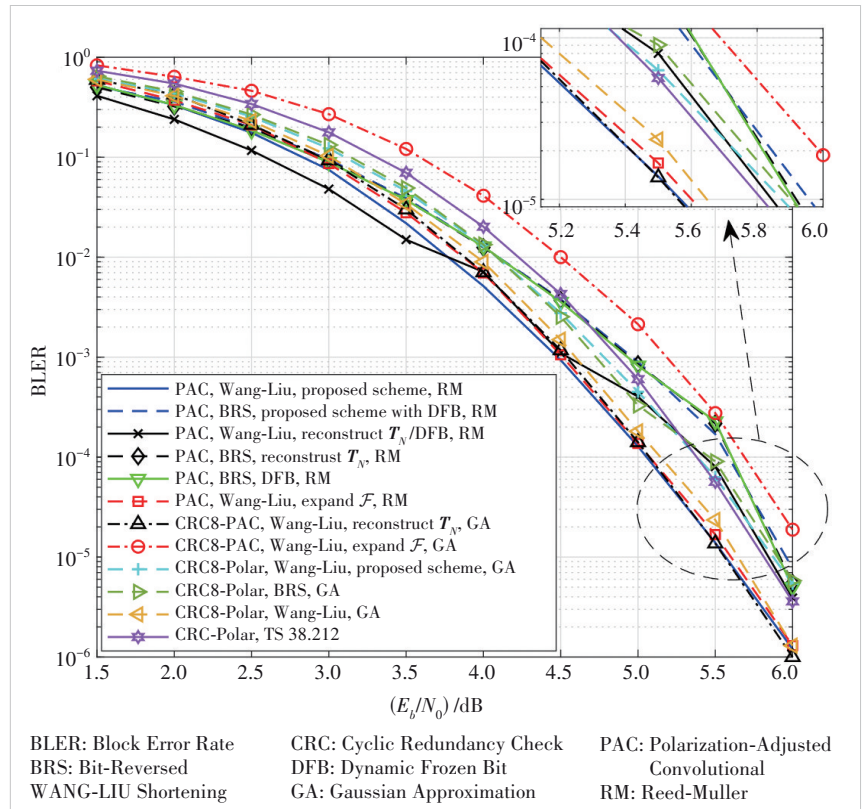
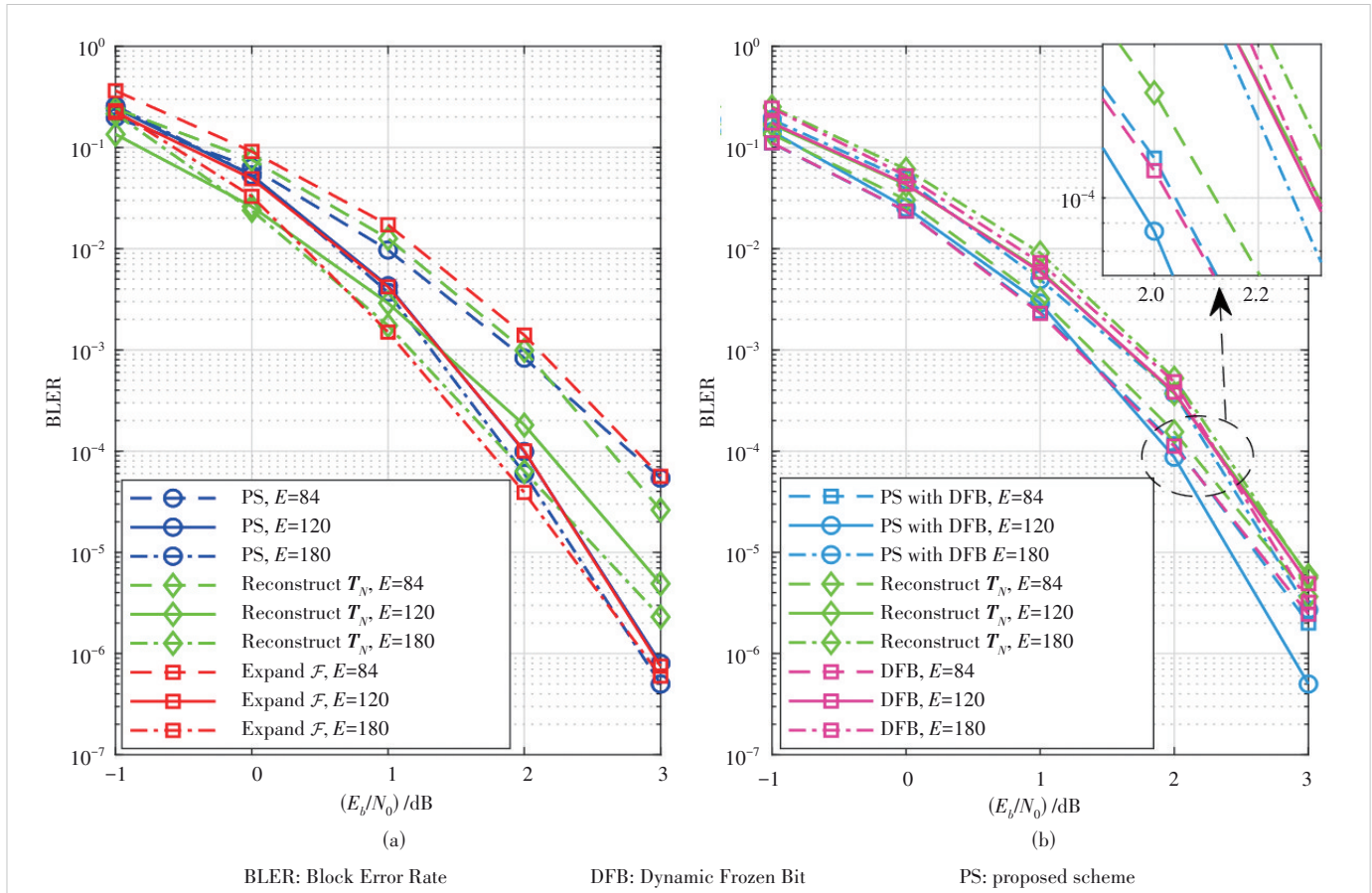


Figure 3. BLER performance of shortened PAC and polar codes with $E=104$ and $K=62$

Figure 4. BLER performance of shortened PAC codes with $R = 1/3$

References

- [1] ARIKAN E. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels [J]. IEEE transactions on information theory, 2009, 55(7): 3051 – 3073. DOI: 10.1109/TIT.2009.2021379
- [2] 3GPP. 3rd generation partnership project; technical specification group radio access network; NR; multiplexing and channel coding (release 17): TS 38.212 [S]. 2022
- [3] NIU K, DAI J, CHEN K, et al. Rate-compatible punctured polar codes: optimal construction based on polar spectra [EB/OL]. [2024-12-03]. <https://arxiv.org/abs/1612.01352>
- [4] ARIKAN E. From sequential decoding to channel polarization and back again [EB/OL]. [2024-09-09]. <https://arxiv.org/abs/1908.09594>
- [5] ROWSHAN M, BURG A, VITERBO E. Polarization-adjusted convolutional (PAC) codes: sequential decoding vs list decoding [J]. IEEE transactions on vehicular technology, 2021, 70(2): 1434 – 1447. DOI: 10.1109/TVT.2021.3052550
- [6] LIU W X, CHEN L, LIU X C. A weighted sum based construction of PAC codes [J]. IEEE communications letters, 2023, 27(1): 28 – 31. DOI: 10.1109/LCOMM.2022.3209082
- [7] NIU K, CHEN K, LIN J R. Beyond turbo codes: rate-compatible punctured polar codes [C]//Proc. IEEE International Conference on Communications (ICC). IEEE, 2013: 3423 – 3427. DOI: 10.1109/ICC.2013.6655078
- [8] LI L P, SONG W, NIU K. Optimal puncturing of polar codes with a fixed information set [J]. IEEE access, 2019, 7: 65965 – 65972. DOI: 10.1109/ACCESS.2019.2918346
- [9] SOLIMAN T, YANG F. Cooperative punctured polar coding (CPPC) scheme based on plotkin's construction [J]. Radioengineering, 2016, 25(3): 482 – 489. DOI: 10.13164/re.2016.0482
- [10] BIOGLIO V, GABRY F, LAND I. Low-complexity puncturing and shortening of polar codes [C]//Proc. IEEE Wireless Communications and Networking Conference Workshops (WCNCW). IEEE, 2017: 1 – 6. DOI: 10.1109/WCNCW.2017.7919040
- [11] ZHAO J H, ZHANG W, LIU Y Y. A novel puncturing scheme of low rate polar codes based on fixed information set [J]. IEEE communications letters, 2021, 25(7): 2104 – 2108. DOI: 10.1109/LCOMM.2021.3072050
- [12] HAN S, KIM B, HA J. Rate-compatible punctured polar codes [J]. IEEE communications letters, 2022, 26(4): 753 – 757. DOI: 10.1109/LCOMM.2022.3144695
- [13] WANG R X, LIU R K. A novel puncturing scheme for polar codes [J]. IEEE communications letters, 2014, 18(12): 2081 – 2084. DOI: 10.1109/LCOMM.2014.2364845
- [14] ARIKAN E. Systematic encoding and shortening of PAC codes [J]. Entropy, 2020, 22(11): 1301. DOI: 10.3390/e22111301
- [15] MILOSLAVSKAYA V. Shortened polar codes [J]. IEEE transactions on information theory, 2015, 61(9): 4852 – 4865. DOI: 10.1109/TIT.2015.2453312
- [16] CHAKI P, KAMIYA N. On the properties of bit-reversal shortening in polar codes [C]//Proc. International Symposium on Information Theory and Its Applications (ISITA). IEEE, 2018: 437 – 441. DOI: 10.23919/ISITA.2018.8664298
- [17] OLIVEIRA R M, DE LAMARE R C. Rate-compatible polar codes based on polarization-driven shortening [J]. IEEE communications letters, 2018, 22(10): 1984 – 1987. DOI: 10.1109/LCOMM.2018.2863375
- [18] OLIVEIRA R M, DE LAMARE R C. Design of rate-compatible polar codes based on non-uniform channel polarization [J]. IEEE access, 2021, 9: 41902 – 41912. DOI: 10.1109/ACCESS.2021.3065816

- [19] JANG M, AHN S K, JEONG H, et al. Rate matching for polar codes based on binary domination [J]. IEEE transactions on communications, 2019, 67(10): 6668 – 6681. DOI: 10.1109/TCOMM.2019.2930502
- [20] ROWSHAN M, VITERBO E. On convolutional precoding in PAC codes [C]//Proceedings of IEEE Globecom Workshops (GC Wkshps). IEEE, 2021: 1 – 6. DOI: 10.1109/gcwkshps52748.2021.9681987
- [21] SUN H, VITERBO E, LIU R K. Analysis of polarization-adjusted convolutional codes (PAC): a source-channel coding method [C]//Proc. IEEE Globecom Workshops (GC Wkshps). IEEE, 2021: 1 – 6. DOI: 10.1109/gcwkshps52748.2021.9682079
- [22] FENG B W, YANG Y, JIAO J, et al. On tail-biting polarization-adjusted convolutional (TB-PAC) codes and small-sizes list decoding [J]. IEEE communications letters, 2023, 27(2): 433 – 437. DOI: 10.1109/LCOMM.2022.3225177
- [23] ROWSHAN M, DAU S H, VITERBO E. On the formation of min-weight codewords of polar/PAC codes and its applications [J]. IEEE transactions on information theory, 2023, 69(12): 7627 – 7649. DOI: 10.1109/TIT.2023.3319015

Biographies

LIU Aolin received his BS degree in communication engineering from Harbin Institute of Technology (Shenzhen), China in 2023, where he is currently pursuing the PhD degree in information and communication engineering. He is also affiliated with Pengcheng Laboratory, China. His research interests include coding theory, error control codes, and satellite communications.

FENG Bowen (fengbowen@hit.edu.cn) received his BS, MS, and PhD degrees in information and communication engineering from the Harbin Institute of Technology, China in 2014, 2016, and 2021, respectively. From 2021 to 2024,

he was a postdoctoral research fellow with Harbin Institute of Technology (Shenzhen) (HITSZ), China. He is currently an associate professor with the Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, HITSZ. His research interests include error control codes, satellite communications, and machine-to-machine communications.

LIANG Chulong received his BE degree in communication engineering and PhD degree in communication and information systems from Sun Yat-sen University, China in 2010 and 2015, respectively. He was a post-doctoral fellow with the Department of Electronic Engineering, City University of Hong Kong, China from July 2015 to June 2018, where he was a research fellow, from June 2018 to May 2019. He is currently a senior engineer with the Algorithm Department, ZTE Corporation. His current research interests include channel coding theory and its applications to communication systems.

XU Jin received his PhD degree from the School of Electronic Information, Wuhan University, China. He is currently a senior pre-research engineer with ZTE Corporation. His research interests include error correction codes, digital signal processing, semantic communication, and machine learning.

ZHANG Qinyu received his bachelor's degree in communication engineering from the Harbin Institute of Technology (HIT), China in 1994 and PhD degree in biomedical and electrical engineering from the University of Tokushima, Japan in 2003. From 1999 to 2003, he was an assistant professor with the University of Tokushima. He has been with Harbin Institute of Technology (Shenzhen) (HITSZ), China since 2003; he is currently a full professor and serves as the Vice President of HITSZ. He has received the National Science Fund for Distinguished Young Scholars, and been recognized as a Young and Middle-Aged Leading Scientist of China and a Chinese New Century Excellent Talent in University, among other honors. He has also obtained three scientific and technological awards from government authorities. His research interests include aerospace communications and networks, and wireless communications and networks.



Full-Duplex Massive MIMO Self-Interference Suppression Based on Beamforming

ZHANG Boyu, ZHANG Ling, LI Zijiang, SHEN Ying

(University of Electronic Science and Technology of China, Chengdu 611731, China)

DOI: 10.12142/ZTECOM.202504011

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251107.1127.004.html>,
published online November 7, 2025

Manuscript received: 2024-04-10

Abstract: The complexities of hardware and signal processing make it especially challenging to develop self-interference cancellation (SIC) techniques for full-duplex (FD) massive multiple-input-multiple-output (MIMO) systems. This paper examines an FD massive MIMO system featuring multi-stream transmission. Specifically, it adopts an architecture where a single transmit or receive radio frequency (RF) channel is connected to three antennas in the same polarization direction, effectively reducing the number of transmit and receive RF channels by half. The SoftNull algorithm serves as the primary method for SI suppression, leveraging digital precoding during transmission. Additionally, this study outlines a design strategy to enhance SIC in the proposed system. Simulation results highlight the efficacy of the SoftNull algorithm, which achieves a remarkable total SIC of up to 64 dB. Furthermore, combined with measures such as antenna isolation and increased transceiver array spacing, the resulting sum rate can be twice that of a half-duplex system.

Keywords: full-duplex; massive MIMO; self-interference; beamforming; precoding

Citation (Format 1): ZHANG B Y, ZHANG L, LI Z J, et al. Full-duplex massive MIMO self-interference suppression based on beamforming [J]. *ZTE Communications*, 2025, 23(4): 97 – 109. DOI: 10.12142/ZTECOM.202504011

Citation (Format 2): B. Y. Zhang, L. Zhang, Z. J. Li, et al., “Full-duplex massive MIMO self-interference suppression based on beamforming,” *ZTE Communications*, vol. 23, no. 4, pp. 97 – 109, Sept. 2025. doi: 10.12142/ZTECOM.202504011.

1 Introduction

In recent years, the full-duplex (FD) and massive multiple-input-multiple-output (MIMO) technologies have emerged as key research focuses in wireless communications^[1-4].

As a key technology of 5G, FD technology allows radios to simultaneously transmit and receive in the same frequency band, which theoretically doubles the spectrum utilization rate compared with the traditional duplex system^[5-6]. As another key technology of 5G, massive MIMO can improve the spectral efficiency and reliability of a system by equipping a large number of antennas^[7-11]. In 5G wireless technology implementations, the combination of various technologies has become a trend^[12-13]. Both FD and massive MIMO technologies have the characteristics of high spectral efficiency, so the combination of the two can further improve the spectral efficiency.

In FD communications, signals sent by the wireless transceiver bring self-interference (SI) to its reception^[14]. Strong SI signals may overwhelm useful information and cause errors in the received data. For a low noise amplifier (LNA) at the receive radio frequency (RF) channel, it may also exceed the dynamic range of its input, bringing harm to receiver hardware. Therefore, SI cancellation (SIC) is critical for realizing FD communications^[15]. For FD massive MIMO systems, each re-

ceive antenna will be interfered by multiple transmit antennas, and SIC becomes more challenging as the number of antennas increases.

Numerous SIC algorithms have been proposed to solve the issues of strong SI in FD massive MIMO systems. In Ref. [16], an adaptive filter structure with analog least mean square (ALMS) loops is proposed, which only realizes SIC in the analog domain. Although the number of taps used is greatly reduced, additional taps are still needed and the hardware complexity is high. To further reduce hardware complexity, some studies have put forward a combination of digital beamforming and analog SIC to reduce SI. For FD MIMO systems, Ref. [17] introduces a novel analog SI canceller which utilizes the flexible signal routing of Mux/DeMux to reduce the taps' number in the analog SI canceller. It also develops a novel optimization framework for the joint design of the analog canceller and TX/RX digital beamforming parameters. The burden of SIC is shared between digital beamforming and analog SI cancellers, which greatly reduces the hardware complexity. In Ref. [18], a joint design framework of TX-RX beamforming and SIC scheme based on the minimum mean square error (MMSE) is introduced to realize SIC under the constraint of analog canceller tap limits. Hybrid precoding/combining (HPC) techniques have also been discussed recently. A novel angular-

based joint hybrid precoding/combining technique is explored in Ref. [19] for FD massive MIMO systems, combining transmit/receive RF beamformers and receiver combiners to achieve SIC. In Ref. [20], the authors propose a hybrid beamforming design for FD massive MIMO systems, which jointly designs the transmit RF beamformer, the precoder, and the combiner to suppress SI. However, all the HPC-based SIC technologies rely on RF beamformers, resulting in an increased signal processing complexity. In Ref. [21], a SIC algorithm, SoftNull, is proposed that uses only transmit digital precoding. It requires no additional analog SI cancellers, nor does it need to be designed with an RF beamformer, which drastically reduces both hardware and signal processing complexity.

Focusing on the multi-stream FD massive MIMO communication system^[19], this paper studies a SIC method based on a 3D geometry-based millimeter wave (mmWave) channel model. In the proposed system model, the transmit and receive antennas are all $\pm 45^\circ$ polarized, and one transmit or receive RF channel is connected with three transmit or receive antennas in the same polarization direction. The number of required RF channels is reduced by half, greatly reducing the volume and cost. At the same time, the 3D geometry-based mmWave channel model is used to model the intended and the SI channels. In view of the above system architecture, the SoftNull algorithm based only on transmit digital precoding is selected to realize SIC, which requires neither additional hardware structure nor RF beamformer processing. Specifically, the proposed method takes advantage of the high degree of freedom provided by large antenna arrays to split the conventional transmit digital precoding matrix into two submatrices: one for traditional digital precoding and the other for SIC. In this way, it can minimize the total SI power while preserving a certain degree of freedom for traditional digital precoding.

In addition, concerning the proposed system architecture, the method of improving SIC ability using the SoftNull algorithm is analyzed theoretically. The analysis results indicate that the ability of SIC can be enhanced by measures such as designing polarization isolation degree between transceiver arrays with different polarization directions, changing the array layout, and increasing the size of transmit/receive uniform rectangular arrays (URA). Simulation results demonstrate that the SoftNull algorithm can achieve up to 64 dB of SIC under the proposed system architecture. To achieve higher system performance, joint antenna isolation and the SoftNull algorithm are proposed. Simulation verifies that when the antenna isolation is increased to 50 dB, the sum rate is 1.7 times higher than that of the standalone SoftNull algorithm and twice as high as the half-duplex algorithm at low signal-to-noise ratios (SNRs). Furthermore, system performance can be further improved by increasing the size of transmit/receive URAs. Simulation results show that the sum rate of the SoftNull algorithm can be improved by up to 54% compared with the same antenna isolation when the size of the transmit/re-

ceive URA is increased to 5 wavelengths, and it remains nearly double that of the half-duplex state at high SNRs.

The rest of this article is organized as follows. Section 2 presents the system model and channel model. In Section 3, a SoftNull algorithm is introduced and the method of enhancing the SIC effect is analyzed theoretically. Section 4 demonstrates the effectiveness of the algorithm under the proposed system architecture, and simulations are carried out by further improving the system performance. Finally, Section 5 concludes this paper.

2 System Model and Channel Model

We consider a system model for the FD massive MIMO system, where the base station (BS) has a precoder (with a precoding matrix $\mathbf{V}_{\text{BS}} \in \mathbf{C}^{N_{\text{RF},t} \times K_{\text{Down}}}$) and a combiner (with a matrix $\mathbf{U}_{\text{BS}} \in \mathbf{C}^{K_{\text{Up}} \times N_{\text{RF},r}}$), as shown in Fig. 1.

The transmitter and the receiver of BS, which operate in a full-duplex mode, are equipped with N_t and N_r antennas, and by $N_{\text{RF},t}$ and $N_{\text{RF},r}$ RF chains, respectively. Each transmit or receive RF chain is attached to three dedicated transmit or receive antenna elements via a power combiner or power divider, where $N_{\text{RF},t} = N_t/3$, $N_{\text{RF},r} = N_r/3$. Each power divider is applied to make three copies of the signals processed at a TX RF chain before these signals enter the TX antenna elements. Similarly, each power combiner is used to add the signals received by three RX antenna elements before these signals enter the RX RF chain. The proposed system decreases the hardware cost and complexity of traditional FD massive MIMO systems, due to the number of RF channels being reduced from $N_t + N_r$ to $N_{\text{RF},t} + N_{\text{RF},r} = (N_t + N_r)/3$. The half-duplex (HD) single-antenna users in the system are divided into K_{Up} users transmitting in the uplink mode and K_{Down} users receiving in the downlink mode, which are denoted as nodes “Up” and nodes “Down”, respectively, where $K_{\text{Up}}, K_{\text{Down}} \leq \{N_{\text{RF},t}, N_{\text{RF},r}\}$.

The transmitter of BS employs both traditional digital precoding and SI suppression using the matrices \mathbf{B} and \mathbf{A} of size $f_D \times K_{\text{Down}}$ and $N_{\text{RF},t} \times f_D$ to process the baseband signal $\mathbf{S}_{\text{BS},t} \in \mathbf{C}^{K_{\text{Down}} \times 1}$ in the frequency domain, respectively. The beamformed signal undergoes $N_{\text{RF},t}$ transmission RF channels and the N_t transmission antenna, and this generates as a downlink transmission signal \mathbf{S}_d . f_D represents the degree of freedom available for traditional digital precoding, that is, the degree of freedom allocated for downlink traditional beamforming. Additionally, it should be guaranteed that $K_{\text{Down}} \leq f_D \leq N_{\text{RF},t}$. The downlink transmission signal can be mathematically expressed as:

$$\mathbf{x}_{\text{BS},t} = \mathbf{C}_t \mathbf{F}^H \mathbf{V}_{\text{BS}} \mathbf{S}_{\text{BS},t} \quad (1),$$

where $\mathbf{V}_{\text{BS}} = \mathbf{A}\mathbf{B} \in \mathbf{C}^{N_{\text{RF},t} \times K_{\text{Down}}}$ is the equivalent transmission digital precoding matrix, and $\mathbf{F} \in \mathbf{C}^{N_{\text{RF},t} \times N_{\text{RF},t}}$ denotes the fast Fourier transform matrix. The elements $[\mathbf{F}]_{m,k}$ with $m, k =$

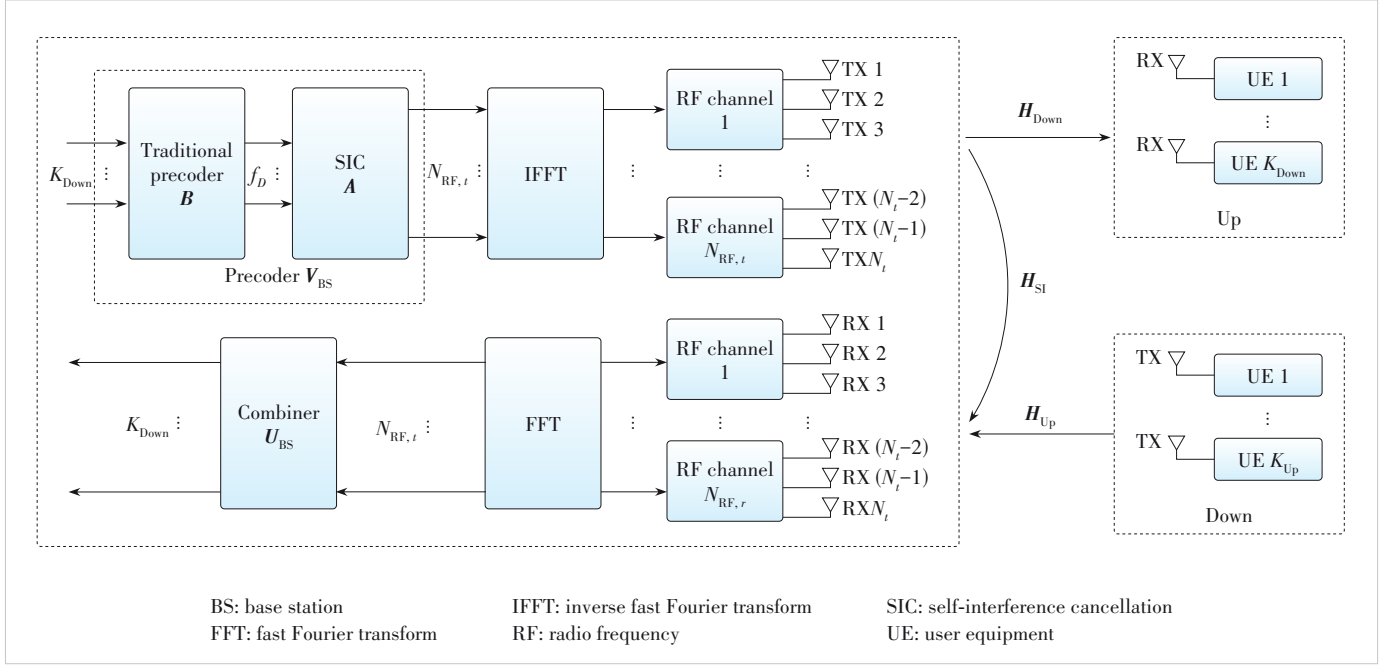


Figure 1. FD massive MIMO communication system with multi-stream transmission

$0, 1, \dots, N-1$ take $e^{-j2\pi mk/N}/\sqrt{N}$. $\mathbf{C}_t \in \mathbf{C}^{N_t \times N_{RF,t}}$ represents the matrix from the transmitting RF channel to the TX antenna on the BS side, which is obtained as:

$$\mathbf{C}_t = \frac{1}{3} \mathbf{E}_{N_{RF,t}} \otimes [\mathbf{E}_{N_{t,x}}, \mathbf{E}_{N_{t,y}}, \mathbf{E}_{N_{t,z}}]^T \quad (2)$$

where \mathbf{E}_N denotes the N -th order unit array.

As shown in Fig. 1, when the BS transmits data to the node Down and the node Up transmits data to the BS, the corresponding intended channel matrix between them is denoted as $\mathbf{H}_{\text{Down}} \in \mathbf{C}^{K_{\text{Down}} \times N_t}$ and $\mathbf{H}_{\text{Up}} \in \mathbf{C}^{N_r \times K_{\text{Up}}}$, respectively. Moreover, $\mathbf{H}_{\text{SI}} \in \mathbf{C}^{N_r \times N_t}$ represents the SI channel at the BS due to the FD transmission. It is assumed that there is no interference between the nodes Down and Up. We model both the intended channel and the SI channel in the following content.

Fig. 2 illustrates the application scenario, where the transmitter and receiver employ a uniform rectangular array (URA) with $\pm 45^\circ$ polarization on the BS side. For convenience, we use polarization 1 and 2 to denote polarization $+45^\circ$ and -45° . The transmit or receive antennas with different polarizations are placed in the same position. The number of transmit or receive antennas with polarization k is $N_u^k = N_{u,x} \times N_{u,y}$, $u \in \{t, r\}$, $k \in \{1, 2\}$, where $u \in \{t, r\}$ represents either the transmitter side for $u = t$ or the

receiver side for $u = r$. Here, $N_{u,x}$ and $N_{u,y}$ represent the antennas along the x -axis and the y -axis, respectively, where $N_t^1 = N_t^2$ and $N_r^1 = N_r^2$.

The transmit and receive URAs of BS located in the x - o - y plane are placed next to each other, where d_1 represents the distance between the transmit and receive URAs along the x -axis.

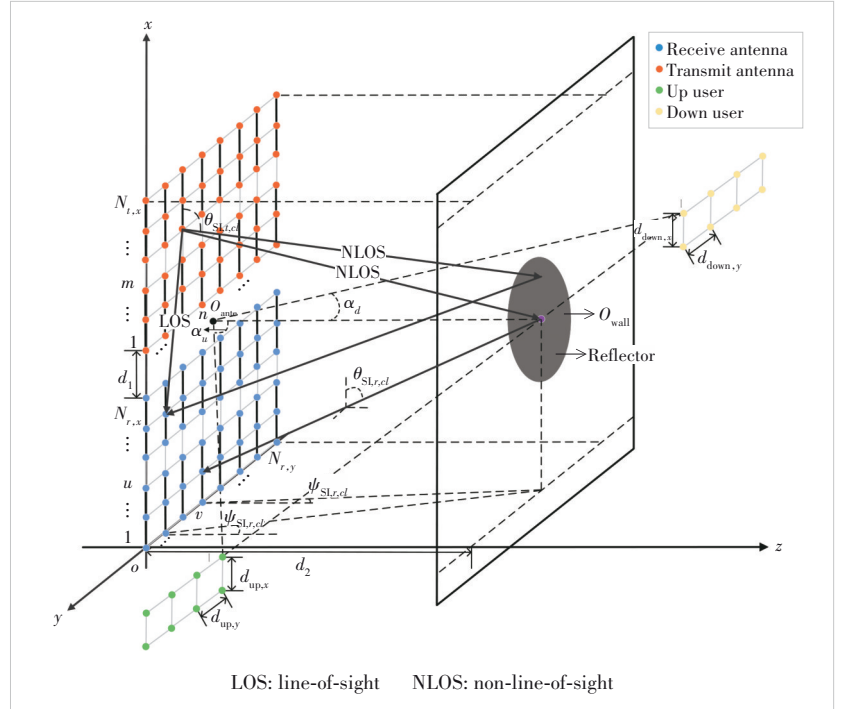


Figure 2. Application scenarios of FD massive MIMO system

Consider a wall parallel to the transmit and receive URAs, with a distance d_2 between them. The wall's center point o_{ante} is at the same height as the center point o_{ante} of transmit and receive URAs. Both the nodes Down and V_{BS} are located on the plane of the wall, which are respectively located on the positive y -axis and the negative y -axis. Uplink user 1 and downlink user 1 are at the same height as o_{ante} at the angle of α_u and α_d .

We assume that there are $K_{\text{Up},x}$ and $K_{\text{Up},y}$ uplink users regularly arranged along the negative x -axis and the positive y -axis with an interval of $d_{\text{Up},x}$ and $d_{\text{Up},y}$. Similarly, there are $K_{\text{Down},x}$ and $N_{\text{RF},d}$ uplink users regularly arranged along the negative x -axis and the positive y -axis with the intervals of $d_{\text{Down},x}$ and $d_{\text{Down},y}$, where $K_{\text{Up}} = K_{\text{Up},x} \times K_{\text{Up},y}$, $\mathbf{H}_{\text{equ}} = \mathbf{H}_{\text{Down},\text{equ}} \mathbf{A}$.

Fig. 3 shows the structure of transmit and receive URAs, where each transmit or receive RF chain is connected to three dedicated transmit or receive antenna elements in the same polarization. $N_{\text{RF},u}^k$, $u \in \{t, r\}$, $k \in \{1, 2\}$ indicates the number of transmit or receive RF chains attached to the antenna with polarization k , where $N_{\text{RF},t}^1 = N_{\text{RF},t}^2$ and $N_{\text{RF},r}^1 = N_{\text{RF},r}^2$.

$N_{t,x}$ and $N_{t,y}$ transmit antenna elements in the same polarization are arranged with equal intervals $d_{t,x}$ and $d_{t,y}$ along the positive x -axis and the negative y -axis, respectively. Similarly, $N_{r,x}$ and $N_{r,y}$ transmit antenna elements in the same polarization are arranged with equal intervals $d_{r,x}$ and $d_{r,y}$ along the positive x -axis and the negative y -axis, respectively.

Intended channels are modeled using the 3D geometry-based mmWave channel model and the system application scenario. We assume that the intended channels contain C_i scattering clusters and $L_{i,c}$ paths in the c -th cluster with $c = 1, 2, \dots, C_i$. There are $L_i = \sum_{c=1}^{C_i} L_{i,c}$ paths between the transmitter and receiver in total. The intended channel matrix is derived as:

$$\mathbf{H}_i = \sum_{c=1}^{C_i} \sum_{l=1}^{L_{i,c}} \tau_{i,c,l}^{-\eta} g_{i,c,l} \phi_{r,j}(\gamma_{r,c,l}^{(x)}, \gamma_{r,c,l}^{(y)}) \phi_{t,i}^H(\gamma_{t,c,l}^{(x)}, \gamma_{t,c,l}^{(y)}) = \Phi_{r,j} \mathbf{G}_i \Phi_{t,i}^H \quad (3),$$

where $\mathbf{H}_i \in \mathbb{C}^{R \times T}$, $(R, T) \in \{(K_{\text{Down}}, N_t), (N_r, K_{\text{Up}})\}$; j and i are the receive and transmit nodes of the intended channel, satisfying $(j, i) \in \{(\text{Down}, \text{BS}), (\text{BS}, \text{Up})\}$. When $(j, i) = (\text{Down}, \text{BS})$, there is $(R, T) = (K_{\text{Down}}, N_t)$. When $(j, i) = (\text{BS}, \text{Up})$, there is $(R, T) = (N_r, K_{\text{Up}})$. $\mathbf{G}_i = \text{diag}(\tau_{i,1}^{-\eta} g_{i,1}, \dots, \tau_{i,L_i}^{-\eta} g_{i,L_i}) \in \mathbb{C}^{L_i \times L_i}$ represents the diagonal path

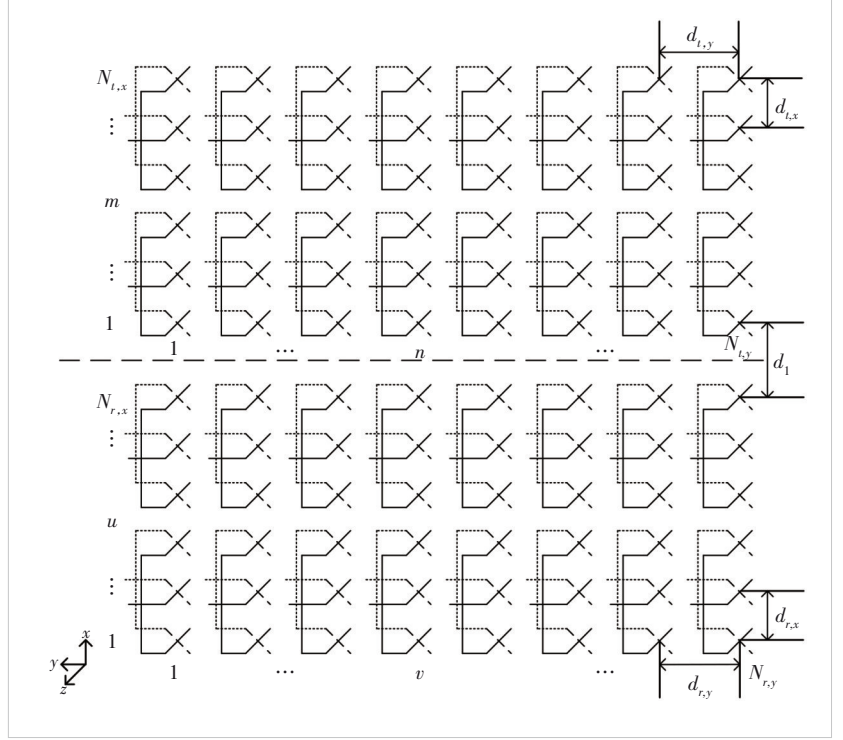


Figure 3. Base station transceiver's uniform rectangular array structure

gain matrix, $\tau_{i,c,l}$ and $g_{i,c,l} \sim \mathcal{CN}(0, 1/L_i)$ denote the distance and complex path gain of the l -th path in the c -th cluster, respectively. η is the path loss exponent. $\Phi_{r,j} \in \mathbb{C}^{R \times L_i}$ and $\Phi_{t,i} \in \mathbb{C}^{L_i \times T}$ respectively represent the receive and transmit phase response matrices of intended channels, which is given by:

$$\Phi_{r,j} = \begin{bmatrix} \phi_{r,j}^H(\gamma_{r,1}^{(x)}, \gamma_{r,1}^{(y)}) \\ \vdots \\ \phi_{r,j}^H(\gamma_{r,L_i}^{(x)}, \gamma_{r,L_i}^{(y)}) \end{bmatrix}^H \quad (4),$$

$$\Phi_{t,i} = \begin{bmatrix} \phi_{t,i}^H(\gamma_{t,1}^{(x)}, \gamma_{t,1}^{(y)}) \\ \vdots \\ \phi_{t,i}^H(\gamma_{t,L_i}^{(x)}, \gamma_{t,L_i}^{(y)}) \end{bmatrix} \quad (5),$$

where $\phi_{u,i}(\gamma_{u,c,l}^{(x)}, \gamma_{u,c,l}^{(y)}) \in \mathbb{C}^{U \times 1}$, $(U, u) \in \{(R, r), (T, t)\}$ denotes the phase response vector of the l -th path in the c -th cluster, which can be expressed as:

$$\phi_{u,i}(\gamma_x, \gamma_y) = \begin{bmatrix} 1, & e^{j2\pi d_{u,x}\gamma_x}, & \dots, & e^{j2\pi d_{u,x}(N_{u,x}-1)\gamma_x} \end{bmatrix}^T \otimes \begin{bmatrix} 1, & e^{j2\pi d_{u,y}\gamma_y}, & \dots, & e^{j2\pi d_{u,y}(N_{u,y}-1)\gamma_y} \end{bmatrix}^T \quad (6).$$

$\gamma_{r,c_l}^{(x)}$ and $\gamma_{r,c_l}^{(y)}$ include the elevation angle-of-arrival (AoA) and azimuth AoA information of the receive antenna of the l -th path in the c -th cluster at node j , which can be written as follows:

$$\gamma_{r,c_l}^{(x)} = \sin(\theta_{r,c_l})\cos(\psi_{r,c_l}) \quad (7),$$

$$\gamma_{r,c_l}^{(y)} = \sin(\theta_{r,c_l})\sin(\psi_{r,c_l}) \quad (8).$$

The elevation AoA θ_{r,c_l} and azimuth AoA ψ_{r,c_l} are the angle between the path and the positive x -axis and that between the projection of the path on the y - o - z plane and the positive z -axis, respectively. $\gamma_{t,c_l}^{(x)}$ and $\gamma_{t,c_l}^{(y)}$ indicate the elevation angles of departure (AoD) and azimuth AoD of the transmit antenna of the l -th path in the c -th cluster at the node i , and can be mathematically expressed as:

$$\gamma_{t,c_l}^{(x)} = \sin(\theta_{t,c_l})\cos(\psi_{t,c_l}) \quad (9),$$

$$\gamma_{t,c_l}^{(y)} = \sin(\theta_{t,c_l})\sin(\psi_{t,c_l}) \quad (10),$$

where θ_{t,c_l} and ψ_{t,c_l} respectively represent the elevation AoD and azimuth AoD of the l -th path in the c -th cluster.

According to the application scenario in Fig. 2, SI channels can be divided into the residual near-field SI channel $\mathbf{H}_{\text{LoS}} \in \mathbf{C}^{N_r \times N_t}$ and the far-field SI channel $\mathbf{H}_{\text{NLoS},i} \in \mathbf{C}^{N_r \times N_t}$ after antenna isolation, which can be obtained as:

$$\mathbf{H}_{\text{SI}} = \mathbf{H}_{\text{LoS}} + \mathbf{H}_{\text{NLoS}} \quad (11).$$

With the transmit and receive URA structure in Fig. 3, \mathbf{H}_{LoS} is derived as:

$$\mathbf{H}_{\text{LoS}} = \begin{bmatrix} \mathbf{H}_{\text{LoS}}^{11} & \mathbf{H}_{\text{LoS}}^{12} \\ \mathbf{H}_{\text{LoS}}^{21} & \mathbf{H}_{\text{LoS}}^{22} \end{bmatrix} \quad (12),$$

where $\mathbf{H}_{\text{LoS}}^{gk} \in \mathbf{C}^{N_r^g \times N_t^k}$, $g, k \in \{1, 2\}$ represents the residual near-field SI channel between the element with polarization g of receive array and the element with polarization k of the transmit array. We assume that the antenna polarization isolation between each transceiver array element is $\mathbf{P}_{\text{XPI,dB}} = -10\log_{10}(\mathbf{K}) \in \mathbf{C}^{N_r^g \times N_t^k}$, $g \neq k$. Eq. (12) is equivalent to:

$$\mathbf{H}_{\text{LoS}} = \begin{bmatrix} \mathbf{H}_{\text{LoS}}^{11} & \mathbf{K} \odot \mathbf{H}_{\text{LoS}}^{11} \\ \mathbf{K} \odot \mathbf{H}_{\text{LoS}}^{11} & \mathbf{H}_{\text{LoS}}^{11} \end{bmatrix} \quad (13).$$

The residual near-field SI channel $\mathbf{H}_{\text{LoS}}^{11}$ between transmit and receive array elements with polarization 1 is detailed in the following. The element of the $[(u-1)N_{r,y} + v]$ row and $[(m-1)N_{t,y} + n]$ column of $\mathbf{H}_{\text{LoS}}^{11}$, which is the channel between the (m,n) -th transmitter and (u,v) -th receiver in Fig. 3,

can be given by

$$\mathbf{H}_{\text{LoS}}^{11}([(u-1)N_{r,y} + v], [(m-1)N_{t,y} + n]) = \frac{\kappa_1}{\Delta_{(m,n) \rightarrow (u,v)}} e^{-j2\pi\Delta_{(m,n) \rightarrow (u,v)}} \quad (14),$$

where κ_1 is a constant satisfying $10\log\{\mathbb{E}[\|\mathbf{H}_{\text{LoS}}^{11}\|_F^2]\} = -P_{\text{IS,dB}}$; $P_{\text{IS,dB}}$ is the antenna isolation, and $\Delta_{(m,n) \rightarrow (u,v)}$ is the distance from the (m,n) -th transmitter to the (u,v) -th receiver. $\Delta_{(m,n) \rightarrow (u,v)}$ is defined as follows:

$$\Delta_{(m,n) \rightarrow (u,v)} = \sqrt{[(m-1)d_{t,x} + (u-1)d_{r,x} + D_1]^2 + (nd_{t,y} - vd_{r,y})^2} \quad (15).$$

Similarly, the far-field SI channel can be obtained as

$$\mathbf{H}_{\text{NLoS}} = \begin{bmatrix} \mathbf{H}_{\text{NLoS}}^{11} & \mathbf{K} \odot \mathbf{H}_{\text{NLoS}}^{11} \\ \mathbf{K} \odot \mathbf{H}_{\text{NLoS}}^{11} & \mathbf{H}_{\text{NLoS}}^{11} \end{bmatrix} \quad (16),$$

where $\mathbf{H}_{\text{NLoS}}^{11} \in \mathbf{C}^{N_t^1 \times N_r^1}$ represents the far-field SI channel between the receive and transmit array elements with polarization 1. We utilize the 3D geometry-based mmWave channel to model $\mathbf{H}_{\text{NLoS}}^{11}$. Suppose there are C_{SI} scattering clusters, and c cluster contains $L_{\text{SL},c}$ non-line-of-sight (NLOS) paths with $c = 1, 2, \dots, C_{\text{SI}}$. In total, there are $L_{\text{SI}} = \sum_{c=1}^{C_{\text{SI}}} L_{\text{SL},c}$ paths between the transmitter and receiver. According to Eq. (3), it can be mathematically expressed as

$$\mathbf{H}_{\text{NLoS}}^{11} = \Phi_{\text{SI},r} \mathbf{G}_{\text{SI}} \Phi_{\text{SI},t} \quad (17),$$

where $\mathbf{G}_{\text{SI}} = \text{diag}(\tau_{\text{SI},1}^{-\eta} g_{\text{SI},1}, \dots, \tau_{\text{SI},L_{\text{SI}}}^{-\eta} g_{\text{SI},L_{\text{SI}}}) \in \mathbf{C}^{L_{\text{SI}} \times L_{\text{SI}}}$ denotes the diagonal path gain matrix, τ_{SI,c_l} and $g_{\text{SI},l} \sim \mathcal{CN}(0, 1/L_{\text{SI}})$ respectively represent the distance and complex path gain of the l -th path in the c -th cluster. $\Phi_{\text{SI},r} \in \mathbf{C}^{N_r \times L_{\text{SI}}}$ and $\Phi_{\text{SI},t} \in \mathbf{C}^{L_{\text{SI}} \times N_t}$ denote the receive and the transmit phase response matrices of the far-field SI channel, respectively. We assume θ_{SI,r,c_l} and ψ_{SI,r,c_l} indicate elevation AoA and azimuth AoA of the l -th NLOS path in the c -th cluster. θ_{SI,t,c_l} and ψ_{SI,t,c_l} respectively represent elevation AoD and azimuth AoD of the l -th NLOS path in the c -th cluster. Therefore, Eq. (11) is equivalent to:

$$\mathbf{H}_{\text{SI}} = \begin{bmatrix} \mathbf{H}_{\text{SI}}^{11} & \mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11} \\ \mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11} & \mathbf{H}_{\text{SI}}^{11} \end{bmatrix} \quad (18),$$

where $\mathbf{H}_{\text{SI}}^{11} = \mathbf{H}_{\text{LoS}}^{11} + \mathbf{H}_{\text{NLoS}}^{11} \in \mathbf{C}^{N_t^1 \times N_r^1}$ is the SI channel between transmit and receive array elements with polarization 1.

In the downlink transmission, the received signals $\mathbf{y}_{\text{UE},r} \in \mathbf{C}^{K_{\text{down}} \times 1}$ at nodes Down are obtained as:

$$\mathbf{y}_{\text{UE},r} = \mathbf{H}_{\text{Down}} \mathbf{x}_{\text{BS},t} + \mathbf{n}_{\text{UE}} \quad (19),$$

where $\mathbf{n}_{\text{UE}} \sim \mathcal{CN}(0, \sigma_{\text{UE}}^2 \mathbf{I}_{K_{\text{Down}}}) \in \mathbb{C}^{K_{\text{Down}} \times 1}$ denotes the additive white Gaussian noise (AWGN) vector at nodes Down.

The BS simultaneously transmits messages to nodes Down in the downlink and receives information from nodes Up in the uplink in the FD mode. Hence, the received signal vector $\mathbf{y}_{\text{BS},r} \in \mathbb{C}^{N_r \times 1}$ at the receiver of BS includes the downlink transmission signals $\mathbf{x}_{\text{BS},d}$ through the SI channels and the uplink transmission signals $\mathbf{x}_{\text{UE},d} \in \mathbb{C}^{K_{\text{Up}} \times 1}$ through the uplink channel. It can be expressed as:

$$\mathbf{y}_{\text{BS},r} = \mathbf{H}_{\text{Up}} \mathbf{x}_{\text{UE},d} + \mathbf{H}_{\text{SI}} \mathbf{x}_{\text{BS},d} + \mathbf{n}_{\text{BS}} \quad (20),$$

where $\mathbf{n}_{\text{BS}} \sim \mathcal{CN}(0, \sigma_{\text{BS}}^2 \mathbf{I}_{N_r}) \in \mathbb{C}^{N_r \times 1}$ is the AWGN vector at the BS receiver.

Upon signal $\mathbf{y}_{\text{BS},r}$ reception at the BS, RX RF chains first process signals received at the RX antenna elements. Assuming that output signals of the RX RF chains are processed in the frequency domain by the received digital combiner $\mathbf{U}_{\text{BS}} \in \mathbb{C}^{K_{\text{Up}} \times N_{\text{RF},r}}$, vector $\mathbf{S}_{\text{BS},r} \in \mathbb{C}^{K_{\text{Up}} \times 1}$ is derived as:

$$\mathbf{S}_{\text{BS},r} = \mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{C}_r \mathbf{y}_{\text{BS},r} \quad (21),$$

where $\mathbf{C}_r = \frac{1}{3} \mathbf{E}_{N_{\text{RF},r}} \otimes [\mathbf{E}_{N_{r,d}}, \mathbf{E}_{N_{r,s}}, \mathbf{E}_{N_{r,s}}] \in \mathbb{C}^{N_{\text{RF},r} \times N_r}$ is the equivalent matrix from the receive antenna to the receive RF channel at the BS. We utilize $\mathbf{H}_{\text{Down},\text{equ}} = \mathbf{H}_{\text{Down}} \mathbf{C}_t \in \mathbb{C}^{K_{\text{Down}} \times N_{\text{RF},d}}$ and $\mathbf{H}_{\text{Up},\text{equ}} = \mathbf{C}_r \mathbf{H}_{\text{Up}} \in \mathbb{C}^{N_{\text{RF},r} \times K_{\text{Up}}}$ to represent the equivalent of downlink channel from the TX RF channel at the BS to nodes Down and the uplink channel from nodes Up to the RX RF channel at the BS, respectively. The equivalent SI channel from the TX RF channel to the RX RF channel at the BS is noted as $\mathbf{H}_{\text{SI},\text{equ}} = \mathbf{C}_r \mathbf{H}_{\text{SI}} \mathbf{C}_t \in \mathbb{C}^{N_{\text{RF},r} \times N_{\text{RF},d}}$. Eq. (19) is derived as:

$$\mathbf{y}_{\text{UE},r} = \mathbf{H}_{\text{Down},\text{equ}} \mathbf{F}^H \mathbf{V}_{\text{BS}} \mathbf{S}_{\text{BS},d} + \mathbf{n}_{\text{UE}} \quad (22).$$

Eq. (21) is obtained as:

$$\begin{aligned} \mathbf{S}_{\text{BS},r} &= \mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{H}_{\text{Up},\text{equ}} \mathbf{x}_{\text{UE},d} + \mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{H}_{\text{SI},\text{equ}} \mathbf{F}^H \mathbf{V}_{\text{BS}} \mathbf{S}_{\text{BS},d} + \\ &\mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{C}_r \mathbf{n}_{\text{BS}} \end{aligned} \quad (23).$$

3 SoftNull Algorithm for FD Massive MIMO System

To achieve SIC, the SoftNull algorithm based on the transmit digital precoding matrix is adopted in this paper. The algorithm has two primary design objectives for the transmit digital precoding matrix: ensuring the useful signal power and minimizing the total SI signal power. By splitting the transmit digital precoding matrix \mathbf{V}_{BS} into two sub-matrices, we have:

$$\mathbf{V}_{\text{BS}} = \mathbf{A} \mathbf{B} \quad (24).$$

The implementation block diagram is shown in Fig. 4. On the BS-side, K_{Down} original parallel transmit data streams $\mathbf{S}_{\text{BS},d}$ are first passed through the conventional digital precoding matrix \mathbf{B} to obtain f_D parallel data streams, and then through the SIC matrix \mathbf{A} to obtain $N_{\text{RF},d}$ parallel data streams, which are subsequently fed into the $N_{\text{RF},d}$ transmit RF chains.

The first-stage matrix \mathbf{B} is used for conventional digital precoding to suppress inter-user interference. At this point, the available degrees of freedom for conventional digital precoding are f_D . Matrix \mathbf{B} can be derived using conventional digital precoding algorithms, such as Zero Forcing (ZF), MMSE, and singular value decomposition. In this paper, ZF is adopted for the design. By computing the pseudo-inverse of the channel matrix, signals in all directions other than the user direction is forced to zero, thus avoiding inter-user interference. Under the system architecture of this paper, the expression of matrix \mathbf{B} is given as:

$$\mathbf{B} = \alpha_{\text{equ}} \mathbf{H}_{\text{equ}}^H (\mathbf{H}_{\text{equ}} \mathbf{H}_{\text{equ}}^H)^{-1} = \alpha_{\text{equ}} \mathbf{W}_{\text{ZF},\text{equ}} \quad (25),$$

where $\alpha_{\text{equ}} = \sqrt{\frac{K_1}{\text{tr}(\mathbf{W}_{\text{ZF},\text{equ}} \mathbf{W}_{\text{ZF},\text{equ}}^H)}}$ is the power control factor

and K_1 is a constant; $\mathbf{H}_{\text{equ}} = \mathbf{H}_{\text{Down},\text{equ}} \mathbf{A}$ is the equivalent channel. Therefore, when designing the first stage matrix \mathbf{B} , only the information of the equivalent channel \mathbf{H} is required, and it is not necessary to obtain the information of the downlink channel \mathbf{H} and matrix \mathbf{A} separately.

The second-stage matrix \mathbf{A} is used to minimize the total SI signal power while preserving the degrees of freedom f_D for the conventional digital precoding matrix \mathbf{B} . Therefore, the design objective of matrix \mathbf{A} is to minimize the total SI signal power while preserving the required degrees of freedom f_D for the conventional digital precoding. Assuming that the equivalent SI channel $\mathbf{H}_{\text{SI},\text{equ}}$ is known, the design problem is formulated as follows:

$$\begin{aligned} \mathbf{A} &= \underset{\mathbf{A}}{\text{argmin}} \left\| \mathbf{H}_{\text{SI},\text{equ}} \mathbf{A} \right\|_F^2 \\ \text{s.t. } \mathbf{A}^H \mathbf{A} &= \mathbf{E}_{f_D \times f_D} \end{aligned} \quad (26),$$

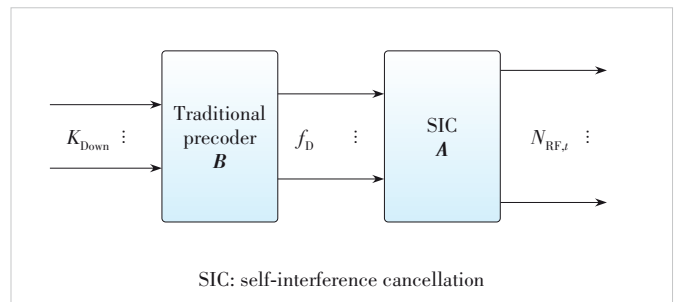


Figure 4. FD massive MIMO system with transmitting digital precoder

where $\|\mathbf{H}_{\text{SI, equ}}\|_F^2$ is the total SI signal power and $\|\cdot\|_F^2$ is the square of the Frobenius parametrization; the constraint $\mathbf{A}^H \mathbf{A} = \mathbf{E}_{f_D \times f_D}$ denotes that \mathbf{A} has F orthogonal columns and is used to ensure that the required degrees of freedom f_D are preserved for conventional digital precoding. The solution to Eq. (26) is:

$$\mathbf{A} = \begin{bmatrix} \mathbf{v}^{(N_{\text{RF},t} - f_D + 1)} & \mathbf{v}^{(N_{\text{RF},t} - f_D + 2)} & \dots & \mathbf{v}^{(N_{\text{RF},t})} \end{bmatrix} \quad (27),$$

where $\text{rank}(\mathbf{H}_{\text{SI, equ}}) = r$ is assumed. $\sigma_1, \sigma_2, \dots, \sigma_r$ are the r positive singular values of $\mathbf{H}_{\text{SI, equ}}$. The singular value decomposition of matrix $\mathbf{H}_{\text{SI, equ}}$ is $\mathbf{H}_{\text{SI, equ}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$, where $\mathbf{U} \in \mathbb{C}^{N_{\text{RF},t} \times N_{\text{RF},t}}$ and $\mathbf{V} \in \mathbb{C}^{N_{\text{RF},t} \times N_{\text{RF},t}}$ are both unitary matrices, $\mathbf{\Sigma} = \text{diag}(\delta_1, \delta_2, \dots, \delta_r, 0, 0, \dots, 0) \in \mathbb{C}^{N_{\text{RF},t} \times N_{\text{RF},t}}$ is a diagonal matrix, and δ_i is a complex number satisfying $|\delta_i| = \sigma_i$; $\mathbf{v}^{(i)}$ is the i -th column of the matrix \mathbf{V} .

It can be seen that SI suppression matrix \mathbf{A} is constructed by projecting onto the f_D right singular vectors corresponding to the f_D least singular values of the equivalent SI channel $\mathbf{H}_{\text{SI, equ}}$. In essence, this process involves identifying the f_D -dimensional subspace in the original transmission space $\mathbb{C}^{N_{\text{RF},t}}$ that minimizes the total SI received by the BS. Therefore, the greater the number of zero singular values of matrix $\mathbf{H}_{\text{SI, equ}}$ (i.e., the smaller the rank r of the matrix $\mathbf{H}_{\text{SI, equ}}$), the closer the f_D -dimensional subspace that minimizes the total SI received by the BS is to the zero space, and the better the SIC effect will be.

Next, the value of r is discussed for the proposed system architecture. Assume that $N_{\text{RF}} = N_{\text{RF},t} = N_{\text{RF},r}$, $N = N_t = N_r$, $N^1 = N_t^1 = N_r^1$, $N_{\text{RF}}^1 = N_{\text{RF},t}^1 = N_{\text{RF},r}^1$, $N_x = N_{t,x} = N_{r,x}$, and $N_y = N_{t,y} = N_{r,y}$. At this point, $\mathbf{C}_r = \mathbf{C}_t^T$, and $\text{rank}(\mathbf{C}_t) = N_{\text{RF}}$. We let $\mathbf{C} = \mathbf{C}_t$ and $r_{\text{SI}}^{11} = \text{rank}(\mathbf{H}_{\text{SI}}^{11})$. First, we discuss the case of channel matrix $\mathbf{H}_{\text{SI}}^{11}$ with a full rank, i.e., $r_{\text{SI}}^{11} = N^1$. When $\mathbf{K} = \beta \mathbf{J}$ (where $\mathbf{J} \in \mathbb{C}^{N^1 \times N^1}$ is an all-ones matrix), the polarization isolation between each transceiver array element in different polarization directions is given by $P_{\text{XPI, dB}} = -10 \log_{10}(\beta)$ with $0 < \beta \leq 1$. Eq. (18) is equivalent to:

$$\mathbf{H}_{\text{SI}} = \begin{bmatrix} \mathbf{H}_{\text{SI}}^{11} & \beta \mathbf{H}_{\text{SI}}^{11} \\ \beta \mathbf{H}_{\text{SI}}^{11} & \mathbf{H}_{\text{SI}}^{11} \end{bmatrix} \quad (28).$$

The values of r are discussed for each of the three cases $0 < \beta < 1$, $\beta \rightarrow 0$ and $\beta = 1$.

When $0 < \beta < 1$, with chunk matrix multiplication, the equivalent self-interfering channel matrix expression is obtained as:

$$\mathbf{H}_{\text{SI, equ}} = \begin{bmatrix} \mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11} & \beta \mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11} \\ \beta \mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11} & \mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11} \end{bmatrix} \quad (29),$$

where $\mathbf{C}_{11} \in \mathbb{C}^{N^1 \times N_{\text{RF}}^1}$, and $\text{rank}(\mathbf{C}_{11}) = N_{\text{RF}}^1$. It can be obtained that:

$$r = 2 \times \text{rank}(\mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11}) \leq N_{\text{RF}} \quad (30).$$

Similarly, when $\beta \rightarrow 0$,

$$r = 2 \times \text{rank}(\mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11}) \leq N_{\text{RF}} \quad (31).$$

When $\beta = 1$,

$$r = \text{rank}(\mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11}) \leq N_{\text{RF}}^1 \quad (32).$$

In summary,

$$r \leq \begin{cases} N_{\text{RF}}, & 0 < \beta < 1 \\ N_{\text{RF}}, & \beta \rightarrow 0 \\ N_{\text{RF}}^1, & \beta = 1 \end{cases} \quad (33).$$

For $0 < \beta < 1$, i.e., when the polarization isolation between the transmit and receive elements of different polarization directions is greater than 0 dB, the matrix $\mathbf{H}_{\text{SI, equ}}$ may not have any zero singular values. That is to say, the f_D -dimensional subspace constructed by the f_D right singular vectors that minimize the total SI of BS reception is not completely close to the zero space, and the SIC effect is not good. When $\beta \rightarrow 0$, i.e., the polarization isolation between the transmit and receive elements of different polarization directions approaches infinity, the number of $\mathbf{H}_{\text{SI, equ}}$ singular values of zero does not change significantly compared with $0 < \beta < 1$. In other words, the SIC performance does not improve significantly as polarization isolation increases. When $\beta = 1$, i.e., the polarization isolation between each transmit and receive array element with different polarization directions is 0 dB, it is evident from the structure of the transceiver URA that $N_{\text{RF}} > N_{\text{RF}}^1$. At this time, the number of $\mathbf{H}_{\text{SI, equ}}$ singular values taking zero is at least $N_{\text{RF}} - N_{\text{RF}}^1 > 0$, which means that the f_D -dimensional subspace constructed by the f_D right singular vectors is closer to the zero space, and the SI cancellation effect is obviously enhanced. Therefore, the polarization isolation can be reduced to 0 dB to improve the SIC capability, if the polarization isolation between each transceiver array element in different polarization directions is the same.

According to the actual measurement of the engineering prototype, the polarization isolation between each transmit and receive array element with different polarization directions at different locations is not the same, indicating that the elements in the matrix \mathbf{K} are not exactly equal. Therefore, the case of r at $\mathbf{K} \neq \beta \mathbf{J}$ is further discussed as follows. At this point, $\text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) \leq r_{\text{SI}}^{11} = N^1$. For matrix $\mathbf{K} = (k_{ij})_{N^1 \times N^1}$, we express it as:

$$\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_{N^1}] = [\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_{N^1}^T]^T \quad (34),$$

where $\mathbf{k}_j \in \mathbf{C}^{N^1 \times 1}, j = 1, 2, \dots, N^1$, and $\mathbf{g}_i \in \mathbf{C}^{1 \times N^1}, i = 1, 2, \dots, N^1$. We let $\mathbf{H}_{\text{SI}} = (\mathbf{h}_{ij})_{N \times N}$. At this point, the expression of the equivalent SI channel $\mathbf{H}_{\text{SI, equ}}$ is:

$$\mathbf{H}_{\text{SI, equ}} = \begin{bmatrix} \mathbf{H}_{\text{SI, equ}}^{11} & \mathbf{H}_{\text{SI, equ}}^{12} \\ \mathbf{H}_{\text{SI, equ}}^{12} & \mathbf{H}_{\text{SI, equ}}^{11} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11} & \mathbf{C}_{11}^T (\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) \mathbf{C}_{11} \\ \mathbf{C}_{11}^T (\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) \mathbf{C}_{11} & \mathbf{C}_{11}^T \mathbf{H}_{\text{SI}}^{11} \mathbf{C}_{11} \end{bmatrix} \quad (35),$$

where $\text{rank}(\mathbf{H}_{\text{SI, equ}}^{11}) \leq N_{\text{RF}}^1$, $\text{rank}(\mathbf{H}_{\text{SI, equ}}^{12}) \leq \begin{cases} \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) & \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) < N_{\text{RF}}^1 \\ N_{\text{RF}}^1 & \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) \geq N_{\text{RF}}^1 \end{cases}$. Denote the j -th column of $\mathbf{H}_{\text{SI, equ}}^{11}$ and $\mathbf{H}_{\text{SI, equ}}^{12}$ as $\mathbf{h}_{\text{SI, equ}, j}^{11}$ and $\mathbf{h}_{\text{SI, equ}, j}^{12}$ respectively, where the elements of $\mathbf{h}_{\text{SI, equ}, j}^{11}$ and $\mathbf{h}_{\text{SI, equ}, j}^{12}$ are $\sum_{k,l=0, N_y, 2N_y} h_{i+kj+l}$ and $\sum_{k,l=0, N_y, 2N_y} k_{i+kj+l} h_{i+kj+l}$ ($i = 1, 2, \dots, N_y, 3N_y + 1, 3N_y + 2, \dots, 4N_y, \dots, N^1; j = 1, 2, \dots, N_{\text{RF}}^1$).

To minimize the rank r of the matrix $\mathbf{H}_{\text{SI, equ}}$, it is necessary that the j -th column of $\mathbf{H}_{\text{SI, equ}}^{12}$ in Eq. (35) is equal to the j -th column of $\mathbf{H}_{\text{SI, equ}}^{11}$, i.e., the following condition must be satisfied.

$$\sum_{k,l=0, N_y, 2N_y} k_{i+kj+l} h_{i+kj+l} = \sum_{k,l=0, N_y, 2N_y} h_{i+kj+l} \quad (36),$$

which can be rewritten as:

$$k_{i+kj+l} = 1 \quad (37),$$

where $i = 1, 2, \dots, N_y, 3N_y + 1, 3N_y + 2, \dots, 4N_y, \dots, N^1$, $k, l = 0, N_y, 2N_y$. Eq. (37) is equivalent to:

$$\mathbf{k}_j = \mathbf{k}_{j+N_y} = \mathbf{k}_{j+2N_y} = [1, 1, \dots, 1]^T \quad (38).$$

In other words, the polarization isolation between each transmit and receive array element with different polarization directions at the corresponding position of Eq. (38) is close to 0 dB. According to the structure of the transmit-receive URA, the polarization isolation between the transmit array element connected to the j -th transmit RF channel and all the receive array elements is 0 dB under this condition. Similarly, r can be reduced when the i -th row of $\mathbf{H}_{\text{SI, equ}}^{12}$ in Eq. (35) is equal to that of $\mathbf{H}_{\text{SI, equ}}^{11}$, i.e.,

$$\mathbf{g}_i = \mathbf{g}_{i+N_y} = \mathbf{g}_{i+2N_y} = [1, 1, \dots, 1] \quad (39).$$

At this point, the polarization isolation between the receive

array connected to the i -th receive RF channel and all transmit arrays is 0 dB. Note that W is the set of column numbers of the column vector \mathbf{k}_j satisfying Eq. (38), P is the set of row numbers of the row vector \mathbf{g}_i satisfying Eq. (39), $w = |W|$ is the number of elements in the set, and $p = |P|$ is the number of elements in the set P . We have:

$$r = \text{rank}([\mathbf{H}_{\text{SI, equ}}^{11} \quad \mathbf{H}_{\text{SI, equ}}^{12}]) - w - p \quad (40),$$

where

$$\text{rank}([\mathbf{H}_{\text{SI, equ}}^{11} \quad \mathbf{H}_{\text{SI, equ}}^{12}]) \leq \begin{cases} N_{\text{RF}}^1 + \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) & \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) < N_{\text{RF}}^1 \\ N_{\text{RF}}^1 & \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) \geq N_{\text{RF}}^1 \end{cases} \quad (41).$$

At this point, r is related to the rank of matrix $\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}$ and the number of unit row (or column) vectors in matrix \mathbf{K} . Therefore, in practical engineering implementations, \mathbf{K} can be designed so that either the rank of $\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}$ is less than N_{RF}^1 , or the polarization isolation between the transmit (or receive) array elements connected to as many transmit (or receive) RF channels as possible and all the corresponding receive (or transmit) array elements is 0 dB, in order to achieve better SIC performance. It should be noted that when using the \mathbf{K} method, if the rank of $\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}$ is only reduced but remains greater than N_{RF}^1 , it will not yield a significant improvement in SIC capability.

In practice, the antenna array placement of massive MIMO does not always result in completely uncorrelated and independent channels. In such cases, channel matrix $\mathbf{H}_{\text{SI}}^{11}$ is non-full rank, i.e., $r_{\text{SI}}^{11} < N^1$. The corresponding value of rank r when $r_{\text{SI}}^{11} < N^1$ is analyzed below. When $\mathbf{K} = \beta \mathbf{J}$, the same three cases are discussed in terms of $0 < \beta < 1$, $\beta \rightarrow 0$ and $\beta = 1$.

When $r_{\text{SI}}^{11} < N_{\text{RF}}^1$,

$$\text{rank}(\mathbf{H}_{\text{SI, equ}}) \leq \begin{cases} 2r_{\text{SI}}^{11}, & 0 < \beta < 1 \\ 2r_{\text{SI}}^{11}, & \beta \rightarrow 0 \\ r_{\text{SI}}^{11}, & \beta = 1 \end{cases} \quad (42).$$

When $r_{\text{SI}}^{11} \geq N_{\text{RF}}^1$,

$$\text{rank}(\mathbf{H}_{\text{SI, equ}}) \leq \begin{cases} N_{\text{RF}}, & 0 < \beta < 1 \\ N_{\text{RF}}, & \beta \rightarrow 0 \\ N_{\text{RF}}^1, & \beta = 1 \end{cases} \quad (43).$$

Therefore, when $N_{\text{RF}}^1 \leq r_{\text{SI}}^{11} < N^1$, r does not change much compared with the case where $\mathbf{H}_{\text{SI}}^{11}$ is full rank. When $r_{\text{SI}}^{11} < N_{\text{RF}}^1$, r decreases, and at this time, the number of zero singular values of $\mathbf{H}_{\text{SI, equ}}$ is at least $N_{\text{RF}} - 2r_{\text{SI}}^{11}$ ($0 < N_{\text{RF}} - 2r_{\text{SI}}^{11} < N_{\text{RF}}$). Thus, the f_D -dimensional subspace constructed by f_D

right singular vectors is closer to the zero space, and the SI canceling performance is significantly enhanced. Consequently, in practical engineering implementation, to enhance SIC performance by changing the rank r_{SI}^{11} of $\mathbf{H}_{\text{SI}}^{11}$, r_{SI}^{11} should be reduced to at least less than the number of transmit or receive RF channels connected to the antenna array elements with the same polarization direction.

When $\mathbf{K} \neq \beta \mathbf{J}$, it can be obtained similarly that

$$r = \text{rank}([\mathbf{H}_{\text{SI, equ}}^{11} \quad \mathbf{H}_{\text{SI, equ}}^{12}]) - w - p \quad (44),$$

where,

$$\text{rank}([\mathbf{H}_{\text{SI, equ}}^{11} \quad \mathbf{H}_{\text{SI, equ}}^{12}]) \leq \begin{cases} r_{\text{SI}}^{11} + \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}), & r_{\text{SI}}^{11} < N_{\text{RF}}^1 \\ N_{\text{RF}}^1 + \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}), & \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) < N_{\text{RF}}^1 \leq r_{\text{SI}}^{11} \\ N_{\text{RF}}^1, & \text{rank}(\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}) \geq N_{\text{RF}}^1 \end{cases} \quad (45).$$

Thus, when $\mathbf{K} \neq \beta \mathbf{J}$, the rank r is related not only to the rank of $\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}$ and the number of unit row (or column) vectors in \mathbf{K} , but also to the rank r_{SI}^{11} of $\mathbf{H}_{\text{SI}}^{11}$. Accordingly, in practical engineering implementations, to enhance the SI cancellation capability, in addition to designing \mathbf{K} to minimize the rank of $\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}$ (i.e., ensuring it is less than N_{RF}^1), or maximizing the number of transmit (receive) array elements connected to transmit (receive) RF chains for which the polarization isolation with all corresponding receive (transmit) array elements is 0 dB, r_{SI}^{11} can also be reduced to less than the number of transmit or receive RF channels connected to the antenna array elements with the same polarization direction.

In this paper, we also consider whether the SIC performance can be further enhanced by changing the array placement. The transmitting and receiving URAs placed vertically along the x -axis in Fig. 3 are reconfigured to a diagonal placement along the same axis. The equivalent SI channel matrix $\mathbf{H}_{\text{SI, equ}}$ obtained at this point is the same as that in Eq. (38). This is because using different array placements only changes the values of the elements of the sub-matrix $\mathbf{H}_{\text{SI}}^{11}$, but the transmit or receive array elements of both polarization directions are still in the same position. Thus the form of $\mathbf{H}_{\text{SI, equ}}$ will not change. Therefore, the rank r_{SI}^{11} of the channel matrix $\mathbf{H}_{\text{SI}}^{11}$ can be minimized by changing the array placement, so that the SIC performance can be improved.

According to the above analysis, the SI cancellation effect can be enhanced from three perspectives: 1) by designing matrix \mathbf{K} so that the rank of $\mathbf{K} \odot \mathbf{H}_{\text{SI}}^{11}$ is less than the number of transmit or receive RF channels connected to the antennas with the same polarization direction; 2) by making the polarization isolation between the transmit (receive) array elements connected to the same transmit (receive) RF channel and all receive (transmit) array elements 0 dB; 3) by reducing the rank r_{SI}^{11} of $\mathbf{H}_{\text{SI}}^{11}$ to be smaller than the number of transmit or

ceive RF channels connected to antennas with the same polarization direction, such as changing the array placement and widening the transceiver URA spacing. Although the channel is modeled as narrowband in the above analysis, beamforming is also applicable to the wideband channel model. This is because beamforming targets frequency-domain data and is implemented at the subcarrier level.

4 Illustrative Results

In this section, the SoftNull algorithm performance for the proposed system is evaluated through simulations from the sum rate and SIC. The downlink achievable rate can be given by:

$$R_{\text{DL}} = \log_2 |\mathbf{I}_{K_{\text{Down}}} + \mathbf{R}_{\text{UE}}^{-1} \mathbf{W}_{\text{DL}}| \quad (46),$$

where $\mathbf{R}_{\text{UE}} = \mathbf{E}\{\mathbf{n}_{\text{UE}} \mathbf{n}_{\text{UE}}^H\}$, and $\mathbf{E}\{\cdot\}$ denotes the statistical expectation. \mathbf{W}_{DL} represents the received covariance matrices of the downlink channel, which is defined as:

$$\mathbf{W}_{\text{DL}} = \mathbf{H}_{\text{Down, equ}} \mathbf{G}_{\text{BS}} \mathbf{H}_{\text{Down, equ}}^H \quad (47),$$

where $\mathbf{G}_{\text{BS}} = \mathbf{F}^H \mathbf{V}_{\text{BS}} \mathbf{E}\{\mathbf{S}_{\text{BS}, J} \mathbf{S}_{\text{BS}, J}^H\} \mathbf{V}_{\text{BS}}^H \mathbf{F}$. Similarly, the uplink (UL) rate is expressed as:

$$R_{\text{UL}} = \log_2 |\mathbf{I}_{K_{\text{Up}}} + (\mathbf{R}_{\text{BS}} + \mathbf{Q}_{\text{BS}})^{-1} \mathbf{W}_{\text{UL}}| \quad (48),$$

where $\mathbf{R}_{\text{BS}} = \mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{E}\{\mathbf{n}_{\text{BS}} \mathbf{n}_{\text{BS}}^H\} \mathbf{F}^H \mathbf{U}_{\text{BS}}^H$ denotes the power of received noise, \mathbf{Q}_{BS} and \mathbf{W}_{UL} are the downlink receive covariance matrices for the SI and intended signal on the BS side, respectively, defined as:

$$\mathbf{Q}_{\text{BS}} = \mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{H}_{\text{SI, equ}} \mathbf{G}_{\text{BS}} \mathbf{H}_{\text{SI, equ}}^H \mathbf{F}^H \mathbf{U}_{\text{BS}}^H \quad (49),$$

$$\mathbf{W}_{\text{UL}} = \mathbf{U}_{\text{BS}} \mathbf{F} \mathbf{H}_{\text{Up, equ}} \mathbf{G}_{\text{UE}} \mathbf{H}_{\text{Up, equ}}^H \mathbf{F}^H \mathbf{U}_{\text{BS}}^H \quad (50),$$

where $\mathbf{G}_{\text{UE}} = \mathbf{E}\{\mathbf{x}_{\text{UE}, J} \mathbf{x}_{\text{UE}, J}^H\}$. Accordingly, the achievable rate can be written as:

$$R_{\text{sum}} = R_{\text{DL}} + R_{\text{UL}} = \log_2 |\mathbf{I}_{K_{\text{Down}}} + \mathbf{R}_{\text{UE}}^{-1} \mathbf{W}_{\text{DL}}| + \log_2 |\mathbf{I}_{K_{\text{Up}}} + (\mathbf{R}_{\text{BS}} + \mathbf{Q}_{\text{BS}})^{-1} \mathbf{W}_{\text{UL}}| \quad (51).$$

The parameters are summarized in Table 1, where λ is the wavelength. It is assumed that the NLOS path between each transmit and receive antenna pair is reflected by the point o_{wall} .

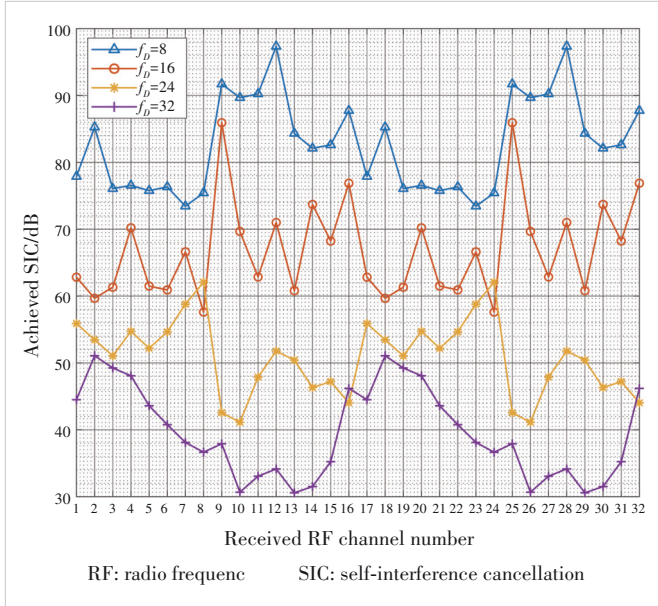
In Fig. 5, we investigate the SIC of each receiving RF channel versus the degrees of freedom f_D reserved for traditional digital precoding under the SoftNull scheme. We assume that the antenna isolation is 0 dB and the URA spacing is 0.2 m, which is 1.7 times the wavelength.

As can be observed, a decrease in f_D corresponds to an in-

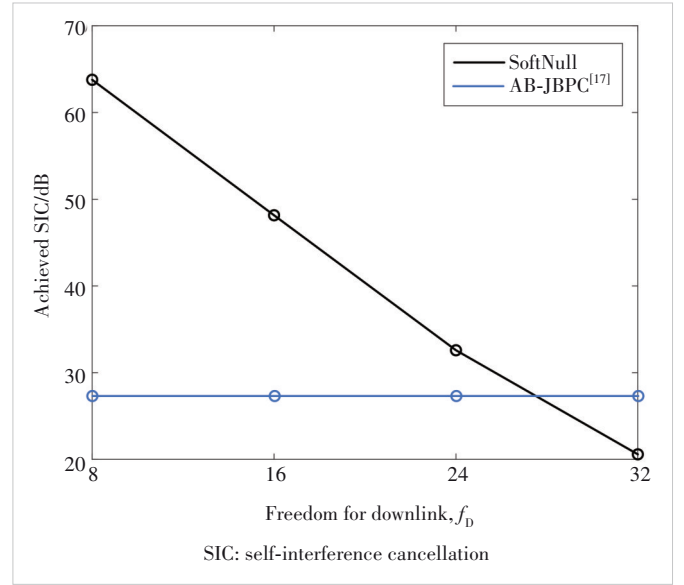
Table 1. Simulation parameters of full-duplex massive MIMO system

Parameter	Value
Bandwidth	100 MHz
Center frequency	2.595 GHz
Transmit power of base station	30 dBm
Transmit power of user	30 dBm
Number of base station antennas	$N_t = N_r = 6 \times 8 \times 2 = 96$
Number of base station RF channels	$N_{RF,d} = N_{RF,r} = 32$
Number of uplink users	$K_{Up} = 2 \times 4 = 8$
Number of downlink users	$K_{Down} = 2 \times 4 = 8$
Antenna spacing along x-axis	$d_{t,x} = d_{r,x} = 0.67\lambda$
Antenna spacing along y-axis	$d_{t,y} = d_{r,y} = 0.5\lambda$
URAs spacing	$d_i = 0.5\lambda$
User spacing along x-axis	$d_{Up,x} = d_{Down,x} = 2.6795 \text{ m}$
User spacing along y-axis	$d_{Up,y} = d_{Down,y} = 2.6795 \text{ m}$
Cross-polarization isolation	$P_{XPL,dB} = 25 \text{ dB}$
Path loss factor	$\eta = 2.92$
Number of uplink channel multipaths	$C_{Up} = 1, L_{Up,e} = 1, L_{Up} = 1$
Number of downlink channel multipaths	$C_{Down} = 1, L_{Down,e} = 1, L_{Down} = 1$
Number of SI channel multipaths	$C_{SI} = 1, L_{SI,e} = 1, L_{SI} = 1$

URA: Uniform rectangular array

**Figure 5. Impact of f_D on SIC in received RF channels**

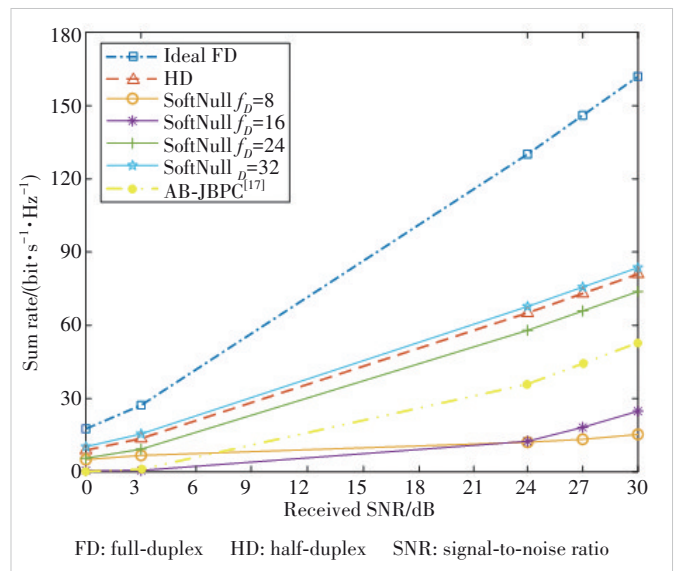
crease in the SIC performance for each receive RF chain. This is because a reduction in the degrees of freedom assigned to traditional digital precoding invariably increases the degrees of freedom available for SIC, resulting in enhanced SIC performance. Additionally, the SIC performance of different receive RF channels fluctuates to varying degrees, since the design goal of the SI suppression matrix is to minimize the total SI signal power rather than the SI signal power of each receiving RF

**Figure 6. Impact of f_D on total SIC**

channel. In Fig. 6, we further investigate f_D versus the total SIC.

However, when we utilize the SoftNull scheme in the proposed system architecture, the total SI channel power can be suppressed by up to 64 dB as shown in Fig. 6, including a path loss of 21 dB. The total SIC performance exhibits an increasing trend with a decrease in the degrees of freedom used for downlinks, similar to the SIC for each receive RF channel. Compared with AB-JHPC^[17] with only transmit beamforming, the proposed SoftNull scheme achieves higher SIC performance when the reserved degrees of freedom are less than 24.

Accordingly, SoftNull can achieve different SI suppression effects by changing the value of f_D within the proposed system architecture. However, the degrees of freedom available for

**Figure 7. Impact of f_D on sum rate**

downlink traditional beamforming also change. We further investigate the impact of f_D adjustment on the sum rate. Fig. 7 plots the relationship between f_D and the sum rate under Soft-Null, along with a performance comparison of HD, ideal FD, and AB-JHPC with only transmit beamforming.

It can be observed that the achievable sum rate increases with the increase of f_D . The maximum sum rate is improved by 16% compared with that of HD, when all available degrees of freedom are used for traditional digital precoding. However, the achieved SIC is only 21 dB from path loss, showing that the SI power on the UL reception is higher than the received desired signal power. It not only causes RF saturation of the receiver but also induces harm to the receiver hardware. Thus,

we should choose a reasonable value of f_D to balance the trade-off between the achieved SIC and the sum rate. In this way, a higher sum rate can be achieved without saturating the receive RF channels on the BS side.

In addition, the achievable sum rate is much smaller than that of the ideal FD. This is because reducing f_D can enhance SIC performance to improve the signal-to-interference plus noise ratio (SINR) and achieve a higher uplink rate, while it has a negative impact on the downlink rate. The uplink rate gain achieved by enhancing SIC performance is unable to compensate for the downlink rate loss. Thus, even when a certain SIC is achieved, the sum rate is low.

To improve the performance of the proposed full-duplex mas-

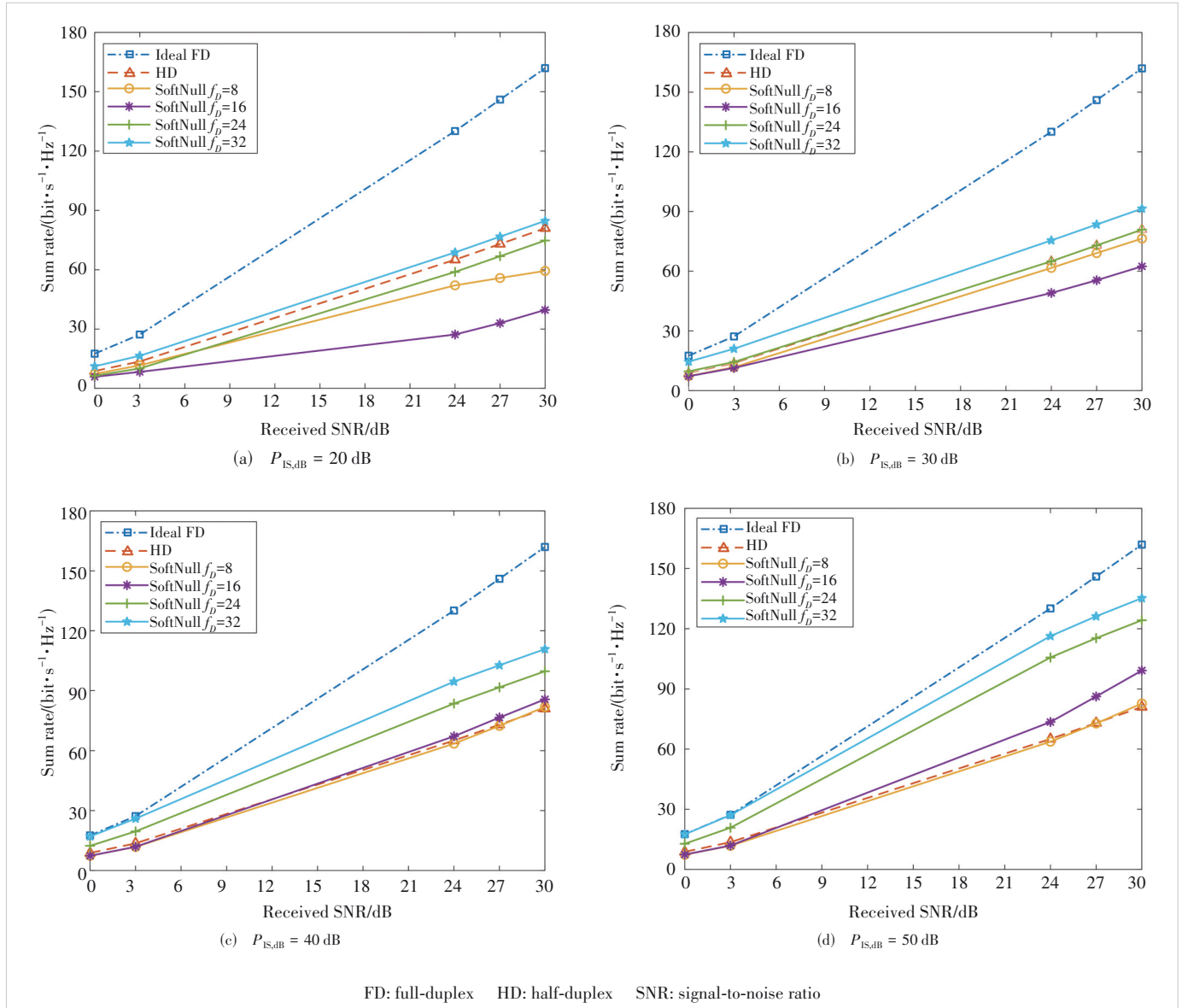


Figure 8. Impact of antenna isolation on sum rate

sive MIMO system with large antenna arrays, antenna isolation is used to further suppress the received SI, thereby enabling a higher sum rate while enhancing the total SIC performance. Fig. 8 depicts the sum rate with different antenna isolation.

Increasing antenna isolation further reduces the received SI signal, so that the uplink rate is improved. When more antenna isolation is utilized, the sum rate increases under different f_D . The received SI power is much smaller than the received intended signal power. And the achieved sum rate is higher than that of HD and approaches the ideal FD, when the antenna isolation is increased to 50 dB and $f_D = 32$. The results with the sum rate are increased by 26.7% compared with those using only the SoftNull algorithm, and 67% to 100% higher than the HD.

As an increase in antenna isolation causes a corresponding decrease in the received SI power, the received UL SINR is infinitely close to the received SNR. Thus, the sum rate under SoftNull cannot increase linearly with the increase of antenna isolation. The lower the received SNR, the smaller the total SIC required to achieve the same rate situation.

The SIC performance and the sum rate can also be further enhanced by increasing the spacing between the transmit and receive URAs. Fig. 9 depicts the sum rate with different antenna isolation when d_1 is increased to 5λ .

As can be observed, an increase in the distance between transmit and receive URAs under the same antenna isolation corresponds to an increase in the sum rate compared with that in Figs. 8c and 8d. When the antenna isolation is 50 dB, the sum rate can be increased by up to 54%, which is 90% to 100% higher than that of HD. The f_D -dimensional subspace found under the same available degrees of freedom and antenna isolation is closer to the null space by increasing the distance of the transceiver URAs, which reduces the rank r_{SI}^{11} . Thus, the received UL SINR and the sum rate performance are improved. It is necessary to reasonably select f_D , antenna isolation, and d_1 with the requirements when using SoftNull to achieve SI elimination under the proposed system architecture, so as to improve the sum rate while achieving the required SIC.

5 Conclusions

In this paper, a structure for the FD massive MIMO communication system with multi-stream transmission, where one transmit or receive RF channel is connected with three transmitting or receiving antennas in the same polarization direction, is proposed. This structure greatly reduces power dissipation and cost. To avoid LNA saturation caused by SI signals, the SoftNull algorithm based on transmit digital precoding is proposed. In addition, the method of enhancing the SIC under the algorithm is analyzed. We consider the method of antenna isolation and increasing the distance between receive and transmit URAs to improve system performance. Simulation results show that the SoftNull algorithm can achieve SI suppression in the ar-

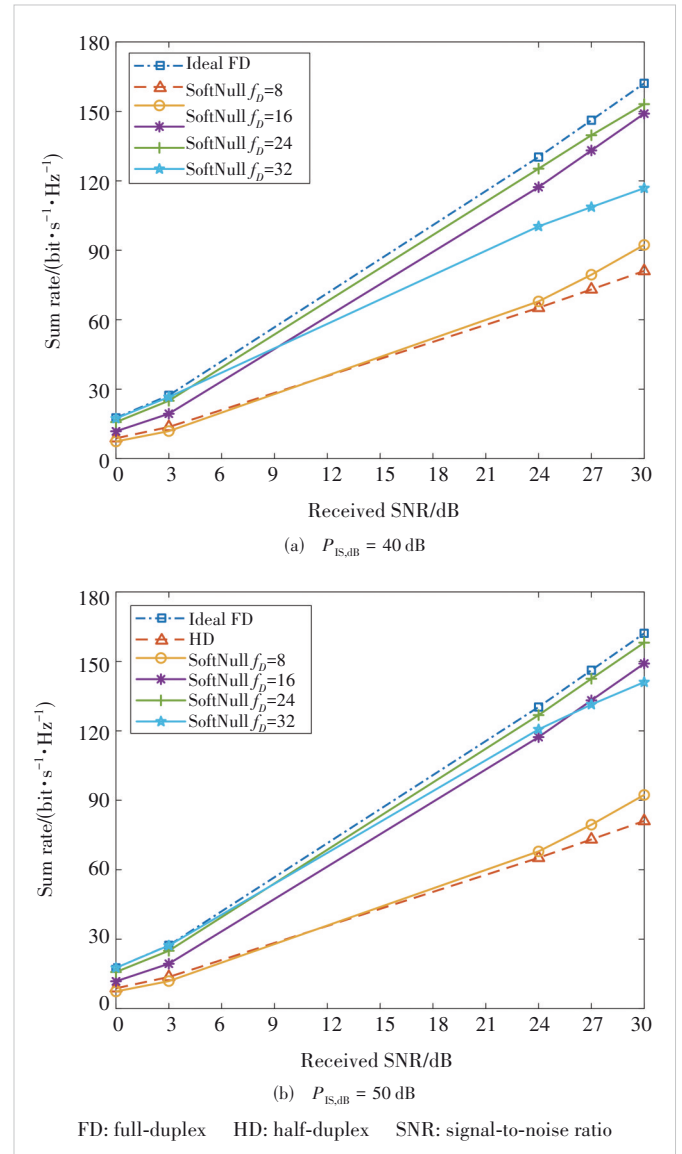


Figure 9. Impact of antenna isolation on sum rate with transceiver uniform rectangular array spacing of 5λ

chitecture. Moreover, the method of joint antenna isolation and increasing the distance between the transmit-receive URAs can further enhance system performance.

References

- [1] BOCCARDI F, HEATH R W, LOZANO A, et al. Five disruptive technology directions for 5G [J]. IEEE communications magazine, 2014, 52(2): 74 – 80. DOI: 10.1109/MCOM.2014.6736746
- [2] AGI WAL M, ROY A, SAXENA N. Next generation 5G wireless networks: a comprehensive survey [J]. IEEE communications surveys & tutorials, 2016, 18(3): 1617 – 1655. DOI: 10.1109/COMST.2016.2532458
- [3] HU Y F, WANG D M, LIANG C L, et al. Massive random access in cell-free massive MIMO system [J]. ZTE technology journal, 2024, 30(1): 26 – 32. DOI: 10.12142/ZTETJ.202401006

- [4] HAN Y, ZHANG J Y, JIN S. U6G extra-large scale MIMO technology [J]. ZTE technology journal, 2024, 30(3): 67 – 71. DOI: 10.12142/ZTETJ.202403011
- [5] ZHANG Z S, CHAI X M, LONG K P, et al. Full-duplex techniques for 5G networks: SI cancellation, protocol design, and relay selection [J]. IEEE communications magazine, 2015, 53(5): 128 – 137. DOI: 10.1109/MCOM.2015.7105651
- [6] AGHABABAEETAFRESHI M, KORPI D, KOSKELA M, et al. Software defined radio implementation of a digital SI cancellation method for in-band full-duplex radio using mobile processors [J]. Journal of signal processing systems, 2018, 90(10): 1297 – 1309. DOI: 10.1007/s11265-017-1312-0
- [7] NGO H Q, LARSSON E G, MARZETTA T L. Energy and spectral efficiency of very large multiuser MIMO systems [J]. IEEE transactions on communications, 2013, 61(4): 1436 – 1449. DOI: 10.1109/TCOMM.2013.020413.110848
- [8] HUANG Y M, HE S W, WANG J H, et al. Spectral and energy efficiency tradeoff for massive MIMO [J]. IEEE transactions on vehicular technology, 2018, 67(8): 6991 – 7002. DOI: 10.1109/TVT.2018.2824311
- [9] TRAN T X, TEH K C. Spectral and energy efficiency analysis for SLNR precoding in massive MIMO systems with imperfect CSI [J]. IEEE transactions on wireless communications, 2018, 17(6): 4017 – 4027. DOI: 10.1109/TWC.2018.2819184
- [10] PIRZADEH H, SWINDLEHURST A L. Spectral efficiency of mixed-ADC massive MIMO [J]. IEEE transactions on signal processing, 2018, 66(13): 3599 – 3613. DOI: 10.1109/TSP.2018.2833807
- [11] ZHANG X, ZHONG L, SABHARWAL A. Directional training for FDD massive MIMO [J]. IEEE transactions on wireless communications, 2018, 17(8): 5183 – 5197. DOI: 10.1109/TWC.2018.2838600
- [12] YIN B, WU M, STUDER C, et al. Full-duplex in large-scale wireless systems [C]//Asilomar Conference on Signals, Systems and Computers. IEEE, 2013: 1623 – 1627. DOI: 10.1109/ACSSC.2013.6810573
- [13] NGO H Q, SURAWEEA H A, MATTHAIU M, et al. Multipair full-duplex relaying with massive arrays and linear processing [J]. IEEE journal on selected areas in communications, 2014, 32(9): 1721 – 1737. DOI: 10.1109/JSAC.2014.2330091
- [14] GOLDSMITH A. Wireless Communications [M]. Cambridge, UK: Cambridge University Press, 2005
- [15] AHMED E, ELTAWIL A M. All-digital SI cancellation technique for full-duplex systems [J]. IEEE transactions on wireless communications, 2015, 14(7): 3519 – 3532. DOI: 10.1109/TWC.2015.2407876
- [16] LE A T, TRAN L C, HUANG X J, et al. Beam-based analog SI cancellation in full-duplex MIMO systems [J]. IEEE transactions on wireless communications, 2020, 19(4): 2460 – 2471. DOI: 10.1109/TWC.2020.2965441
- [17] ALEXANDROPOULOS G C, DUARTE M. Joint design of multi-tap analog cancellation and digital beamforming for reduced complexity full duplex MIMO systems [C]//Proceedings of IEEE International Conference on Communications (ICC). IEEE, 2017: 1 – 7. DOI: 10.1109/ICC.2017.7997175
- [18] IIMORI H, THADEU FREITAS DE ABREU G. Two-way full-duplex MIMO with hybrid TX-RX MSE minimization and interference cancellation [C]//IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2018: 1 – 5. DOI: 10.1109/SPAWC.2018.8445776
- [19] KOC A, LE-NGOC T. Full-duplex mmWave massive MIMO systems: a joint hybrid precoding/combining and SI cancellation design [J]. IEEE open journal of the communications society, 2021, 2: 754 – 774. DOI: 10.1109/OJCOMS.2021.3069672
- [20] SATYANARAYANA K, EL-HAJJAR M, KUO P H, et al. Hybrid beamforming design for full-duplex millimeter wave communication [J]. IEEE transactions on vehicular technology, 2019, 68(2): 1394 – 1404. DOI: 10.1109/TVT.2018.2884049
- [21] EVERETT E, SHEPARD C, ZHONG L, et al. SoftNull: many-antenna full-duplex wireless via digital beamforming [J]. IEEE transactions on wireless communications, 2016, 15(12): 8077 – 8092. DOI: 10.1109/TWC.2016.2612625

Biographies

ZHANG Boyu received his bachelor's degree in electronic information engineering from University of Electronic Science and Technology of China (UESTC) in 2022. He is currently working towards a master's degree in communication and information engineering at UESTC. His research interests include wireless communication and full duplex for massive MIMO systems.

ZHANG Ling received her master's degree from University of Electronic Science and Technology of China (UESTC) in 2023. She is currently a chip and device design engineer. Her research interests include wireless communications, massive MIMO systems, chip development.

LI Zijiang received her master's degree in electronic information engineering from University of Electronic Science and Technology of China (UESTC) in 2023. She is currently working towards a PhD degree in communication and information engineering at UESTC. Her research interests include SI suppression and integrated communication and sensing for massive MIMO systems.

SHEN Ying (shenyings@uestc.edu.cn) received his PhD degree in communication and information systems from University of Electronic Science and Technology of China (UESTC) in 2009. He is currently working at the National Key Laboratory of Wireless Communications, UESTC. His research interests include full-duplex, MIMO, and broadcasting systems.



A Root Cause Analysis Framework for Microservice Systems with Multimodal Data

LI Yingke¹, HAN Jing², SUN Yongqian¹, SHI Binpeng¹,
GONG Zican²

(1. Nankai University, Tianjin 300071, China;

2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202504012

<https://kns.cnki.net/kcms/detail/34.1294.TN.20251110.1139.002.html>,
published online November 10, 2025

Manuscript received: 2024-01-26

Abstract: In recent years, microservice architecture has gained increasing popularity. However, due to the complex and dynamically changing nature of microservice systems, failure detection has become more challenging. Traditional root cause analysis methods mostly rely on a single modality of data, which is insufficient to cover all failure information. Existing multimodal methods require collecting high-quality labeled samples and often face challenges in classifying unknown failure categories. To address these challenges, this paper proposes a root cause analysis framework based on a masked graph autoencoder (GAE). The main process involves feature extraction, feature dimensionality reduction based on GAE, and online clustering combined with expert input. The method is experimentally evaluated on two public datasets and compared with two baseline methods, demonstrating significant advantages even with 16% labeled samples.

Keywords: root cause analysis; multimodal data; self-supervised learning; online clustering

Citation (Format 1): LI Y K, HAN J, SUN Y Q, et al. A root cause analysis framework for microservice systems with multimodal data [J]. ZTE Communications, 2025, 23(4): 110 – 119. DOI: 10.12142/ZTECOM.202504012

Citation (Format 2): Y. K. Li, J. Han, Y. Q. Sun, et al., “A root cause analysis framework for microservice systems with multimodal data,” *ZTE Communications*, vol. 23, no. 4, pp. 110 – 119, Dec. 2025. doi: 10.12142/ZTECOM.202504012.

1 Introduction

The cloud-native system deploys various types of system instances, including network devices (e.g., routers, switches, and firewalls), servers, containers, microservices, middleware, etc. These system instances generate diverse modalities of data, such as traces, logs, and metrics, enabling comprehensive system monitoring from different perspectives. In the event of a failure in one or more system instances of the cloud-native system, the failure gradually propagates to parts or even the entire cloud platform, leading to performance degradation or interruption and posing significant risks to service stability^[1-3]. In order to promptly conduct root cause analysis, operations personnel continuously gather operational data like traces, logs, and metrics from each system instance to solve failure events.

Traditional anomaly detection methods based on single-modal data, such as traces^[4-5], logs^[6-9], and metrics^[10-12], have been widely studied. Nevertheless, recent research acknowledges that a more comprehensive understanding can be achieved by integrating the information from all three modalities^[13] (see Section 2.1 for details). This is attributed to their

ability to offer a holistic perspective on the overall system status^[3]. A failure in one instance of a cloud-native system can propagate to other instances or even the entire system, potentially leading to user dissatisfaction and revenue losses for service providers. In recent years, some methods^[13-17] have also been proposed to integrate multimodal data for more effective root cause analysis results.

However, the practical application of multimodal intelligent diagnosis still faces the following challenges:

1) Multimodal data often suffers from high-dimensional raw data, leading to low-quality features. The high dimensionality of raw data results in sparse data features and an excessive number of ineffective features, thereby impairing feature quality.

2) Most existing root cause analysis methods employ supervised learning, requiring a substantial number of labeled samples for training. Since these models are trained on a limited set of samples from known failure categories, they will achieve low F1 scores when classifying unknown failure categories in practice.

To address the limitations of existing methods, this paper proposes a framework based on a masked Graph Autoencoder (GAE).

The main contributions of this paper are summarized as follows:

- To tackle the first challenge, we use a masked GAE for

This work was supported by ZTE Industry-University-Institute Cooperation Funds under Grant No. HC-CN-20221123003.

feature dimension reduction. GAE effectively mitigates the impact of noise and compresses high-dimensional data into a lower-dimensional space that better captures the essence of the data.

- To address the second challenge, we have devised an online streaming clustering and classification framework. Specifically, we employ a cosine similarity metric to diagnose failures based on feature vectors, resulting in several clusters. These clusters are then presented to operations personnel, who investigate and assign corresponding normal or failure types to each cluster. This enables accurate identification of failure categories, providing targeted troubleshooting guidance for operations personnel, enhancing operational efficiency, and simultaneously reducing the workload of expert feedback annotations.

We conduct experiments on datasets collected from two benchmark microservice systems. Our method is compared with two industry-adopted approaches, demonstrating its effectiveness and efficiency.

2 Background

2.1 Multimodal Data Description

Compared with data from a single modality, the combination of multimodal data gains more valuable insights and delivers a more comprehensive view of the overall system status^[13]. In this section, we provide a detailed introduction to the definition of multimodal data. As shown in Fig. 1, in cloud-native systems, various types of instances such as physical machines, virtual machines, containers, microservices, and middleware coexist, generating a vast number of metrics, logs, and trace data.

Metric data is a type of time series data with specific significance obtained through timed sampling, in the format of (*timestamp*, *value*). Service-level metric data is usually a minute-level statistical metric at the internet business level (e.g., the number of online users of the service, the average service response time for users), most of which are seasonally smoothed metric data used to reflect the quality and scale of services. Machine metric data is usually a second-level statistical metric that reflects the health status

of the machine (e.g. CPU utilization and memory utilization), and most of it is more complex fine-grained metric data. The metric data in this paper refers to the second type of data.

Logs are a type of semi-structured text, produced by running programs and containing information such as timestamps, verbosity levels, and raw messages. Generated by developers using logging modules, they record the richest source data. Therefore, operators can discover the causes of issues by examining the logs.

For microservice systems, traces are a critical type of monitoring data. Unlike logs or metrics that focus on individual services, traces concentrate on interactions between services. Since distributed tracing^[4-5] was proposed, it has become an important component of the microservice architecture. For each user request, several microservices are called, and a trace is formed to record the tree-like call dependencies between microservices and the time consumption of each call node. Trace-based anomaly detection can detect anomalies in system execution based on the structure of traces (e.g., missing service calls) and metrics (e.g., latency). Due to its ability to better reflect the correlation between response times and call paths, this method has been widely used.

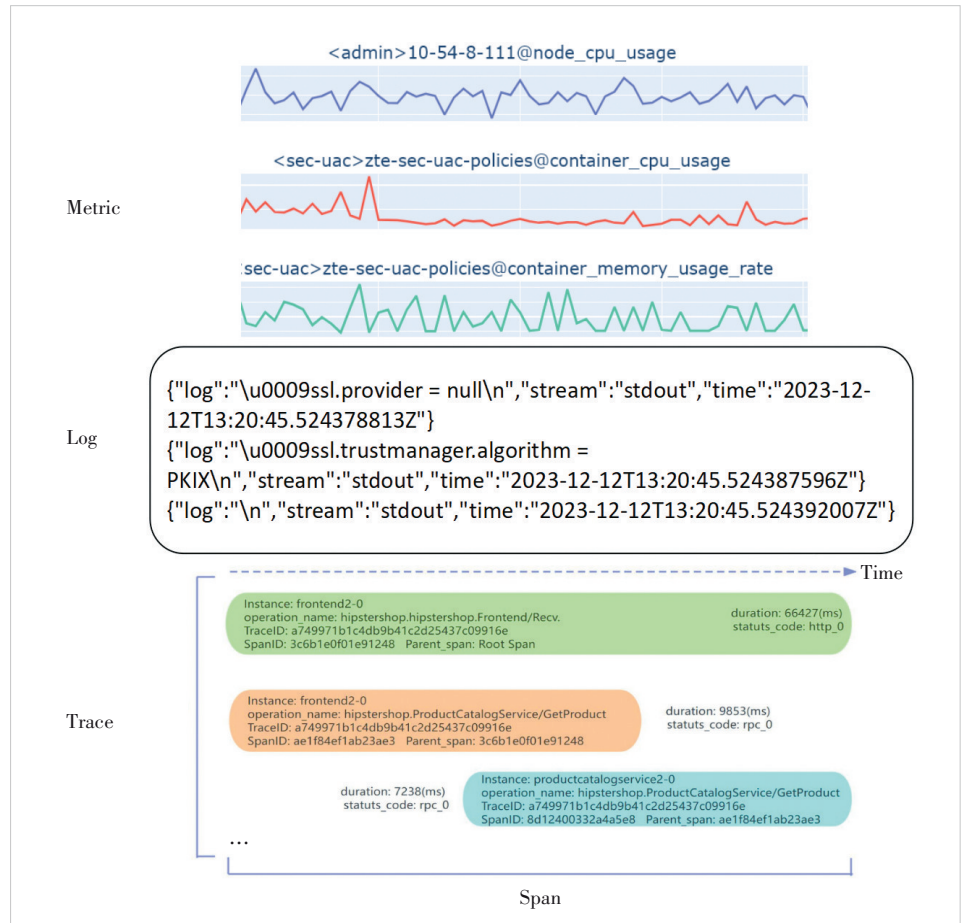


Figure 1. Multimodal data examples

2.2 Challenges and Analysis

1) Challenge 1: extracting high-quality features from multimodal data. Multimodal data, due to its diverse sources, often presents challenges such as high dimensionality and significant noise. This paper utilizes a streaming peaks-over-threshold (SPOT) algorithm^[10] to calculate thresholds for each type of time series data and extracts two features: the degree of deviation and the duration of anomalies in the time series data, both calculated based on these thresholds. If the original features are directly used as input, there is a challenge of excessive data noise and high dimensionality, significantly diminishing the accuracy of the root cause analysis model. To address this challenge, the paper employs a GAE for feature dimensionality reduction. GAE is chosen because it can mitigate the impact of noise, compressing high-dimensional data into a lower-dimensional space that better reflects the essence of the data.

2) Challenge 2: achieving more accurate classification results. Most existing methods rely on supervised learning, requiring a large number of annotated samples for training. Addressing this challenge is particularly crucial for establishing a robust and accurate failure classifier. Recent methods, such as Eadro and DiagFusion, struggle to effectively classify unseen failure categories since the models are trained on limited samples of known failure categories. Consequently, the lack of sufficient training samples for newly emerging failure types could lead to a degradation in classifier performance, hindering accurate failure predictions. To tackle this, we employ a semi-supervised learning approach, where clustering is integrated with annotations from operations personnel, and a mapping layer is maintained to form a closed loop. This contributes to enhancing the classifier's generalization capability, enabling effective performance in novel failure scenarios.

2.3 Preliminaries

1) GAE^[11]: It can reconstruct inputs given a context and has demonstrated successful applications in prediction tasks and graph clustering. GAE integrates graph neural networks (GNNs) and autoencoders (AEs) to facilitate representation learning and the reconstruction of graph data. The training goal of GAE is to minimize reconstruction error, which measures the discrepancy between reconstructed and original graphs. The earliest works, such as GAE and VGAE^[12], take a 2-layer GCN as the encoder and dot product for link prediction decoding. Subsequent GAE mostly adopts the structural reconstruction (e.g., ARVGA^[13]) following VGAE, or a combination of structural and feature reconstruction (e.g., MGAE^[18], GALA^[17], and GATE^[19]) as their objectives. Inspired by GraphMAE^[20], we introduce masks into the traditional GAE design and replace mean squared error (MSE) with softmax cross-entropy (SCE) as the loss function to address the design shortcomings of GAE in graph representation learning.

2) Generative self-supervised learning

Self-supervised learning extracts supervisory information from a large amount of unsupervised data and constructs pseudo-labels using inherent information of the dataset. In representation learning, it holds significant potential to replace traditional supervised learning^[21]. Contrastive learning has dominated self-supervised graph learning, yet the success of most such methods often relies on high-quality data augmentation^[22]. Alternatively, training stability can be achieved through additional strategies^[23–24]. Generative self-supervised learning aims to recover missing parts of input data and reconstruct the intrinsic features of the data, without relying on the aforementioned factors.

3 Approach

In this section, we introduce our multimodal root cause analysis framework. After processing multimodal data into time-series data and generating a topological graph, we employ a generative graph autoencoder to perform masked encoding on graph data to obtain high-quality features. Subsequently, we cluster the features, and experts label the resulting clusters. A mapping layer corresponding to the labeled clusters is then generated, forming a closed-loop root cause analysis system. This framework is primarily divided into three parts: SPOT threshold-based feature extraction, generative GAE training, and clustering with expert labeling. To address challenge 1, we utilize a masked graph autoencoder to extract high-quality feature vectors. For challenge 2, we design an online streaming clustering classification method. After clustering, experts label and generate a mapping layer, where each label corresponds to a failure cause for each cluster. Our framework is illustrated in Fig. 2.

3.1 Feature Extraction

3.1.1 Data Preprocessing

In multimodal data processing, we standardize the logs, metrics, and invocation chain data of the cases into time-series data. Metrics are inherently time-series data that require no additional processing.

Log data processing is based on the existing method Drain3^[25]. Drain is an online real-time log parsing method that utilizes a fixed-length tree model, commonly used for extracting log templates. It transforms unstructured log information into structured log templates and template parameters. This part of the work involves two main aspects of log processing. First, regular expressions are used for masking in Drain. They are primarily employed to mask frequent occurrences such as IP addresses, dates, URLs, port numbers, and IDs. Frequent occurrences of these elements may cause the Drain model to mistakenly consider them as template constants when, in reality, they are template parameters. Second, to present the templates more concisely, consecutive masked parameters are merged into one. Due to the large and heterogeneous nature of

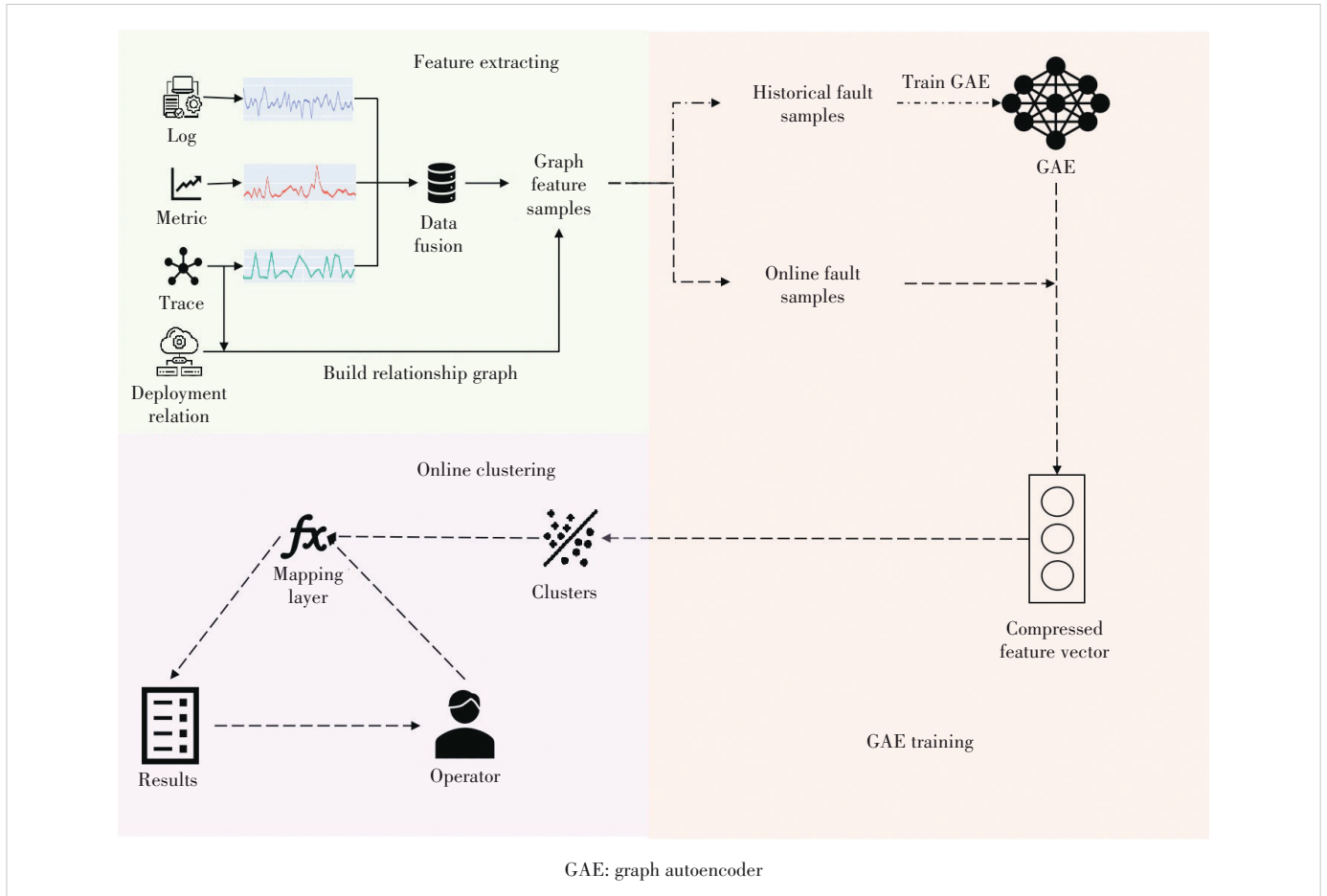


Figure 2. Framework of proposed multimodal root cause analysis method

the collected historical log data, there is a risk of significant noise in the extracted templates. Many templates, such as “*request starting http/2 post <*> application/grpc*”, may interfere with root cause analysis. To mitigate this situation, it is necessary to filter the extracted log templates. For example, in the 2022 AIOps Challenge dataset, this work filters out log templates containing useful information, such as “*severity: info message: <*>*”, “*severity: debug message: <*>*”, “*severity: error message: <*>*”, “*severity: warning message: <*>*” and “*file <*> line <*> in <*>*”. Noise templates like “*request starting http/2 post <*> application/grpc*” are uniformly defined as an “other” template.

Based on preliminary research and analysis of actual failure cases, this paper observes that different categories of failures result in significant differences in the types and frequencies of log entries. For example, log entries with templates like “*file <*> line <*> in <*>*” may only be recorded during certain failure scenarios, while templates like “*severity: warning message: <*>*” may be extensively recorded in specific types of failure cases. Based on this observation, each log template is treated as a type of time-series data, and the frequency of appearance per minute is recorded as a numerical

value. This transforms logs into time-series data for subsequent processing.

For call chain data, the focus is primarily on the call latency, call frequency, and returned status codes recorded. To unify it into time-series data, we first aggregate the call chains by microservice instances. We then calculate the count and average latency per minute for each execution segment. Subsequently, we filter out abnormal return status codes (such as `rpc_9` and `http_13`) and record the frequency of each type per minute. This yields time series data derived from the call chain data.

3.1.2 Feature Extraction Based on SPOT Dynamic Thresholds

We utilize the SPOT algorithm based on extreme value theory to compute thresholds and subsequently calculate anomaly scores for the windowed data of each metric. This approach effectively filters out irrelevant metrics without anomalies while extracting root cause features for anomalous metrics. Initially, historical data from all instances within the same microservice system over the past day are collected. The SPOT algorithm is applied to the entire dataset for initialization, calculating thresholds for each metric. Each instance

then uses the corresponding metric threshold during subsequent anomaly score calculations. Two types of anomaly scores are defined in this paper: deviation severity and time-weighted abnormal duration scores, which measure the degree of metric deviation from normal values and the time-weighted duration of abnormalities, respectively.

1) Deviation severity score: It measures the dimensionless deviation of an indicator from its normal value. A higher value indicates a higher degree of anomaly, suggesting a higher probability of the indicator being relevant to a failure category. The calculation for deviation severity is shown in Eq. (1), where T represents the duration of the failure, t represents a specific moment within the failure duration, x_t represents the indicator value at time t , and z_q represents the SPOT threshold for that indicator.

$$\text{Score} = \max_{t \in T} \left\{ \max \left(0, \frac{x_t - z_q}{z_q} \right) \right\} \quad (1).$$

2) Time-weighted abnormal duration score: Based on the analysis of historical failures and empirical findings, indicators relevant to failure categories exhibit two characteristics in the time dimension: longer abnormal durations and earlier anomaly occurrences. Hence, abnormal duration is incorporated as a feature. Additionally, to highlight the universality of the empirical observation that earlier anomalies are more likely to be root cause indicators, time weighting is incorporated into the calculation of the abnormal duration score. The feature is defined in Eq. (2), where T represents the failure duration, t_{inject} the moment of failure occurrence, t is the moment within the failure duration when the indicator value exceeds the corresponding SPOT threshold, γ is the time-weighting distribution exponent (with a default value of 1 assumed in this

paper), and I represents the time gap between two consecutive data records of the indicator.

$$\text{Duration} = \sum_{t \in T} \frac{1}{\left(|t - t_{\text{inject}}| \right)^\gamma} \times I \quad (2).$$

3.2 GAE-Based Feature Dimensionality Reduction

The network architecture of the graph autoencoder selected in this paper is illustrated in Fig. 3.

Details regarding model inputs, outputs, and the computation process are as follows:

1) Model input: Each feature vector of microservice instances obtained through feature extraction and the graph feature samples constructed from the topology graph are used as the model inputs.

2) Masked feature reconstruction: Inspired by GraphMAE^[20], we adopt a masked autoencoder. Actually, the idea of employing masking as the corruption strategy in masked autoencoders has found wide applications in computer vision (CV) and natural language processing (NLP). Formally, we sample a subset of nodes $\tilde{\nu} \subset \nu$ and mask each of their features with a token [MASK], e.g., a learnable vector $x_{[M]}$. The node feature \tilde{x}_i of $v_i \in \nu$ in the masked feature matrix \tilde{X} can be defined in Eq. (3).

$$\tilde{x}_i = \begin{cases} x_{[M]}, & v_i \in \tilde{\nu} \\ x_i, & v_i \notin \tilde{\nu} \end{cases} \quad (3).$$

The objective is to reconstruct the masked features of nodes in $\tilde{\nu}$ given the partially observed node signals \tilde{X} and the input adjacency matrix A .

3) SAGEConv layer: This layer computes the vector repre-

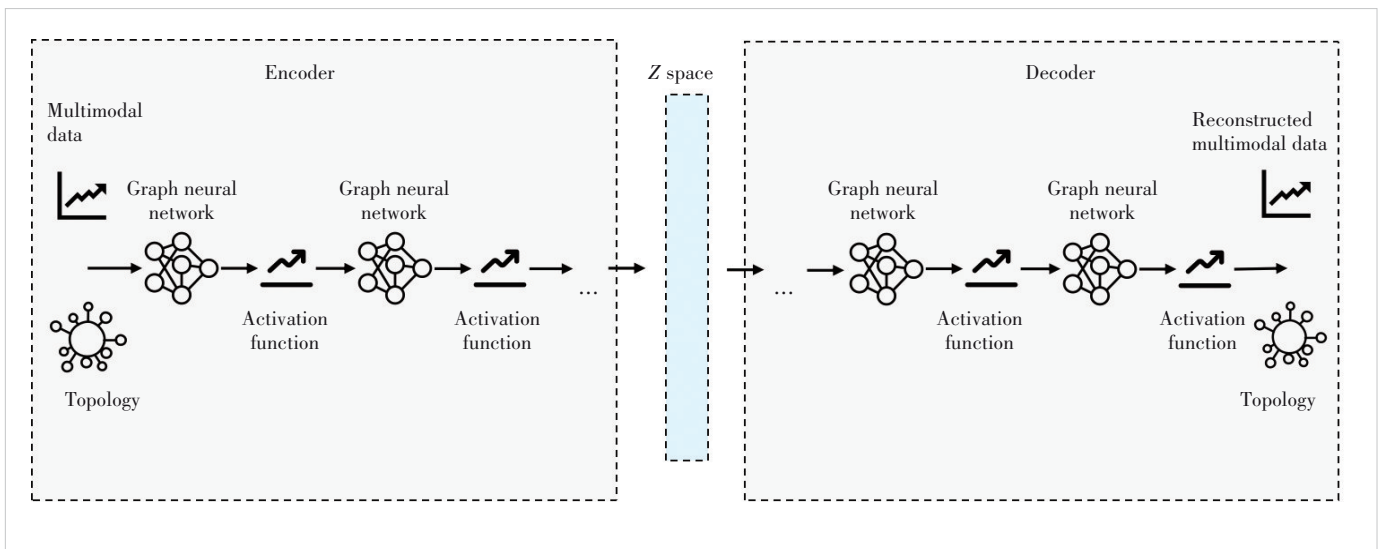


Figure 3. Graph autoencoder structure

sensation of each instance in the topology graph by aggregating information from neighboring instances. The computation process is expressed in Eqs. (4) and (5). Here, i denotes the instance node to be computed, N_i represents the set of neighboring instances of the instance node i , and “aggregate” denotes the aggregation method for neighboring instances; h_i^{l+1} is the $(l+1)$ -th feature vector representation for instance node i .

$$h_{N_i}^{l+1} = \text{aggregate}\left(\{h_j^l, \forall j \in N_i\}\right) \quad (4),$$

$$h_i^{l+1} = \sigma\left(W \cdot \text{concat}\left(h_i^l, h_{N_i}^{l+1}\right)\right) \quad (5).$$

4) Scaled cosine error: It is adopted in our framework instead of MSE used by existing GAE models to tackle imbalance between easy and hard samples during reconstruction. The intuition is that we can down weight easy samples’ contribution in training by scaling the cosine error with a power of $\gamma \geq 1$. Formally, given the original feature X and the reconstructed output $Z = f_D(A, \tilde{H})$, SCE is defined in Eq. (6):

$$L_{\text{SCE}} = \frac{1}{\tilde{V}} \sum_{v_i \in \tilde{V}} \left(1 - \frac{x_i^T z_i}{\|x_i\| \cdot \|z_i\|}\right)^\gamma, \gamma \geq 1 \quad (6),$$

which is averaged over all masked nodes. The scaling factor γ is a hyper-parameter adjustable over different datasets. This scaling technique can also be viewed as adaptive sample re-weighting, and the weight of each sample is adjusted with the reconstruction error. This error is also known in supervised ob-

ject detection as the focal loss.

3.3 Online Clustering

As shown in Fig. 4, to accurately identify failure categories and facilitate targeted troubleshooting for operations personnel, an online clustering process is designed based on the feature vectors previously obtained. The main components of the process include distance measurement, maintenance of cluster center sample caching, and mapping of clusters to failure categories based on expert feedback. This method does not require continuous attention from operations personnel; instead, only when new categories appear in the cluster cache, they are required to be annotated, improving operational efficiency while reducing the workload of expert feedback annotations.

The maintenance strategy for cluster center sample caching is as follows: If the current sample is assigned to a cluster where the number of center samples is less than the buffer limit (set to 5 in the algorithm), it is directly added to the cache; otherwise, the algorithm compares its distance with the existing samples in the cluster center cache. If the distance is smaller, it replaces the sample with the largest distance from the cluster center in the cache; otherwise, there is no need to update the cluster center sample cache.

To ensure more precise classification results and meaningful clusters, we introduce a mapping of clusters to failure categories based on expert feedback in the process. This involves experts labeling each cluster and maintaining a mapping layer from cluster categories to failure categories.

For each input failure to be detected, if its clustering result maps to a known failure category, the failure category can be

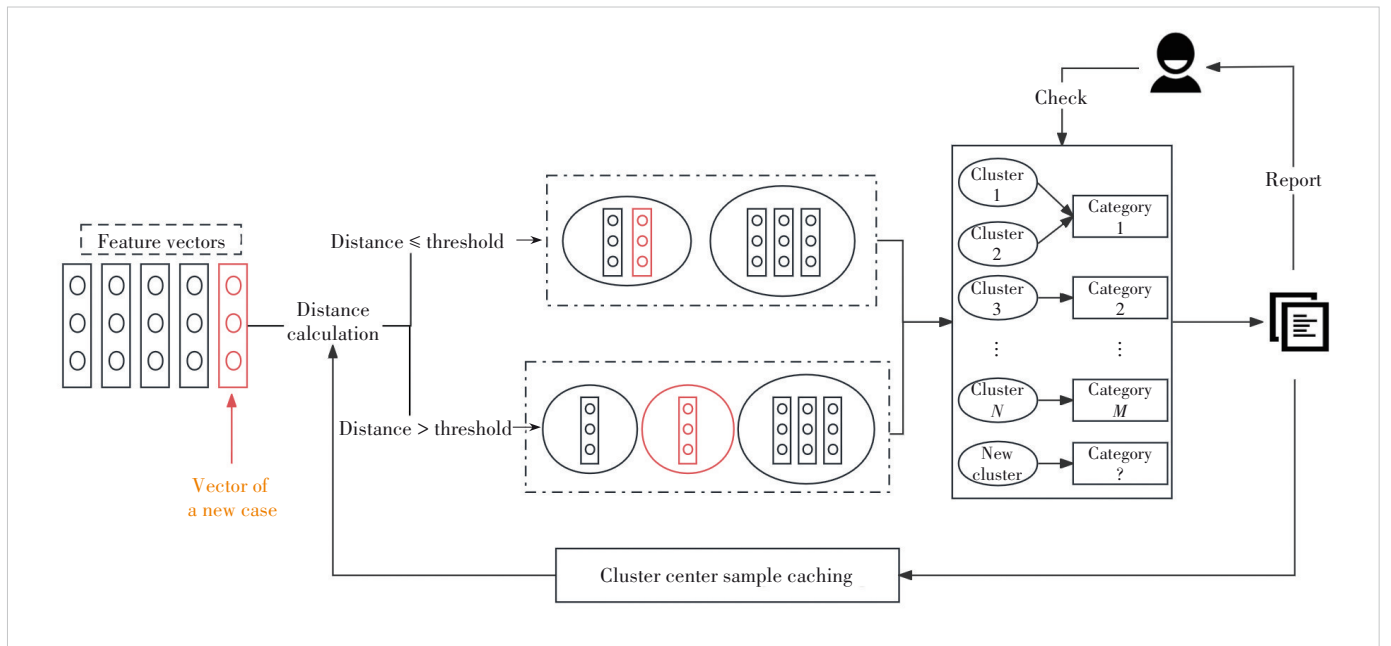


Figure 4. Online clustering process

directly output as the diagnostic result. If it belongs to an unknown failure category, a new cluster is created, reported to the operations personnel for failure category labeling, and the cluster center sample cache and mapping layer are simultaneously updated according to a predefined strategy.

4 Experiment and Evaluation

In this section, we evaluate the performance of our framework with the following research questions (RQs):

- RQ1: How effective is our framework in multi-modal root cause analysis?
- RQ2: How effective is the improvement of the model's performance with multimodal data and the use of GAE?

4.1 Experimental Setup

1) Datasets

This paper evaluates the performance of the method for microservice systems using two datasets: 22AIOps (we define it as D1) and 21AIOps (D2).

D1 is derived from the 2022 AIOps challenge, focusing on "failure identification and classification in microservice architecture e-commerce systems". The dataset was adapted by the organizers based on the open-source HipsterShop from Google, simulating a microservice architecture in an e-commerce system. The system comprises 10 microservices, each deploying 4 instances. Each microservice has its corresponding container monitor, and the deployed virtual machines have associated monitoring data. The failure scenarios in this dataset are derived from real system failures and are replayed in batches. To collect failure records, operators conducted a failure replay in the system for several days in May 2022. The recorded failures were then labeled with their respective actual root cause instances. Injected failures represent five common failure types identified in the industry, i.e., CPU-related, memory-related, Java virtual machine (JVM)-CPU-related, JVM-memory-related, and input/output (IO)-related failures, with corresponding annotations for injected failure root causes. The dataset consists of 174 microservice-level failure cases, which are divided into training and testing sets with a ratio of 8:2.

D2 is sourced from the 2021 AIOps Challenge, focusing on "real-time failure detection and root cause localization in commercial bank application systems in cloud environments". This dataset is generated by simulating real applications from two large commercial banks. The system includes 9 microservices, each deploying 2 instances. Similar to the 2022AIOps dataset, the injected failures represent five common failure types with corresponding root cause annotations. The dataset comprises 159 microservices-level failure cases, which are divided into training and test sets with a ratio of 8:2. The details are shown in Table 1.

2) Baselines

We compare our framework with two baselines: CloudRCA^[17] and DiagFusion^[13]. They both use GNNs to capture and learn the topological features of microservices. DiagFusion integrates and processes traces, logs, and metrics from multiple sources, transforms them into structured events, generates unified embeddings for diagnosis, and learns from historical failure cases to identify root cause instances and failure types. CloudRCA^[17] uses both metrics and logs. It uses the PC algorithm to learn the causal relationship among anomaly patterns of metrics, anomaly patterns of logs, and types of failure, and constructs a hierarchical Bayesian network to infer the failure type.

3) Evaluation metrics

We use Precision, Recall, and F1-Score as our evaluation metrics, which are defined as follows. Taking the failure type "A" as an example: True Positive (TP) represents the framework correctly predicts the presence of a fault. False Positive (FP) indicates the model incorrectly identifies a fault as type A when the real cause is type B. False Negative (FN) refers to cases where the model fails to classify a fault as type A when the real cause is type A. Precision represents the proportion of actual positive samples among those predicted as positive; Recall indicates the proportion of actual positive samples among all positive samples in the entire dataset; F1-Score combines Precision and Recall, with values ranging between 0 and 1, where a higher value indicates better classification performance of the model.

$$\text{Precision} = \frac{1}{N} \times \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i} \times \text{num}_i \quad (7),$$

$$\text{Recall} = \frac{1}{N} \times \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i} \times \text{num}_i \quad (8),$$

$$\text{F1-Score} = \frac{1}{N} \times \sum_{i=1}^N \frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \times \text{num}_i \quad (9).$$

4.2 Experimental Results

1) RQ1: How effective is our framework in multi-modal root

Table 1. Datasets' details

Dataset	D1		D2	
	Training Set	Testing Set	Training Set	Testing Set
Number of failure Cases	139	35	127	32
Number of system microservices	10		9	
Number of microservice instances	20		18	
Number of involved failure categories	5		5	

cause analysis?

We compare our framework with two baselines: DiagFusion and CloudRCA. The results are shown in Table 2, where both DiagFusion and CloudRCA show poor precision and recall performance, which can be ascribed to the utilization of datasets with low label ratio. They demand a substantial quantity of well-labeled failure cases for effective training; otherwise, achieving satisfactory performance will be difficult. Furthermore, DiagFusion has a specific constraint that the number of output neurons must be equal to that of nodes in the system. This limits its applicability in scenarios where the number of nodes dynamically changes, such as in systems with dynamic scaling or evolving architectures. CloudRCA, by contrast, uses only metrics and log data (trace modality is not included). Moreover, CloudRCA uses the PC algorithm, which infers causal relationships and learns graph structures based on observational data. This approach usually requires sufficient sample data for conducting conditional independence tests and inferring causal relationship inference; a lack of such data will reduce the accuracy of failure diagnosis.

The experimental results show that our framework has good performance for complex datasets in practice. Our method adopts GAE, which can alleviate the influence of noise by compressing high-dimensional data into a lower-dimensional space that more effectively captures the essence of the data. Moreover, it does not rely on supervised learning methods, eliminating the need for a large number of labeled samples for training. This contributes to the stability of our performance.

In addition, we utilize t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce the dimensionality of feature vectors for classification, aiming to observe the distribution of feature vectors for the same failure type, as shown in Fig. 5. It can be seen that the failure category distribution of the D2 dataset is more concentrated than that of the D1 dataset. When the distribution of feature vectors is more concentrated, the overlap between categories decreases, allowing classifiers to differentiate more easily between different categories based on the values of the feature vectors. This explains the reason why both baselines perform better on the D2 dataset than the D1 dataset.

2) How effective is the improvement in the model's performance achieved by using multimodal data and GAE?

As mentioned above, we use three types of data along with GAE. In this RQ, we will verify the effect of the main parts of

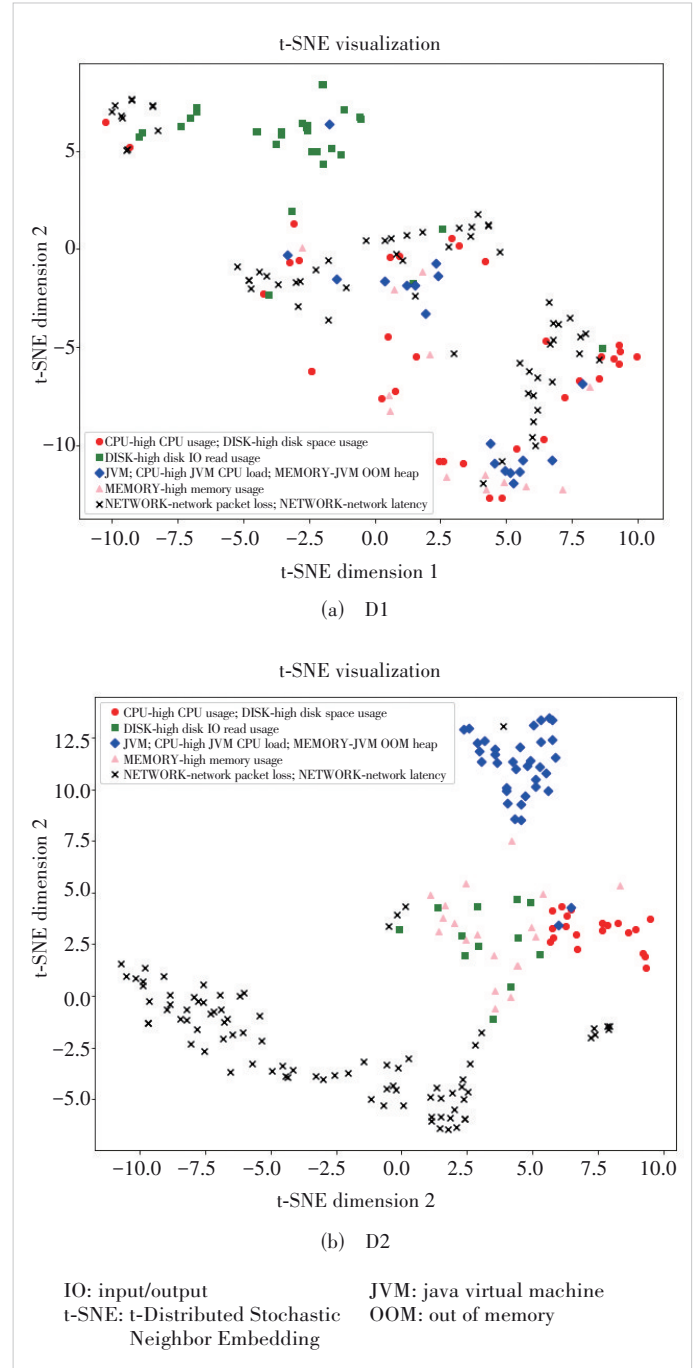


Figure 5. T-SNE visualization of feature vector distribution for the same failure type

Table 2. Comparison of proposed framework with two baselines

Dataset	D1				D2			
Approach	Precision	Recall	F1-Score	Labeled Ratio	Precision	Recall	F1-Score	Labeled Ratio
DiagFusion	0.955	0.239	0.382	0.161	0.879	0.424	0.572	0.163
CloudRCA	0.160	0.154	0.191	0.161	0.382	0.367	0.361	0.163
Our method	0.833	0.879	0.851	0.161	0.839	0.824	0.824	0.163

Table 3. Verification of main part of proposed framework without trace data, metric data, log data, or GAE

Dataset	22 AIOps Dataset		
Approach	Precision	Recall	F1-Score
OM w/o trace data	0.733	0.799	0.754
OM w/o log data	0.710	0.747	0.713
OM w/o metric data	0.256	0.506	0.340
OM w/o GAE	0.738	0.678	0.625
OM	0.833	0.879	0.851

GAE: graph autoencoder OM: our method

the framework on its effectiveness, by removing trace data, metric data, log data, and GAE. The results are shown in Table 3.

The experimental results indicate that each modality of data contains certain failure category information, with the metric data being the most information-rich. Therefore, the combined use of metrics, logs, and call chain data for root cause analysis can significantly enhance the effectiveness of this task. The results also demonstrate that the use of graph autoencoders for dimensionality reduction has the most pronounced impact on model performance, validating the capability of this module to mitigate the influence of noise and compress high-dimensional data into a lower-dimensional space that better reflects the intrinsic nature of the data.

5 Related Work

Previous anomaly detection approaches are usually based on system logs^[6-9], KPIs^[10-12], or both^[26], targeting traditional distributed systems without complex invocation relationships. Recently, combining multimodal data to conduct root cause analysis has drawn increasing attention. PDiagnosis^[19] combines metrics, logs, and traces to identify root causes. It employs lightweight anomaly detection across all three modalities to detect anomalous patterns. Based on a voting strategy, it selects the component with the most severe anomalies as the root cause. CloudRCA^[17], using both metrics and logs, employs the PC algorithm to learn the causal relationships among anomaly patterns of metrics, anomaly patterns of logs, and types of failure. It then constructs a hierarchical Bayesian network to infer the failure type.

In recent years, there has been a growing trend of using GNNs to capture and learn the topological features of microservices. Eadro^[16] unifies multimodal data into vectors and performs joint training for both anomaly detection and root cause localization. DiagFusion^[13], on the other hand, unifies multimodal data into events, generates unified embedding representations, and identifies root cause instances and failure types by learning from historical failure cases. However, both methods share a limitation: they require a large number of high-quality labeled failure cases for method training; otherwise, it is difficult to achieve good performance.

6 Conclusions and Future Work

This paper primarily investigates a root cause analysis method for microservice systems based on a masked graph auto-encoder. It first introduces the background and practical application value of root cause analysis in microservices systems. Subsequently, it reviews the current domestic and international state-of-the-art research on the topic and outlines the contribution of this paper. The proposed method consists of three key components: feature extraction based on the SPOT threshold, feature dimensionality reduction using a masked graph auto-encoder, and online clustering combined with expert feedback. Based on the designed workflow, this paper conducts specific experiments, validates the results, and analyzes the outcomes. It demonstrates the effectiveness of the proposed method for microservice systems based on a masked graph auto-encoder in root cause analysis tasks, showcasing certain advantages over various baseline methods. Nevertheless, the design of our method can be further improved in the following areas:

- The experiments in this paper evaluate the effectiveness of the root cause analysis method on the datasets of the AIOps Challenge in 2021 and 2022. However, its effectiveness has not been validated in large-scale production environments with microservice systems.
- This paper transforms logs and call chains into time series data for the fusion of multimodal data. However, it does not consider the mutual influence between multimodal data. The method simply converts multimodal data inputs into a unified form. How to organically integrate multimodal data and simulate the propagation of failures between data modalities is a future research direction.

References

- [1] YI X X, ZHANG N H, LIU Y C, et al. Key technologies for intelligent computing power network [J]. ZTE technology journal, 2025, 31(2): 31 - 38. DOI: 10.12142/ZTETJ.202502005
- [2] WU H Q. Reflections on AI-empowered network reconstruction [J]. ZTE technology journal, 2025, 31(1): 1 - 3. DOI: 10.12142/ZTETJ.202501001
- [3] ZHANG S L, XIA S B, FAN W Z, et al. Failure diagnosis in microservice systems: a comprehensive survey and analysis [J]. ACM transactions on software engineering and methodology, 2025, (Just Accepted). DOI: 10.1145/3715005
- [4] JIN M X, LV A R, ZHU Y P, et al. An anomaly detection algorithm for microservice architecture based on robust principal component analysis [J]. IEEE access, 2020, 8: 226397 - 226408
- [5] YU G B, CHEN P F, CHEN H Y, et al. MicroRank: end-to-end latency issue localization with extended spectrum analysis in microservice environments [C]//Proceedings of the Web Conference 2021. ACM, 2021: 3087 - 3098. DOI: 10.1145/3442381.3449905
- [6] LIN Q W, ZHANG H Y, LOU J G, et al. Log clustering based problem identification for online service systems [C]//The 38th International Conference on Software Engineering Companion (ICSE-C). IEEE, 2016: 102 - 111
- [7] YUAN Y, SHI W C, LIANG B, et al. An approach to cloud execution failure diagnosis based on exception logs in OpenStack [C]//Proceedings of

- IEEE 12th International Conference on Cloud Computing (CLOUD). IEEE, 2019: 124 – 131. DOI: 10.1109/cloud.2019.00031
- [8] ZHANG X, XU Y, QIN S, et al. Onion: identifying incident-indicating logs for cloud systems [C]//The 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, 2021: 1253 – 1263. DOI: 10.1145/3468264.3473919
- [9] DU M, LI F F, ZHENG G N, et al. DeepLog: anomaly detection and diagnosis from system logs through deep learning [C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017: 1285 – 1298. DOI: 10.1145/3133956.3134015
- [10] MA M, LIN W L, PAN D S, et al. Self-adaptive root cause diagnosis for large-scale microservice architecture [J]. IEEE transactions on services computing, 2022, 15(3): 1399 – 1410. DOI: 10.1109/TSC.2020.2993251
- [11] LIN J J, CHEN P F, ZHENG Z B. Microscope: pinpoint performance issues with causal graphs in micro-service environments [M]//Service-oriented computing. Cham: Springer International Publishing, 2018: 3 – 20. DOI: 10.1007/978-3-030-03596-9_1
- [12] MA M, XU J M, WANG Y, et al. AutoMAP: diagnose your microservice-based web applications automatically [C]//Proceedings of The Web Conference 2020. ACM, 2020: 246 – 258. DOI: 10.1145/3366423.3380111
- [13] ZHANG S, JIN P, LIN Z, et al. Robust root cause analysis of microservice system through multimodal data [J]. IEEE transactions on services computing, 2023:1 – 14
- [14] SUN Y Q, SHI B P, MAO M Y, et al. ART: a unified unsupervised framework for incident management in microservice systems [C]//Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering. ACM, 2024: 1183 – 1194. DOI: 10.1145/3691620.3695495
- [15] SUN Y Q, LIN Z H, SHI B P, et al. Interpretable failure localization for microservice systems based on graph autoencoder [J]. ACM transactions on software engineering and methodology, 2025, 34(2): 52. DOI: 10.1145/3695999
- [16] LEE C, YANG T Y, CHEN Z B, et al. Eadro: an end-to-end troubleshooting framework for microservices on multi-source data [C]//The 45th International Conference on Software Engineering (ICSE). IEEE, 2023: 1750 – 1762. DOI: 10.1109/ICSE48619.2023.00150
- [17] ZHANG Y Y, GUAN Z X, QIAN H J, et al. CloudRCA: a root cause analysis framework for cloud computing platforms [C]//The 30th ACM International Conference on Information & Knowledge Management. ACM, 2021: 4373 – 4382. DOI: 10.1145/3459637.3481903
- [18] WANG C, PAN S R, LONG G D, et al. Mgae: marginalized graph autoencoder for graph clustering [C]//Conference on Information and Knowledge Management. ACM, 2017: 889 – 898
- [19] HOU C J, JIA T, WU Y F, et al. Diagnosing performance issues in microservices with heterogeneous data source [C]//Proceedings of IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking. IEEE, 2021: 493 – 500. DOI: 10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00074
- [20] HOU Z Y, LIU X, CEN Y K, et al. GraphMAE: self-supervised masked graph autoencoders [C]//The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, 2022: 594 – 604. DOI: 10.1145/3534678.3539321
- [21] SIGELMAN B H, BARROSO L A, BURROWS M, et al. Dapper, a large-scale distributed systems tracing infrastructure [EB/OL]. [2024-01-26] <https://research.google.com/archive/papers/dapper-2010-1.pdf>
- [22] Opentracing. Opentracing [EB/OL]. [2024-01-26]. <https://opentracing.io>
- [23] SIFFER A, FOUQUE P A, TERMIER A, et al. Anomaly detection in streams with extreme value theory [C]//The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017: 1067 – 1075. DOI: 10.1145/3097983.3098144
- [24] KIPF T N, WEWELLING M. Variational graph auto-encoders [EB/OL]. (2016-11-21) [2024-01-26]. <https://arxiv.org/abs/1611.07308>
- [25] HE P J, ZHU J M, ZHENG Z B, et al. Drain: an online log parsing approach with fixed depth tree [C]//International Conference on Web Services (ICWS). IEEE, 2017: 33 – 40. DOI: 10.1109/ICWS.2017.13
- [26] PARK J, LEE M, CHANG H J, et al. Symmetric graph convolutional auto-encoder for unsupervised graph representation learning. (2018-09-27) [2024-01-26]. <https://arxiv.org/abs/1908.02441>

Biographies

LI Yingke received her BS degree in software engineering from the School of Information Engineering, Minzu University of China in 2018. She is currently pursuing her master's degree at the College of Software, Nankai University, China. Her research interests include anomaly detection and failure diagnosis.

HAN Jing (han.jing28@zte.com.cn) received her master's degree from Nanjing University of Aeronautics and Astronautics, China. She has been with ZTE Corporation since 2000. She had been engaged in 3G/4G key technologies, from 2000 to 2016, and has become a technical director responsible for intelligent operation of cloud platforms and wireless networks since 2016. Her research interests include machine learning, data mining, and signal processing.

SUN Yongqian received his BS degree in statistical specialty from Northwestern Polytechnical University, China in 2012, and PhD degree in computer science from Tsinghua University, China in 2018. He is currently an assistant professor with the College of Software, Nankai University, China. His research focuses on anomaly detection, root cause analysis, and failure diagnosis in service management.

SHI Binpeng received his BE degree in software engineering from the College of Software, Nankai University, China in 2023, where he is currently pursuing his master's degree. His research interests include anomaly detection and failure diagnosis.

GONG Zican received his master's degree in professional computing and artificial intelligence from the Australian National University in 2019. He has been a machine learning engineer at ZTE Corporation since 2020. His research interests include machine learning, professional computing and system architecture.

ZTE Communications

Table of Contents, Volume 23, 2025

Volume-Number-Page

Special Topic

Efficient Spatio-Temporal Predictive Learning for Massive MIMO CSI Prediction

Editorial	YANG Kun, JIN Shi, XIANG Luping	23-01-01
Efficient Spatio-Temporal Predictive Learning for Massive MIMO CSI Prediction	CHENG Jiaming, CHEN Wei, LI Lun, AI Bo	23-01-03
RIS Enabled Simultaneous Transmission and Key Generation with PPO: Exploring Security Boundary of RIS Phase Shift	FAN Kaiqing, YAO Yuze, GAO Ning, LI Xiao, JIN Shi	23-01-11
Endogenous Security Through AI-Driven Physical-Layer Authentication for Future 6G Networks	MENG Rui, FAN Dayu, XU Xiaodong, LYU Suyu, TAO Xiaofeng	23-01-18
Separate Source Channel Coding Is Still What You Need: An LLM-Based Rethinking	REN Tianqi, LI Rongpeng, ZHAO Mingmin, CHEN Xianfu, LIU Guangyi, YANG Yang, ZHAO Zhifeng, ZHANG Honggang	23-01-30
Exploration of NWDAF Development Architecture for 6G AI-Native Networks	HE Shiwen, PENG Shilin, DONG Haolei, WANG Liangpeng, AN Zhenyu	23-01-45
Device Activity Detection and Channel Estimation Using Score-Based Generative Models in Massive MIMO	TANG Chenyue, LI Zeshen, CHEN Zihan, Howard H. YANG	23-01-53
Efficient PSS Detection Algorithm Aided by CNN	LI Lanlan	23-01-63

Digital Twin Online Channel Modeling for 6G and Beyond

Editorial	WANG Chengxiang, HUANG Chen	23-02-01
Channel Measurement and Analysis of Human Body Radar Cross Section in 26 GHz ISAC Systems	DUAN Hongyu, WANG Mengyang, DUO Hao, HE Danping, MA Yihua, LU Bin, ZHONG Zhangdui	23-02-03
Space Network Emulation System Based on a User-Space Network Stack	LEI Jianzhe, ZHAO Kanglian, HOU Dongxu, ZHOU Fenlin	23-02-11
A Machine Learning-Based Channel Data Enhancement Platform for Digital Twin Channels	AI Bo, ZHANG Yuxin, YANG Mi, HE Ruisi, GUO Rongge	23-02-20
6G Digital Twin Enabled Channel Modeling for Beijing Central Business District	LU Mengyuan, BAI Lu, HAN Zengrui, HUANG Ziwei, LU Shiliang, CHENG Xiang	23-02-31
Channel Knowledge Maps for 6G Wireless Networks: Construction, Applications, and Future Challenges	LIU Xingchen, SUN Shu, TAO Meixia, Aryan KAUSHIK, YAN Hangsong	23-02-46
Air-to-Ground Channel Measurement and Modeling for Low-altitude UAVs: A Survey	CHEN Peng, LIU Yajuan, WEI Wentong, WANG Wei, LI Na	23-02-60

Security of Large Models

Editorial	SU Zhou, DU Linkang	23-03-01
Poison-Only and Targeted Backdoor Attack Against Visual Object Tracking	GU Wei, SHAO Shuo, ZHOU Lingtao, QIN Zhan, REN Kui	23-03-03
VOTI: Jailbreaking Vision-Language Models via Visual Obfuscation and Task Induction	ZHU Yifan, CHU Zhixuan, REN Kui	23-03-15
From Function Calls to MCPs for Securing AI Agent Systems: Architecture, Challenges and Countermeasures	WANG Wei, LI Shaofeng, DONG Tian, MENG Yan, ZHU Haojin	23-03-27
Dataset Copyright Auditing for Large Models: Fundamentals, Open Problems, and Future Directions	DU Linkang, SU Zhou, YU Xinyi	23-03-38
StegoAgent: A Generative Steganography Framework Based on GUI Agents	SHEN QiuHong, YANG Zijin, JIANG Jun, ZHANG Weiming, CHEN Kejiang	23-03-48

New Generation FTTR Communication and Networking Technology

Editorial	GE Xiaohu, ZHONG Yi	23-04-01
PON Monitoring Scheme Based on TGD-OFDR with High Spatial Resolution and Dynamic Range	ZHU Yidai, FAN Xinyu, ZHU Songlin, DONG Jiaxing, LI Guoqiang, HE Zuyuan	23-04-03
Insights on Next Generation WLAN: High Experiences (HEX)	YANG Mao, LI Bo, YAN Zhongjiang	23-04-10
FTTR-MmWave Architecture for Next-Generation Indoor High-Speed Communications	CHEN Zhe, ZHOU Peigen, WANG Long, HOU Debin, HU Yun, CHEN Jixin, HONG Wei	23-04-16
A Transformer-Based End-to-End Receiver Design for Wi-Fi 7 Physical Layer	LIU Yichen, GAO Ruixin, ZENG Chen, LIU Yingzhuang	23-04-27
Root Cause Analysis of Poor FTTR Quality Based on Transformer Mechanisms	YU Weichao, LIU Yang, ZHANG Junxiong, YE Junliang, GE Xiaohu	23-04-37
QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer for FTTR Networks	CAI Jinhan, ZAN Mingyuan, SHEN Gangxiang	23-04-48
C-WAN for FTTR: Enabling Low-Overhead Joint Transmission with Deep Learning	ZHANG Yang, CEN Zihan, ZHAN Wen, CHEN Xiang	23-04-65

Research Papers

A Basis Function Generation Based Digital Predistortion Concurrent Neural Network Model of RF Power Amplifier	SHAO Jianfeng, HONG Xi, WANG Wenjie, LIN Zeyu, LI Yunhua	23-01-71
A Wide Passband Frequency Selective Surface with Angular Stability	TANG xingyang, SUI Jia, FU Jiahui, YANG Kaiwen, ZHAO Zhipeng	23-01-78
Dual-Polarized 2D Beam-Scanning Antenna Based on Reconfigurable Reflective Elements	LIU Zhipeng, LI Kexin, CAI Yuanming, LIU Feng, GUO Jiayin	23-01-85
VFabric: A Digital Twin Emulator for Core Switching Equipment	WANG Qianglin, ZHANG Xiaoning, YANG Yi, FAN Chenyu, YUE Yangyang, WU Wei, DUAN Wei	23-01-90
Precise Location of Passive Inter-Modulation in Long Cables by Fractional Frequencies Based Multi-Range Rulers	DONG Anhua, LIANG Haodong, ZHU Shaohao, ZHANG Qi, ZHAO Deshuang	23-01-101
Measurement and Analysis of Radar-Cross-Section of UAV at 21 – 26 GHz Frequency Band	AN Hao, LIU Ting, HE Danping, MA Yihua, DOU Jianwu	23-01-107

Doppler Rate Estimation for OTFS via Large-scale Antenna Array	SHAN Yaru, WANG Fanggang, HAO Yaxing, HUA Jian, XIN Yu	23-01-115
GaN-Based Optoelectronic Impact Force Sensor	RUAN Junhui, JIANG Chengxiang, XU Shengli, WANG Yongjin, SHI Fan	23-02-96
Intelligent AP Clustering and Receiver Design for Uplink Cell-free Networks	AN Zhenyu, HE Shiwen, YANG Li, ZHAN Hang, HUANG Yongming	23-02-103
Integrated All-Light Network for Air, Space, Land, and Sea	LIANG Yingze, WANG Linning, QI Ziqian, LIU Pengzhan, WANG Yongjin	23-02-109
Real-Time 7-Core SDM Transmission System Using Commercial 400 Gbit/s OTN Transceivers and Network Management System	CUI Jian, GU Ninglun, CHANG Cheng, SHI Hu, YAN Baoluo	23-03-81
Antenna Parameter Calibration for Mobile Communication Base Station via Laser Tracker	LI Junqiang, CHEN Shijun, FENG Yujie, FAN Jiancun, CHEN Qiang	23-03-89
M+MNet: A Mixed-Precision Multibranch Network for Image Aesthetics Assessment	HE Shuai, LIU Limin, WANG Zhanli, LI Jinliang, MAO Xiaojun, MING Anlong	23-03-96
Empowering Grounding DINO with MoE: An End-to-End Framework for Cross-Domain Few-Shot Object Detection	DONG Xiugang, ZHANG Kaijin, NONG Qingpeng, JU Minhan, TU Yaofeng	23-04-77
Shortened PAC Codes and List Decoding	LIU Aolin, FENG Bowen, LIANG Chulong, XU Jin, ZHANG Qinyu	23-04-86
Full-Duplex Massive MIMO Self-Interference Suppression Based on Beamforming	ZHANG Boyu, ZHANG Ling, LI Zijiang, SHEN Ying	23-04-97
A Root Cause Analysis Framework for Microservice Systems with Multimodal Data	LI Yingke, HAN Jing, SUN Yongqian, SHI Binpeng, GONG Zican	23-04-110

Review

Liquid Neural Networks: Next-Generation AI for Telecom from First Principles	ZHU Fenghao, WANG Xinquan, ZHU Chen, HUANG Chongwen	23-02-76
Overview of Cross-Component In-Loop Filters in Video Coding Standards	LI Zhaoyu, MENG Xuwei, ZHANG Jiaqi, HUANG Cheng, JIA Chuanmin, MA Siwei, JIANG Yun	23-02-85
Learned Distributed Query Optimizer: Architecture and Challenges	GAO Jun, HAN Yinjun, LIN Yang, MIAO Hao, XU Mo	23-02-96
Analysis of Feasible Solutions for Railway 5G Network Security Assessment	XU Hang, SUN Bin, DING Jianwen, WANG Wei	23-03-59
Key Techniques and Challenges in NeRF-Based Dynamic 3D Reconstruction	LU Ping, FENG Daquan, SHI Wenzhe, LI Wan, LIN Jiaxin	23-03-71

The 1st Youth Expert Committee

for Promoting Industry-University-Institute Cooperation

Director CHEN Wei, Beijing Jiaotong University
Deputy Director QIN Xiaoqi, Beijing University of Posts and Telecommunications
LU Dan, ZTE Corporation

Members (Surname in Alphabetical Order)

CAO Jin	Xidian University
CHEN Li	University of Science and Technology of China
CHEN Qimei	Wuhan University
CHEN Shuyi	Harbin Institute of Technology
CHEN Siheng	Shanghai Jiao Tong University
CHEN Wei	Beijing Jiaotong University
GUAN Ke	Beijing Jiaotong University
HAN Kaifeng	China Academy of Information and Communications Technology
HE Zi	Nanjing University of Science and Technology
HOU Tianwei	Beijing Jiaotong University
HU Jie	University of Electronic Science and Technology of China
HUANG Chen	Purple Mountain Laboratories
LI Ang	Xi'an Jiaotong University
LIU Chunsen	Fudan University
LIU Fan	Southeast University
LIU Junyu	Xidian University
LU Dan	ZTE Corporation
LU Youyou	Tsinghua University
NING Zhaolong	Chongqing University of Posts and Telecommunications
QI Liang	Shanghai Jiao Tong University
QIN Xiaoqi	Beijing University of Posts and Telecommunications
QIN Zhijin	Tsinghua University
SHI Yinghuan	Nanjing University
TANG Wankai	Southeast University
WANG Jingjing	Beihang University
WANG Xinggang	Huazhong University of Science and Technology
WANG Yongqiang	Tianjin University
WEN Miaowen	South China University of Technology
WU Qingqing	Shanghai Jiao Tong University
WU Yongpeng	Shanghai Jiao Tong University
XIA Wenchao	Nanjing University of Posts and Telecommunications
XU Mengwei	Beijing University of Posts and Telecommunications
XU Tianheng	Shanghai Advanced Research Institute, Chinese Academy of Sciences
YANG Chuanchuan	Peking University
YIN Haifan	Huazhong University of Science and Technology
YU Jihong	Beijing Institute of Technology
ZHANG Jiao	Beijing University of Posts and Telecommunications
ZHANG Yuchao	Beijing University of Posts and Telecommunications
ZHANG Jiayi	Beijing Jiaotong University
ZHAO Yuda	Zhejiang University
ZHAO Zhongyuan	Beijing University of Posts and Telecommunications
ZHOU Yi	Southwest Jiaotong University
ZHU Bingcheng	Southeast University

ZTE COMMUNICATIONS

中兴通讯技术(英文版)

ZTE Communications has been indexed in the following databases:

- Abstract Journal
- China Science and Technology Journal Database
- Chinese Journal Fulltext Databases
- Index Copernicus
- Scopus
- Ulrich's Periodicals Directory
- Wanfang Data
- WJCI 2021-2024

Industry Consultants:

DUAN Xiangyang, GAO Yin, HU Liujun, HUA Xinhai, LIU Xinyang,
SHI Weiqiang, TU Yaofeng, WANG Huitao, XIONG Xiankui, XU Jin,
YAN Xincheng, ZHAO Yajun, ZHU Xiaoguang

ZTE COMMUNICATIONS

Vol. 23 No. 4 (Issue 93)

Quarterly

First Issue Published in 2003

Supervised by:

Anhui Publishing Group

Sponsored by:

Time Publishing and Media Co., Ltd.

Shenzhen Guangyu Aerospace Industry Co., Ltd.

Published by:

Anhui Science & Technology Publishing House

Edited and Circulated (Home and Abroad) by:
Magazine House of ZTE Communications

Staff Members:

General Editor: WANG Xiyu

Editor-in-Chief: TAO Shanyong

Executive Editor-in-Chief: HUANG Xinming

Deputy Editor-in-Chief: LU Dan

Editorial Director: WANG Pingping

Editor-in-Charge: ZHU Li

Editors: REN Xixi, XU Ye, YANG Guangxi

Producer: XU Ying

Circulation Executive: WANG Pingping

Assistant: WANG Kun

Editorial Correspondence:

Add: 12F Kaixuan Building, 329 Jinzhai Road,
Hefei 230061, P. R. China

Tel: +86-551-65533356

Email: magazine@zte.com.cn

Website: <http://zte.magtechjournal.com>

Annual Subscription: RMB 120

Printed by:

Anhui Tianjin Printing Technology Co., Ltd.

Publication Date: December 25, 2025

China Standard Serial Number: ISSN 1673-5188
CN 34-1294/TN