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

ZHANG Yang[1], CEN Zihan[1], ZHAN Wen[2], CHEN Xiang[1]
(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] estimates 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:

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

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 $\mathbb{E}\left[|s|^2\right] = 1$, 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:

$$P_{\text{rx}} = \left| \sum_{k=1}^{K} \sqrt{P_k}\, h_k \right|^2 \tag{2}.$$

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: $P_{\text{rx, non-coop}} = \sum_{k=1}^{K} P_k |h_k|^2$. In contrast, JT achieves additional coherent combining gain, expressed as $P_{\text{rx}}/P_{\text{rx, non-coop}}$, which is typically greater than 1. This gain significantly enhances both spectral efficiency and transmission reliability compared to non-cooperative transmission.

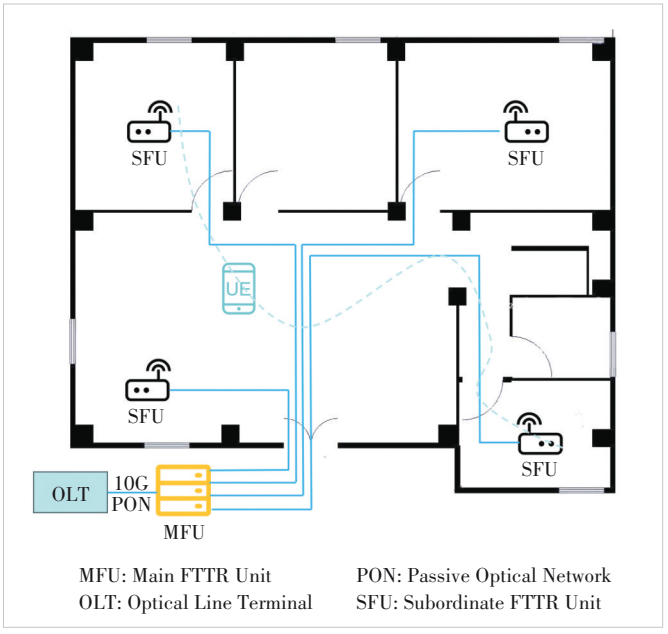


Figure 1. Typical Fiber-to-the-Room (FTTR) system

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:

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

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:

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

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),

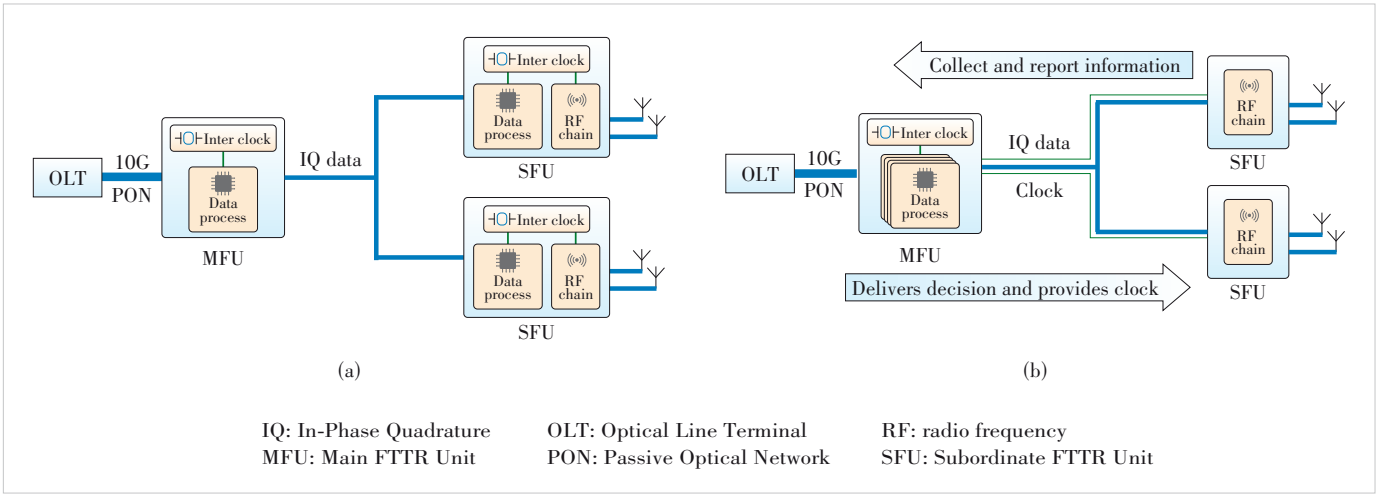


Figure 2. (a) Traditional distributed Fiber-to-the-Room (FTTR) architecture; (b) FTTR/centralized wireless access network (C-WAN) architecture

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:

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

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:

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

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 $\hat{\boldsymbol{H}}(t-L+1),\cdots,\hat{\boldsymbol{H}}(t)$ and the partial antenna observations $\boldsymbol{H}_{N_s}(t+1)$ at time instant $t+1$, the objective is to learn a mapping function $f(\cdot;\boldsymbol{\Theta})$ that accurately estimates the full-dimensional CSI at time $t+1$:

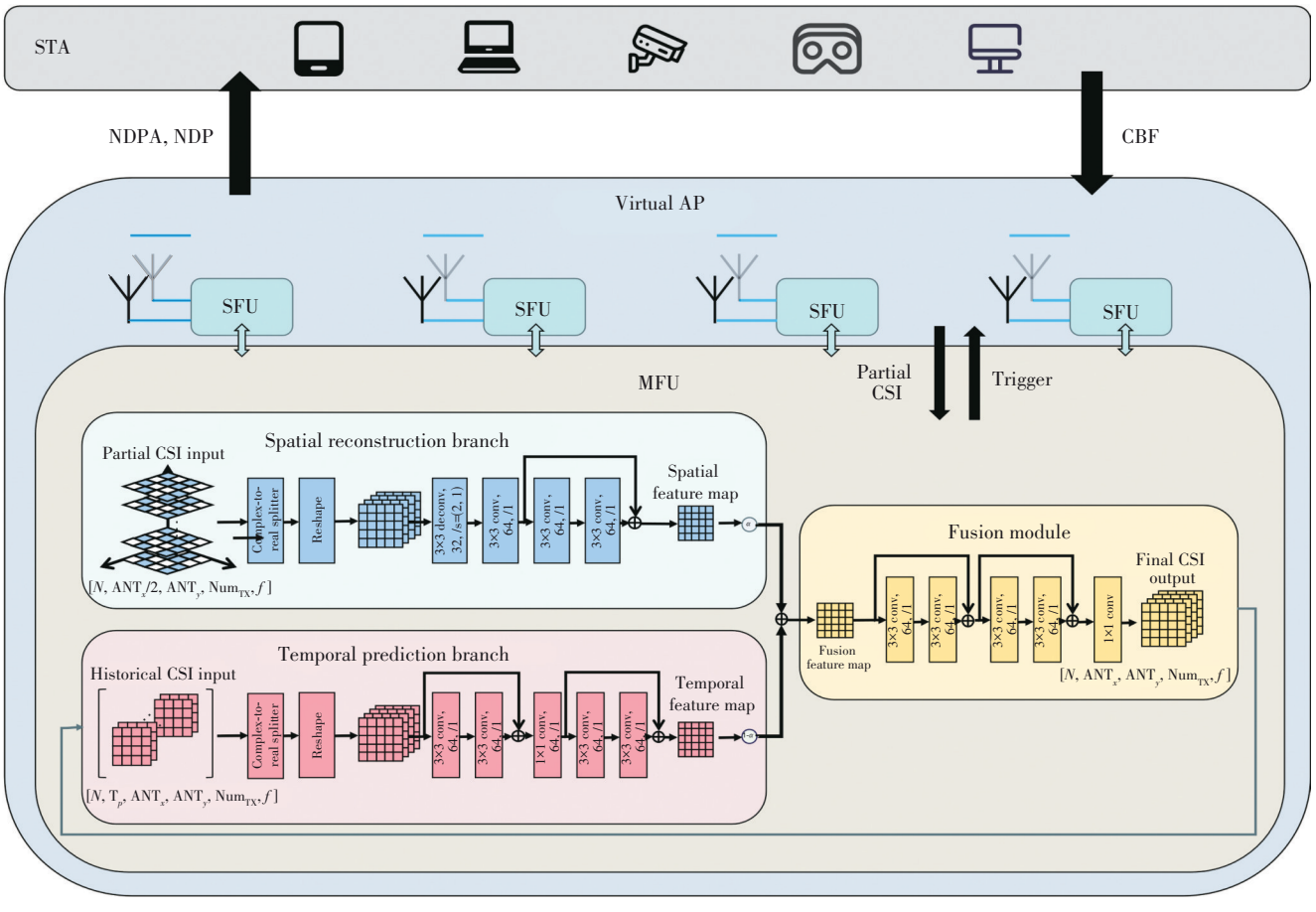
$$\hat{\boldsymbol{H}}(t+1) = f\left(\left\{\hat{\boldsymbol{H}}(t-L+1),\cdots,\hat{\boldsymbol{H}}(t)\right\}, \boldsymbol{H}_{N_s}(t+1);\boldsymbol{\Theta}\right) \tag{7}.$$

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 $\left[N, \mathrm{ANT}_x, \mathrm{ANT}_y, \mathrm{NUM}_{\mathrm{tx}}, f_{\mathrm{subcarrier}}\right]$. After reduction, the matrix contains only partial antenna observations, with dimensions $\left[N, \mathrm{ANT}_x/2, \mathrm{ANT}_y, \mathrm{NUM}_{\mathrm{tx}}, f_{\mathrm{subcarrier}}\right]$. 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; $\mathrm{NUM}_{\mathrm{tx}}$ is the number of transmitters, and $f_{\mathrm{subcarrier}}$ is the number of OFDM subcarriers. After preprocessing steps such as complex number separation and spatial dimension flattening, the



AP: access point
CBF: Compressed Beamforming
CSI: channel state information
MFU: Main FTTR Unit
NDP: Null Data Packet
NDPA: Null Data Packet Announcement
SFU: Subordinate FTTR Unit
STA: Station

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 $[N \cdot T_p, \mathrm{ANT}_x, \mathrm{ANT}_y, \mathrm{NUM}_{\mathrm{tx}}, f_{\mathrm{subcarrier}}]$. 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:

$$\boldsymbol{F}_{\mathrm{reconstructed}} = \alpha \cdot \boldsymbol{F}_{\mathrm{spatial}} + (1 - \alpha) \cdot \boldsymbol{F}_{\mathrm{temporal}} \tag{8},$$

where $\boldsymbol{F}_{\mathrm{spatial}}$ represents the output features of the spatial reconstruction branch, $\boldsymbol{F}_{\mathrm{temporal}}$ 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 $[N, \mathrm{ANT}_x, \mathrm{ANT}_y, \mathrm{NUM}_{\mathrm{tx}}, f]$ 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 $\hat{\boldsymbol{H}}$ and the true CSI $\boldsymbol{H}$. 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 $\boldsymbol{H}$ and the predicted CSI matrix $\hat{\boldsymbol{H}}$, the precoding matrices are calculated using ZF precoding:

$$\boldsymbol{W} = \boldsymbol{H}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1} \tag{9},$$

$$\hat{\boldsymbol{W}} = \hat{\boldsymbol{H}}^{\mathrm{H}}\left(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\right)^{-1} \tag{10}.$$

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

$$P_{\mathrm{sig}} = \left|\left|\boldsymbol{H}\hat{\boldsymbol{W}}\right|\right|_F^2 \tag{11}.$$

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:

$$\mathrm{SINR}_i = \frac{P_{\mathrm{sig},i}}{P_{\mathrm{noise},i}} = \frac{\left|\left|\boldsymbol{H}\hat{\boldsymbol{W}}\right|_{i,j}\right|^2}{\sum_{j \neq i}\left|\left|\boldsymbol{H}\hat{\boldsymbol{W}}\right|_{i,j}\right|^2 + \sigma_n^2} \tag{12},$$

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

$$\mathcal{L} = -\frac{1}{N_{\mathrm{sc}}}\sum_{i=1}^{N_{\mathrm{sc}}}\log_2\left(1 + \mathrm{SINR}_i\right) \tag{13}.$$

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 $\boldsymbol{M} \in C^{N_t \times N_{sc}}$, 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 $\hat{\boldsymbol{H}}$ and the true channel matrix $\boldsymbol{H}$, serving as a key metric to quantify the accuracy of our CSI estimation, with the expression $\text{NMSE} = \frac{\left|\hat{\boldsymbol{H}} - \boldsymbol{H}\right|_F^2}{\left|\boldsymbol{H}\right|_F^2}$.

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 $\boldsymbol{H}_{N_s}(t+1)$ 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 $\boldsymbol{H}_{N_s}(t+1)$ 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

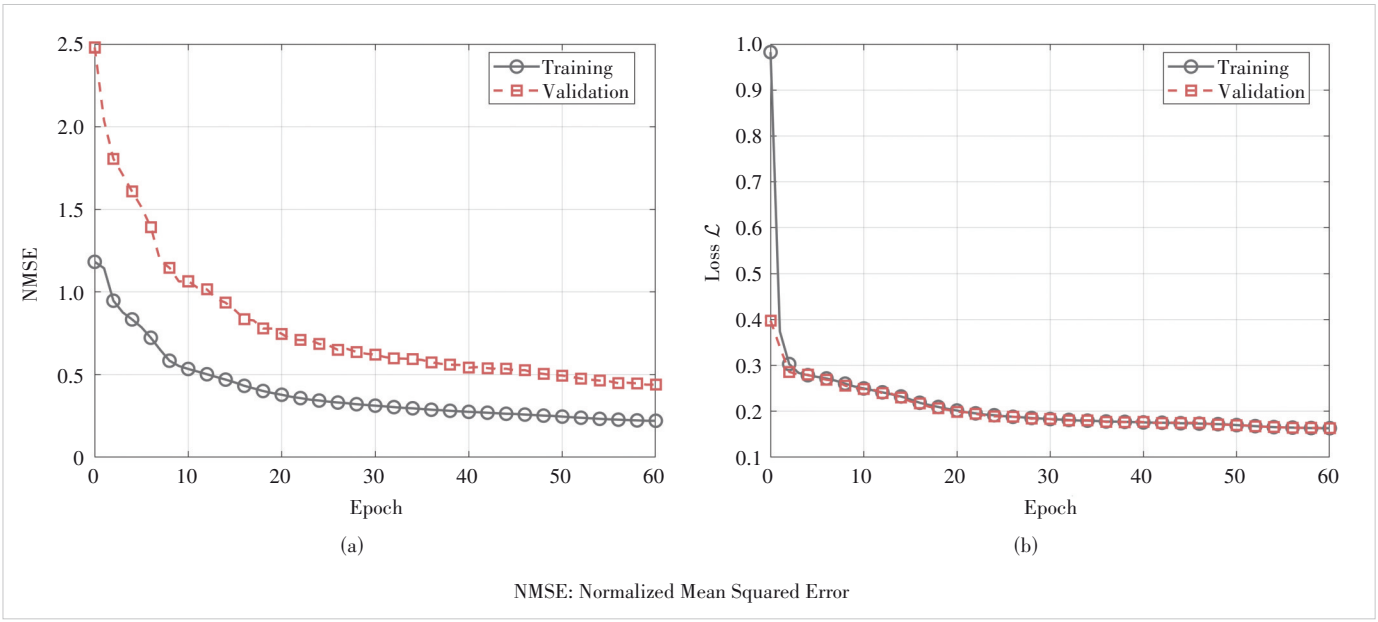


Figure 4. Loss convergence curves: (a) NMSE; (b) the spectral efficiency loss function $\mathcal{L}$

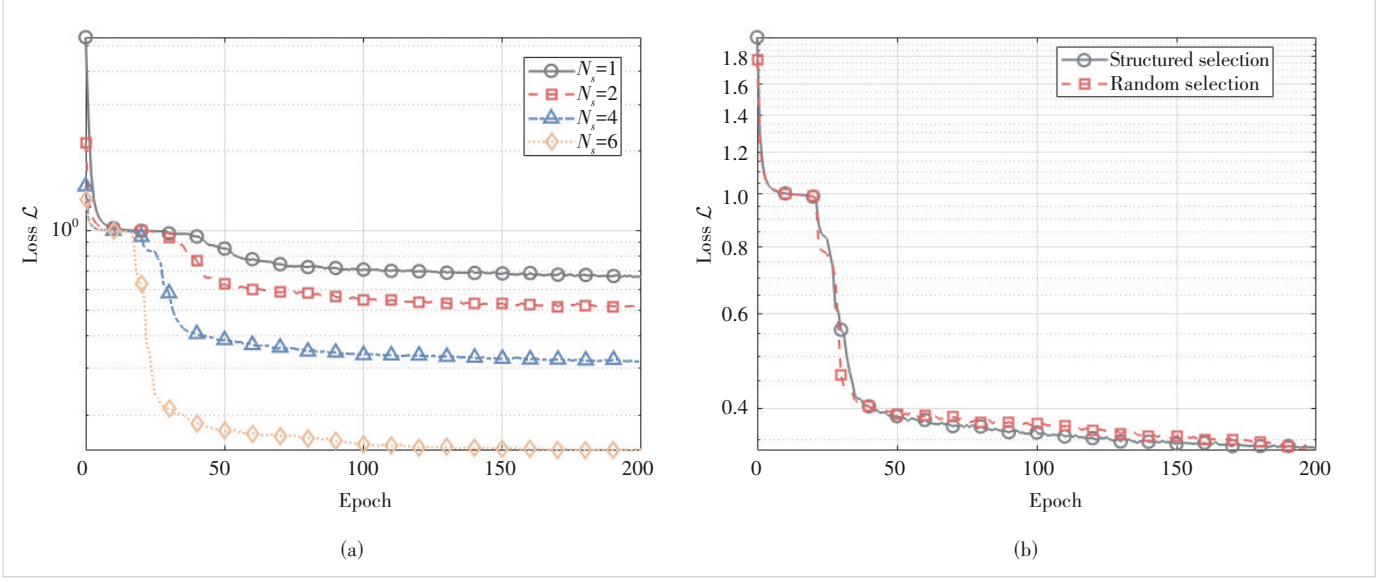


Figure 5. (a) Effect of N_s on the loss function; (b) Effect of N_s selection strategy on the loss function

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-

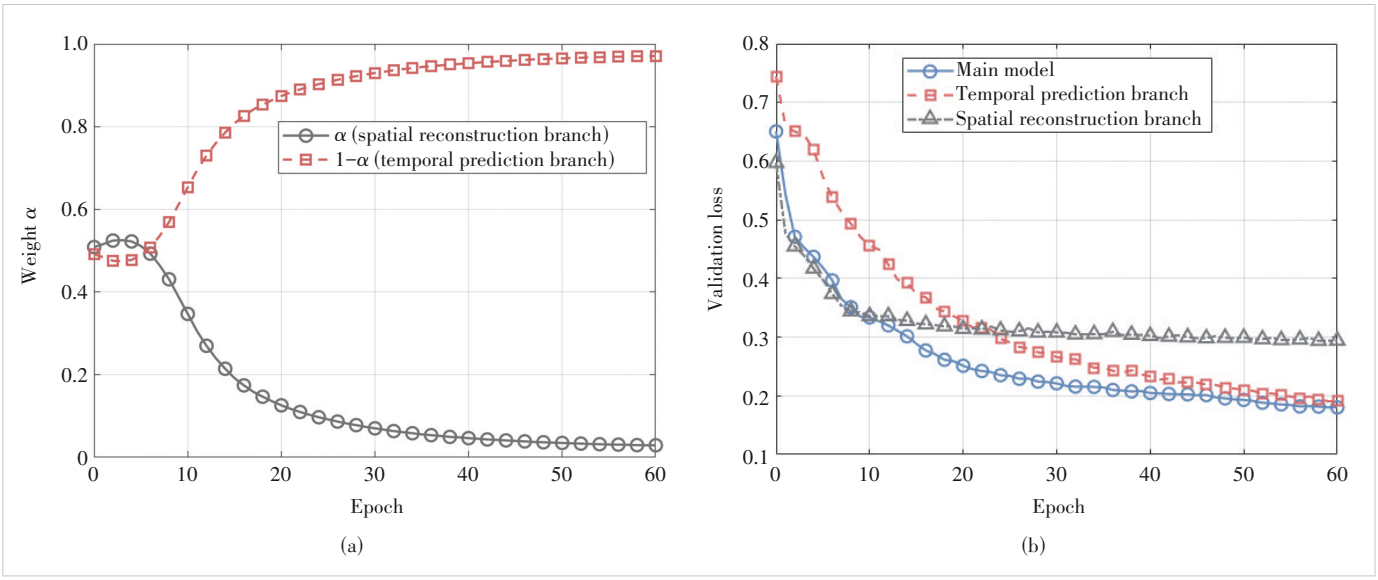


Figure 6. (a) Adaptive fusion weight α for the temporal branch and (b) performance comparison of model variants

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

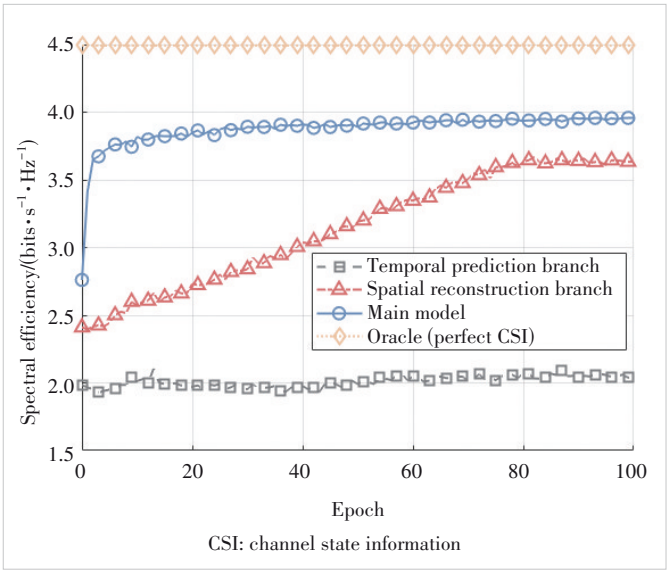


Figure 7. Cumulative Distribution Function (CDF) of achievable spectral efficiency for different CSI acquisition methods

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/ICCWorkshops59551.2024.10615937

[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 WiFi 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. BeamSync: 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 precalibration 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-PPPNet: 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/JIOT.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 interests 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 postdoctoral 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 (WCSP) 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).