Adaptive TTD Configurations for Near-Field Communications: An Unsupervised Transformer Approach

Hsienchih Ting, *Graduate Student Member, IEEE*, Zhaolin Wang, *Member, IEEE*, and Yuanwei Liu, *Fellow, IEEE*

***Abstract*— True-time delayers (TTDs) are popular analog devices for facilitating near-field wideband beamforming subject to the spatial-wideband effect. In this paper, an adaptive TTD configuration is proposed for short-range TTDs. Compared to the existing TTD configurations, the proposed one can effectively combat the spatial-wideband effect for arbitrary user locations and array shapes with the aid of a switch network. A novel end-to-end deep neural network is proposed to optimize the hybrid beamforming with adaptive TTDs for maximizing spectral efficiency. *First*, based on the U-Net architecture, a near-field channel learning module (NFC-LM) is proposed for adaptive beamformer design through extracting the latent channel response features of various users across different frequencies. In the NFC-LM, an improved cross attention (CA) is introduced to further optimize beamformer design by enhancing the latent feature connection between near-field channel and different beamformers. *Second*, a switch multi-user transformer (S-MT) is proposed to adaptively control the connection between TTDs and phase shifters (PSs). In the S-MT, an improved multi-head attention, namely multi-user attention (MSA), is introduced to optimize the switch network by exploring the latent channel relations among various users. *Third*, a multi-feature cross attention (MCA) is introduced to simultaneously optimize the NFC-LM and S-MT by enhancing the latent feature correlation between beamformers and the switch network. Numerical simulation results show that 1) the proposed adaptive TTD configuration effectively eliminates the spatial-wideband effect under uniform linear array (ULA) and uniform circular array (UCA) architectures, and 2) the proposed deep neural network can provide near-optimal spectral efficiency, and solve the multi-user beamformer design and dynamical connection problem in real-time.**

***Index Terms*— Adaptive true-time delayers, hybrid beamforming, near-field, transformer.**

I. Introduction

TO SUPPORT the enormous data requirements, the sixth generation of wireless technology (6G) in the high-frequency band with ultra-broad bandwidth has drawn extensive attention [1]. However, the communication distances in high-frequency bands, such as millimeter-wave (mmWave) and terahertz (THz) bands, will experience severe decreases [2]. Extremely large-scale multiple-input multiple-output (XL-MIMO) technology, with its potential to effectively address path loss and distance challenges, is thereby emerging as a crucial component in the implementation of 6G [3]. Although the integration of XL-MIMO and high-frequency technologies promises vastly improved data speeds and ultra-low latency in 6G, they also bring forth several new challenges. *First*, different from the traditional wireless systems where high-frequency communication takes place in the far-field region, the utilization of XL-MIMO can result in high-frequency communication occurring in the near-field region. As introduced in [3] and [4], the near-field region in 6G can span over hundreds of meters, due to the large aperture of antenna arrays and the extremely high carrier frequencies. Such an extended near-field region in 6G demands a reevaluation of electromagnetic (EM) characteristics, necessitating a shift from conventional modeling methods used in 5G. Particularly, in 6G, the dominance of spherical waves in the near-field region makes the conventional planar-wave model less applicable. *Second*, the increase in both antenna number and bandwidth within high-frequency massive MIMO systems exacerbates the near-field spatial-wideband effect [5]. The resulting frequency-dependent variation in array response poses a significant challenge to the conventional phase shifters (PSs) based hybrid beamforming architectures. Current methods [6], [7], [8] adjust the phase of the signal at each antenna element by generating identical phase shifts across different frequencies, which is inadequate for addressing the near-field spatial-wideband effect in high-frequency massive XL-MIMO systems.

In high-frequency communication, hybrid beamforming integrates analog and digital processing to enhance signal strength towards the intended receiver with reduced hardware complexity [9], overcoming the limitations faced by conventional digital beamforming methods [10]. To address the frequency-dependent spatial-wideband effect in XL-MIMO systems, the implementation of true-time delayers (TTDs) beamforming methods are being explored. TTDs can provide a frequency-dependent phase [11], assisting in the phase alignment of multi subcarriers throughout the bandwidth. Based on the conventional fully-connected (FC) architecture,

Received 19 March 2024; revised 16 August 2024; accepted 19 October 2024. Date of publication 12 November 2024; date of current version 10 January 2025. The associate editor coordinating the review of this article and approving it for publication was G. Zhu. *(Corresponding author: Yuanwei Liu.)*

Hsienchih Ting and Zhaolin Wang are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K. (e-mail: h.ting@qmul.ac.uk; zhaolin.wang@qmul.ac.uk).

Yuanwei Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), E1 4NS London, U.K., and also with the Department of Electronic Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: yuanwei.liu@qmul.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2024.3491419.

Digital Object Identifier 10.1109/TWC.2024.3491419

a straightforward approach involves replacing phase shifters (PSs) with TTDs within the conventional hybrid beamforming framework, as proposed in [11], [12], and [13]. However, adopting this structure in an XL-MIMO system will bring unaffordable costs of hardware complexity. In order to reduce hardware complexity, numerous beamforming architectures incorporating merely a limited number of TTDs have been proposed in recent years [5], [10], [14], [15], [16], [17], [18], which is the main focus of this paper.

A. Prior Works

1) Conventional Algorithm Based Hybrid Beamforming: TTD provides a notable advantage in maintaining consistent beamforming over a broad frequency range. The frequency-dependent characteristic of TTDs can effectively mitigate the spatial-wideband effect. Reference [5] discussed the impact of TTDs in enhancing hybrid beamforming for downlink transmission and integrated TTDs into the PSs-based analog beamformer, offering wideband beamforming capabilities comparable to full-digital arrays. Moreover, the TTD precoding architecture presented in [14] provides a detailed and structured method, allowing adaption to the system of varying bandwidths by tuning the number of TTDs according to the maximum subcarrier and center frequency ratio. With a similar TTD hybrid beamforming structure, [15] utilized a piecewise-far-field wideband channel model to approximate the near-field channel model, addressing the near-field beam split challenge in the XL-MIMO system. In particular,, the entire array is divided into several small subarrays to manage phase discrepancies in near-field channels by decomposing them into inter-array near-field and intra-array far-field discrepancies. These discrepancies are compensated by PSs and TTDs, respectively. However, the above methods ignore the finite time delay and phase delay constraints of TTDs and PSs, as well as the implications on system size, complexity, and cost. As a balanced solution, a switch network is designed to dynamically control the connection between TTDs and PSs, offering a compromise between system simplicity and spectral efficiency. Reference [10] presents the dynamic subarray with fixed true-time-delay (DS-FTTD) architecture, blending low-cost FTTD elements with a dynamic subarray strategy, but at the cost of reduced spectral efficiency and difficulty in effective beamforming. Similarly, [16] fix the connection between TTDs and PSs while dynamically controlling the connection between the RF chains and TTDs. The proposed hybrid bemforming algorithm optimizes the PSs and TTDs for maximum spectral efficiency, disregarding power consumption, then randomly selecting the connection between RF chains and TTDs to meet the power constraint. Different from the dynamic connection strategy, [19] employs baseband TTDs rather than the RF TTDs [14], as in this framework they can compute the optimal results of a fully digital structure [20] and then apply a linear layer to minimize the mean square error (MSE) between the optimal results and their proposed method. Expanding on the cascaded TTDs structure proposed by [21] for angular coverage expansion, [17] tackles the spatial-wideband effect by accumulating time delays. Moreover, [18] explored the advantages of using a serial TTD configuration over a parallel one in the near-field region. To counteract the reduction in independent control associated with a serial setup, a hybrid configuration is introduced for single-user systems, and for broader coverage in multi-user systems, a hybrid-forward-and-backward (HFB) configuration is proposed.

2) Deep Learning Based Hybrid Beamforming: Deep learning (DL) based hybrid beamforming has drawn growing attention. Reference [22] implemented a model-driven DL approach for hybrid beamforming by incorporating the iterative discrete estimation (IDE2) precoder [23] into a neural network through an unfolding approach. Reference [24] proposed a deep neural network (DNN) to learn the SVD process, which is trained by factorizations derived from SVD. Different from above-mentioned methods that employ DL to enhance the performance of traditional algorithms, [25] introduced a convolutional neural network (CNN) for direct mapping from input channel matrices to hybrid beamformers, employing supervised learning with a digital codebook as the label. Moreover, to address the issue of short-range beamformers design in single user MIMO system, [26] treated antenna selection as a classification task, using two CNNs for joint hybrid beamforming and antenna selection design. To address the complexities of multi-user MISO system beamformer design, [27] integrated a double-loop algorithm with a DNN to accelerate antenna selection, mapping beamforming vectors to configurations more efficiently. However, as the complexity of the system increases, it is hard to transform the problem into a convex problem and calculate the optimal results as training labels. Given the limitations inherent to the MSE framework, the suboptimal training labels produced by traditional algorithms pose a significant challenge for DL models to surpass these conventional methods in performance. In [28], [29], and [30], an unsupervised learning approach is employed for MIMO hybrid beamforming, wherein the optimization function and hardware constraints are directly incorporated into the loss function. Furthermore, leveraging a DNN architecture that utilizes the channel matrix as input, the aforementioned DL methods demonstrate the capability to achieve near-optimal outcomes without the necessity for training labels. Moreover, [31] utilized two residual networks [32] which are trained separately by unsupervised learning to address beamformer design and antenna selection, surpassing traditional algorithms in performance.

B. Motivation and Contributions

As mentioned above, TTD-based hybrid beamforming has been widely investigated to combat the spatial-wideband effect in mmWave or THz wireless communication. However, current research mainly focuses on the planar-wave model, an approximation of the spherical-wave, and is often restricted to one specific antenna structure. Moreover, given a finite time delay, the independent use of multiple TTDs is hard to effectively alleviate the spatial-wideband effect. Therefore, it is important to introduce an effective TTD-based hybrid beamforming

structure that is adaptable to the spherical-wave model and arbitrary antenna structures. According to [17], [18], [33], and [34], serial TTD configuration can provide adequate time delay for hybrid beamforming with low hardware complexity. However, a larger array aperture results in a narrower beam, a fixed number of TTDs with predetermined time delays or static connections between TTDs and PSs cannot sustain high beamforming performance. Moreover, the complexity of TTD-based hybrid beamforming system poses computational challenges for conventional algorithms, hindering real-time application capabilities. In addition, supervised learning methods struggle to achieve near-optimal results. Motivated by this, we propose an adaptive TTD configuration hybrid beamforming method for arbitrary antenna structure that cascade multiple TTDs and compensate the time delay adaptively through controlling the connection between TTDs and PSs dynamically. Furthermore, a novel unsupervised network is introduced to optimize the hybrid beamforming with adaptive TTDs for maximizing spectral efficiency. The main contributions of this paper can be summarized as follows:

- We first introduce an serial TTD configuration with adaptive switch network for arbitrary user locations and antenna shapes in near-field region. We then propose a U-Net [35] structure based near-field channel feature learning module (NFC-LM) for beamformer design. In addition, an improved cross attention (CA) [36] is proposed to enhance the correlation between the latent features [37] of near-field channel and corresponding beamformers.
- To realize the adaptive switch network, we further propose a switch multi-user transformer (S-MT) with Hungarian algorithm to adaptively control the connection between the TTDs and PSs. Moreover, in order to effectively enhance the spectral efficiency, we propose an improved multi-head self-attention [38] called multi-user attention (MSA) to model the channel relationships among various users.
- To simultaneously optimize the beamformers design and TTD selection in an end-to-end way, we propose a multi-feature cross-attention (MCA) module further enhancing the latent feature correlations between beamformers and switch network.
- We provide numerical results for both uniform linear array (ULA) and uniform circular array (UCA) architectures to evaluate the performance of our proposed adaptive TTD configuration hybrid beamforming method. The results demonstrate that our method can effectively combat the spatial-widebandd effect for both ULA and UCA, regardless of user locations. In addition, the proposed unsupervised deep neural network can provide near optimal spectral efficiency.

C. Organization and Notations

The rest of the paper is organized as follows: Section II intoduces the strucutre of adaptive TTD configuration and formulate the spherical-wave model hybrid beamforming problem. Section III introduces the proposed DL based algorithm for beamformer design and dynamic PSs selection. Section IV presents the numerical results of different array structures and ablation studies of our proposed method. Section V concludes this paper.

Notations: We use lower-case, bold-face lower-case and bold-face uppercase letters to represent scalars, vectors and matrices, respectively. The transpose, conjugate transpose of a matrix are denoted by $(\cdot)^T$ and $(\cdot)^H$, respectively. The hadamard product is denoted by $\odot$. The convolution operate is denoted by $*$. The Euclidean norm of vector $\mathbf{x}$ is denoted as $\|\mathbf{x}\|$, while the Frobenius norm of matrix $\mathbf{X}$ is denoted as $\|\mathbf{X}\|_F$. A block diagonal matrix with diagonal blocks $\mathbf{x}_1, \ldots, \mathbf{x}_N$ is denoted as blkdiag $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. $\mathcal{CN}(\mu, \sigma^2)$ denotes the circularly symmetric complex Gaussian random distribution with mean μ and variance σ^2. $\mathcal{I}(M)$ denotes the set $\{1, 2, \ldots, M\}$.

II. System Model and Problem Formulation

In this paper, we study a near-field wideband XL-MIMO communication system. A base station (BS) equipped with N antennas serves K single-antenna users situated in the near-field region. The system operates in the mmWave or THz bands, adopting orthogonal frequency division multiplexing (OFDM) with M subcarriers to address the inter-symbol interference caused by the frequency-wideband effect. The bandwidth and central frequency of system are denoted by B and f_c, respectively, therefore each OFDM subcarrier having a frequency $f_m = f_c + \frac{B(2m-1-M)}{2M}$. To facilitate the near-field wideband beamforming with short-range TTDs for arbitrary user locations and array shapes, we propose an adaptive TTD-based hybrid beamforming architecture, depicted in Fig. 1. Our proposed architecture introduces an additional switch network which is positioned between the PS network and the TTD network. The adaptive compensation of time delays among different users is achieved through the dynamic interconnection of TTDs and PSs.

A. Signal Model

Let N_{RF} denote the number of RF chains employed in the adaptive TTD-based beamforming architecture. For the m-th subcarrier, let $\mathbf{A}_m \in \mathbb{C}^{N \times N_{\mathrm{RF}}}$ denotes the frequency-dependent analog beamformer jointly achieved by PSs and TTDs, and $\mathbf{D}_m \in \mathbb{C}^{N_{\mathrm{RF}} \times K}$ denote the baseband digital beamformer for K users. Then, the transmit signal can be expressed as

$$\mathbf{x}_m = \mathbf{A}_m \mathbf{D}_m \mathbf{c}_m = \sum_{k=1}^{K} \mathbf{A}_m \mathbf{d}_{m,k} c_{m,k}, \tag{1}$$

where $\mathbf{d}_{m,k} \in \mathbb{C}^{N_{\mathrm{RF}} \times 1}$ is the vector at the k-th column of matrix $\mathbf{D}_m$ and also denotes the baseband digital beamformer for the k-th user. Vector $\mathbf{c}_m = [c_{m,1}, \ldots, c_{m,K}]^T \in \mathbb{C}^{K \times 1}$ denote the unit-power information symbols delivered to the K users, which are modelly as independent and identically distributed (i.i.d.) random variables, i.e., $\mathbb{E}[\mathbf{c}_m \mathbf{c}_m^H] = \mathbf{I}_K$.

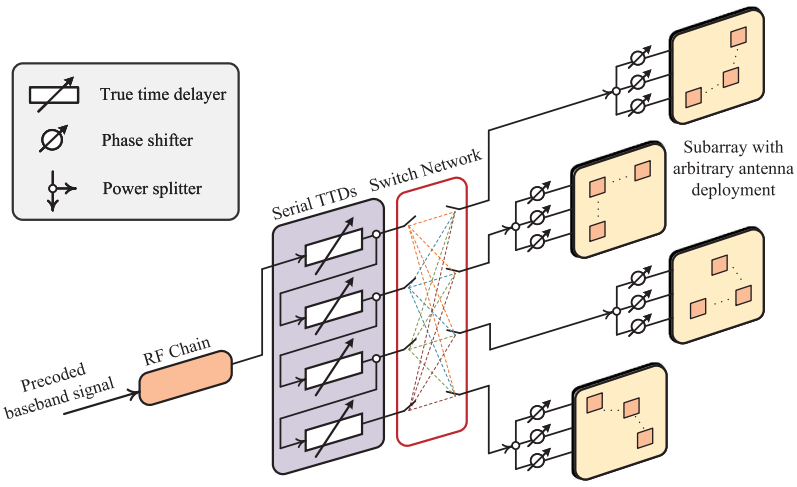


Fig. 1. Proposed adaptive-serial configuration for TTD-based hybrid beamforming.

Accordingly, the signal received at user k on the m-th subcarrier can be modelled as

$$y_{m,k} = \mathbf{h}_{m,k}^H \mathbf{A}_m \mathbf{d}_{m,k} c_{m,k} + \sum_{i \neq k} \mathbf{h}_{m,k}^H \mathbf{A}_m \mathbf{d}_{m,i} c_{m,i} + z_{m,k}, \tag{2}$$

where $\mathbf{h}_{m,k} \in \mathbb{C}^{N \times 1}$ denotes the baseband communication channel for user k on the m-th subcarrier and $z_{m,k} \sim \mathcal{CN}(0, \sigma_{m,k}^2)$ denotes the additive complex Gaussian noise with power of $\sigma_{m,k}^2$.

In this paper, the spherical-wave-based near-field channel model is considered. The adopted model accounts for the effects of both direct line-of-sight (LOS) and L_k indirect non-line-of-sight (NLOS) paths for user k, which arise due to scattering. Hence, the channel $\mathbf{h}_{m,k}$ can be expressed as [39]

$$\mathbf{h}_{m,k} = \beta_{m,k} \mathbf{b}^*(f_m, r_k, \theta_k) + \sum_{l=1}^{L_k} \tilde{\beta}_{m,k,l} \mathbf{b}^*(f_m, \tilde{r}_{k,l}, \tilde{\theta}_{k,l}), \tag{3}$$

where $\beta_{m,k}$ and $\tilde{\beta}_{m,k}$ denote the complex channel gain for LOS and NLOS paths, respectively, r_k and θ_k denote the distance and direction of user k with respect to the BS, $\tilde{r}_{k,l}$ and $\tilde{\theta}_{k,l}$ denote the distance and direction of the l-th resolvable scatter in user k's NLOS path with respect to the BS, and vector $\mathbf{b}^*(f, r, \theta) \in \mathbb{C}^{N \times 1}$ denotes the array response at distance r and direction θ. Due to the spherical wave propagation in the near-field region, $\mathbf{b}^*(f, r, \theta)$ should be modelled accurately as

$$\mathbf{b}(f, r, \theta) = e^{-j\frac{2\pi f}{c}\mathbf{r}(r,\theta)}, \tag{4}$$

where $\mathbf{r}(r, \theta) = [r_1, \ldots, r_N]^T$ denotes the vector of distances from each antenna to a given point, where r_n specifies the distance from the n-th antenna and the location point (r, θ). There are two key observations from the expression of array response. On the one hand, the array response vector is frequency-dependent, which thus requires exploiting TTDs to facilitate frequency-dependent analog beamforming. On the other hand, the distance vector $\mathbf{r}(r, \theta)$ not only depends on r and θ, but also depends on the location of the antenna elements. In other words, different shapes of antenna arrays can lead to different channel characteristics. Based on these observations, in the following, we propose an adaptive TTD configuration to compensate for the spatial wideband effect regarding different array shapes and short-range TTDs.

B. Proposed Adaptive-Serial TTD Configuration

In this subsection, we introduce the detailed configuration of the analog beamforming matrix, $\mathbf{A}_m$, within the structure of the proposed adaptive-serial TTDs beamforming method. As shown in Fig. 1, we assume each RF chain connects to all antennas via L TTDs and N PSs.[1] Furthermore, each TTD is connected to a sub-array with size $Q = N/L$. Let $t_{l,i}$ and $\phi_{n,i}$ denote the time delay at the output of the l-th TTD and the phase adjustment of the n-th PS connected to the i-th RF chain. Given that PSs only allow constant-modulus, the PSs are subject to the following constraint

$$|\phi_{n,i}| = 1. \tag{5}$$

To address the short-range limitation, the TTDs are connected in an serial manner [18]. More particularly, the output time delay $t_{l,i}$ of l-th TTD is an accumulative result containing all time delays provided by previous TTDs. Hence, $t_{l,i}$ can be expressed as

$$t_{l,i} = \sum_{j=1}^{l} \tilde{t}_{j,i}, \tag{6}$$

where $\tilde{t}_{j,i}$ denotes the time delay realized by the j-th TTD and is subject to a maximum delay constraint, i.e., $\tilde{t}_{j,i} \in [0, t_{\max}]$. For each subcarrier indexed by m, the frequency domain phase shift realized by the cumulative time delay $t_{l,i}$ is given by $e^{-j2\pi f_m t_{l,i}}$. We further assume that power is distributed equitably across all TTDs, with adjustments to the power

[1]In this work, we focus on two-dimensional (2D) beamforming within a 2D coordinate system. For three-dimensional (3D) beamforming, such as that achieved by uniform planar arrays (UPAs), a two-tier TTD configuration should be utilized [40], which is left for our future work.

being made through the coefficients of the power divider. Additionally, the dynamic connection between the l-th TTD and the n-th PS is implemented by the switch network through the adjustment of the binary switch coefficient $\tilde{s}_{n,l,i} \in \{0,1\}$. Consequently, with this architecture, the overall analog adaptive TTD-based hybrid beamforming matrix $\mathbf{A}_m \in \mathbb{C}^{N\times N_{\mathrm{RF}}}$ can be expressed as

$$\mathbf{A}_m = \begin{bmatrix} a_{m,1,1} & \cdots & a_{m,1,N_{\mathrm{RF}}} \\ \vdots & \ddots & \vdots \\ a_{m,N,1} & \cdots & a_{m,N,N_{\mathrm{RF}}} \end{bmatrix}, \tag{7}$$

where

$$a_{m,n,i} = \phi_{n,i} \sum_{l=1}^{L} \tilde{s}_{n,l,i} e^{-j2\pi f_m t_{l,i}}. \tag{8}$$

To make the adaptive-serial TTD configuration viable in practical systems, it is important to ensure that each PS is connected exclusively to one corresponding TTD. This constraint can be expressed as

$$\sum_{l=1}^{L} \tilde{s}_{n,l,i} = 1. \tag{9}$$

Furthermore, since each TTD in a serial configuration can support a larger time delay compared to a parallel configuration [18], we distribute the N antennas into L groups. Each TTD then exclusively connects to one sub-array with size $Q = N/L$. In this case, we have:

$$\sum_{n=1}^{N} \tilde{s}_{n,l,i} = Q. \tag{10}$$

We note that based on the above configuration, if we consider N antennas with different PSs should partitioned into L different TTDs and each TTD at least has one connection, then the total number of combinations can be expressed by the Stirling number of the second kind [41]

$$L! \times S(N,L) = L! \times \frac{1}{L!} \sum_{i=0}^{L} (-1)^{L-i} \binom{L}{i} (i)^N. \tag{11}$$

This number is substantial even for a modest number of TTDs. For instance, with a single RF chain, the count reaches 3.4032×10^{38} for a configuration of 64 antennas and 4 TTDs. Moreover, if we assume that each TTD is connected to non-ordered, equal-sized subsets of antennas, with each subset comprising $Q = N/L$ antennas, as written in (10) then the total number of combinations, calculated as $\frac{N!}{(N/L!)^L L!}$, remains significantly large. For instance, in a configuration with 64 antennas and 4 TTDs, this leads to approximately 2.67×10^{34} combinations. To find the near optimal solution, we methodically preallocate the antennas into L organized groups, with each group containing Q antennas, as shown in Fig. 1. In this case, the overall analog beamforming matrix $\mathbf{A}_m$ reduces to

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{a}_{m,1,1} & \cdots & \mathbf{a}_{m,1,N_{\mathrm{RF}}} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{m,L,1} & \cdots & \mathbf{a}_{m,L,N_{\mathrm{RF}}} \end{bmatrix}. \tag{12}$$

Here, $\mathbf{a}_{m,p,i} \in \mathbb{C}^{Q\times 1}$ denotes the analog beamformer for the p-th ordered, equal-sized subarray and the i-th RF chain. It can be written specifically as

$$\mathbf{a}_{m,p,i} = \boldsymbol{\phi}_{p,i} \sum_{l=1}^{L} s_{p,l,i} e^{-j2\pi f_m t_{l,i}}, \tag{13}$$

where $\boldsymbol{\phi}_{p,i} = [\phi_{(p-1)Q+1,i}, \ldots, \phi_{pQ,i}]^T$ denotes the PS-based analog beamformer for the p-th ordered, equal-sized subarray and the i-th RF chain, and $s_{p,l,i} \in \{0,1\}$ is the new binary switch coefficient. In particular, $s_{p,l,i} = 1$ implies that the l-th TTD is connected to the p-th sub-array with the i-th RF chain. Furthermore, we assume that each TTD is exclusively linked to a single sub-array, and vice vesa, each sub-array is uniquely connected to one TTD, resulting in the following constraint:

$$\sum_{l=1}^{L} s_{p,l,i} = 1, \quad \sum_{p=1}^{L} s_{p,l,i} = 1. \tag{14}$$

C. Problem Formulation

In this work, we aim to maximize spectral efficiency by jointly optimizing the analog and digital beamformers. According to (2), the achievable rate for user k at the m-th subcarrier can be calculated based on the Shannon formula as follows:

$$R_{m,k} = \log_2 \left(1 + \frac{|\mathbf{h}_{m,k}^H \mathbf{A}_m \mathbf{d}_{m,k}|^2}{\sum_{i=1,i\neq k}^{K} |\mathbf{h}_{m,k}^H \mathbf{A}_m \mathbf{d}_{m,i}|^2 + \sigma_{m,k}^2} \right). \tag{15}$$

The spectral efficiency of the considered multi-user OFDM system is thus given by

$$R = \frac{1}{M + L_{\mathrm{CP}}} \sum_{m=1}^{M} \sum_{k=1}^{K} R_{m,k}. \tag{16}$$

The spectral efficiency maximization problem is thus given by

$$\max_{\boldsymbol{\Phi},\mathbf{S},\mathbf{T},\mathbf{D}_m} \quad \sum_{m=1}^{M} \sum_{k=1}^{K} R_{m,k} \tag{17a}$$

$$\text{s.t.} \quad \|\mathbf{A}_m \mathbf{D}_m\|_F^2 \leq P_t, \forall m, \tag{17b}$$

$$\mathbf{a}_{m,p,i} = \boldsymbol{\phi}_{p,i} \sum_{l=1}^{L} s_{p,l,i} e^{-j2\pi f_m t_{l,i}}, \forall m,p,i, \tag{17c}$$

$$\sum_{l=1}^{L} s_{p,l,i} = 1, \quad \sum_{p=1}^{L} s_{p,l,i} = 1, \forall i, \tag{17d}$$

$$t_{l,i} = \sum_{j=1}^{l} \tilde{t}_{j,i}, \quad \tilde{t}_{j,i} \in [0, t_{\max}], \forall l,i, \tag{17e}$$

$$|\phi_{n,i}| = 1, \forall n,i, \tag{17f}$$

where P_t denotes the maximum transmit power for each subcarrier. Matrices $\boldsymbol{\Phi} \in \mathbb{C}^{L\times N_{\mathrm{RF}}}$, $\mathbf{S} \in \mathbb{C}^{L\times L\times N_{\mathrm{RF}}}$, and $\mathbf{T} \in \mathbb{C}^{L\times N_{\mathrm{RF}}}$ represent the coefficients of PSs, switches, and TTDs, respectively. Their entries are given by

$$[\boldsymbol{\Phi}]_{p,i} = \phi_{p,i}, \quad [\mathbf{S}]_{p,l,i} = s_{p,l,i}, \quad [\mathbf{T}]_{l,i} = t_{l,i}. \tag{18}$$

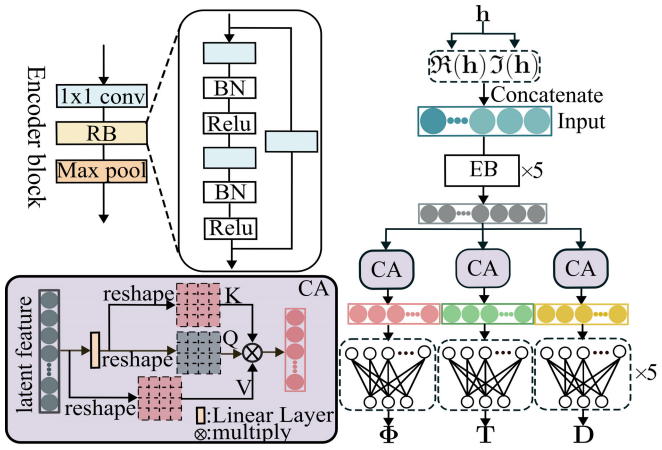


Fig. 2. The network structure of channel feature learning module.

However, jointly optimizing $\mathbf{\Phi}$, $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{D}_m$ is intractable, as $\mathbf{S}$ directly influences the structure of $R_{m,k}$ in (17a). The aim of this paper is to develop a solution for problem (17a) that is not only low in complexity but also feasible for real-time implementation. To enhance the clarity of this paper, we introduce an end-to-end DL approach and a more detailed discussion will be provided in Section III.

III. DL-Based Adaptive TTD Hybrid Beamforming

To address the aforementioned problems, in this section, we will introduce our proposed DL-Based Adaptive TTD Hybrid Beamforming algorithm including the network structure, training procedure, and the loss function design. We break down the optimization problem into two phases. In the first phase, we design a near-field channel learning module (NFC-LM) to construct the hybrid beamforming matrices $\mathbf{\Phi}$, $\mathbf{T}$, and $\mathbf{D}_m$. In the second phase, we introduce the S-MT module to manage the connection matrix $\mathbf{S}$. These two networks are then trained in an unsupervised end-to-end way.

A. Near-Field Channel Latent Feature Learning

Unlike most existing NN-based hybrid beamforming algorithms [19], [28], [29], [30], which typically employ several linear layers to directly construct relationships among channel response, analog beamformers and digital beamformers for specific antenna shape, our approach introduces a novel unsupervised Encoder-Decoder structure for arbitrary antenna shapes. This structure aims to explore latent feature relations among $\mathbf{h}_{m,k}$, $\mathbf{\Phi}$, $\mathbf{T}$ and $\mathbf{D}_m$. Based on the U-Net structure [35], our near-field channel learning module (NFC-LM) is divided into two symmetrical parts: a convolutional encoder and multi-linear decoders, as illustrated in Fig. 2. Each convolution of convolutional encoder can be expressed as

$$\mathrm{Conv}_{i,j}(\mathbf{X}) = \mathbf{W}_{i,j} * \mathbf{X} + \mathbf{B}_{i,j}, \tag{19}$$

where $\mathbf{X}$ represents the input tensor, $\mathbf{W}_{i,j}$ and $\mathbf{B}_{i,j}$ represents the weights and biases of the j-th 1D convolution of i-th convolutional layer, respectively. Each linear layer of multi-linear decoders can be expressed as

$$\mathrm{Linear}_{i,j}(\mathbf{X}) = \mathbf{W}_{i,j}\mathbf{X} + \mathbf{B}_{i,j}, \tag{20}$$

where $\mathbf{X}$ represents the input tensor, $\mathbf{W}_{i,j}$ and $\mathbf{B}_{i,j}$ represents the weights and biases of j-th linear projection of i-th linear decoder layer, respectively.

To enhance latent feature representation learning in NFC-LM, we initially separate complex channel response matrices into their real and imaginary components. These components are treated as multi-modal features, which are then transformed into tensors. Thus, the channel response $\mathbf{h}_{m,k}$ can be further expressed as follow:

$$\Re(\mathbf{h}_{m,k}) = \sum_{l=1}^{L_k} \left(\Re(\beta_{m,k,l})\Re(\mathbf{b}^*_{m,k,l}) - \Im(\beta_{m,k,l})\Im(\mathbf{b}^*_{m,k,l})\right), \tag{21}$$

$$\Im(\mathbf{h}_{m,k}) = \sum_{l=1}^{L_k} \left(\Re(\beta_{m,k,l})\Im(\mathbf{b}^*_{m,k,l}) + \Im(\beta_{m,k,l})\Re(\mathbf{b}^*_{m,k,l})\right), \tag{22}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts, respectively. Subsequently, the real and imaginary components of the channel responses across different users and frequencies are combined and transformed into a tensor representation. The process can be expressed as follows:

$$\mathbf{H} = \mathrm{Tens}(\mathrm{con}(\Re(\mathbf{h}_{1,1}), \Im(\mathbf{h}_{1,1}), \ldots, \Re(\mathbf{h}_{M,K}), \Im(\mathbf{h}_{M,K}))), \tag{23}$$

where $\mathbf{H} \in \mathbb{R}^{K\times M\times 2N}$ represents the combined responses for all users K and frequencies M, expanding each antenna response into $2N$ elements, thereby accommodating both its real and imaginary components. Here $\mathrm{Tens}(\cdot)$ represents the tensorization operation flattening a matrix into a tensor by column order and $\mathrm{con}(\cdot)$ denotes the concatenation operation. Additionally, to directly construct the latent feature correlations among channel response, PSs, TTDs and baseband digital beamformer, we also predict the real and imaginary parts of $\mathbf{\Phi} \in \mathbb{C}^{L\times N_{RF}}$, $\mathbf{T} \in \mathbb{C}^{L\times N_{RF}}$ and $\mathbf{D} \in \mathbb{C}^{K\times M\times N_{RF}}$. Here $\mathbf{D}$ denotes the concatenation of $\mathbf{D}_m$ across all subcarriers and all RF chains. By denoting the input channel response tensor as $\mathbf{H}$ and the output as $\mathbf{\Phi}$, $\mathbf{T}$ and $\mathbf{B}$, the end-to-end relationship of the NFC-LM can be expressed as

$$\left[\Re(\mathbf{\Phi}), \Im(\mathbf{\Phi}), \Re(\mathbf{T}), \Im(\mathbf{T}), \Re(\mathbf{D}), \Im(\mathbf{D})\right] = \mathrm{NFC-LM}(\mathbf{H}), \tag{24}$$

where $\mathrm{NFC-LM}(\cdot)$ represents the mapping function of NFC-LM.

As showed in Fig. 2, the feature extraction part of our method consists of five encoder blocks denoted as $\mathrm{EB}(\cdot)$, which are integrated with batch normalization ($\mathrm{BN}(\cdot)$) and $\mathrm{ReLU}(\cdot)$ activation. Each encoder block includes a 1D convolution layer with a kernel size of 3 and a stride of 1, followed by a residual block denoted as $\mathrm{RB}(\cdot)$, before the max pooling operation. Considering the input channel response tensor, $\mathbf{H}$, the feature extraction process for the output latent channel response feature, $\mathbf{L_H}$, from the encoder can be written as follows:

$$\begin{aligned}
\mathrm{RB}_i(\mathbf{X}_i) &= \mathrm{Conv}_{i,3}(\mathrm{ReLU}((\mathrm{Conv}_{i,2}(\mathbf{X}_i)))) + \mathbf{X}_i,\\
\mathrm{EB}_i(\mathbf{X}_i) &= \mathrm{Conv}_{i,4}(\mathrm{RB}_i(\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{i,1}(\mathbf{X}_i))))),\\
\mathbf{L}_{\mathbf{H},i} &= \mathrm{EB}_i(\mathbf{L}_{\mathbf{H},i-1}),
\end{aligned} \tag{25}$$

where i indexes the i-th encoder block with being 5 in this paper, $\mathbf{X}_i$ represents the input tensor to the i-th encoder block, the initial input $\mathbf{L}_{\mathbf{H},0}$ is set to $\mathbf{H}$ and $\mathbf{L}_{\mathbf{H},i} \in \mathbb{R}^{(K\cdot 2^{2+i})\times(M/2^i)\times(2N/2^i)}$. We also introduce an improved cross-attention (CA) [36] module in NFC-LM to further enhance the latent feature representation. The extracted latent channel response feature $\mathbf{L}_{\mathbf{H},5}$ is first flattened into $\mathbb{R}^{K\cdot 2^{2+5}\times(M/2^5)(2N/2^5)}$, where $K\cdot 2^{2+5}$ represents the channel dimension and $(M/2^5)(2N/2^5)$ reprensents the data length. In the CA module, the process begins with employing a linear layer to map $\mathbf{L}_{\mathbf{H},5}$ onto the latent features associated with the various beamformers, namely $\mathbf{L}_{\mathbf{\Phi}}$, $\mathbf{L}_{\mathbf{T}}$ and $\mathbf{L}_{\mathbf{D}}$. After that, three distinct linear layers are applied to generate the query ($\mathbf{Q}$) from latent channel response feature and key ($\mathbf{K}$) and value ($\mathbf{V}$) from corresponding latent beamformer feature. Subsequently, the cross-attention weights are calculated by applying the softmax function, denoted by $\mathrm{softmax}(\cdot)$, to the scaled dot-product of $\mathbf{Q}$ and $\mathbf{K}$. This calculation aims to model the interrelations between the latent channel response feature and corresponding latent beamformer feature across the data dimension. Finally, the cross-attention weights are multiplied with $\mathbf{V}$ to further enhance the connection between channel response and corresponding beamformer. The incorporation of the CA module enable beamformers to tackle arbitrary user locations and antenna shapes through establishing data dimension connections between latent channel response feature and latent beamformer features. To clarify, the operation of the Cross-Attention (CA) module, denoted by $\mathrm{CA}(\cdot)$, can be expressed as follows:

$$\mathbf{Y} = \mathrm{CA}\left(\mathbf{X}\right) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{26}$$

where

$$\mathbf{Q} = \mathrm{Linear}_{0,2}(\mathbf{X}), \tag{27a}$$
$$\mathbf{K} = \mathrm{Linear}_{0,3}(\mathbf{X}), \tag{27b}$$
$$\mathbf{V} = \mathrm{Linear}_{0,4}(\mathbf{X}). \tag{27c}$$

Moreover, $\mathbf{X} = \mathbf{L}_{\mathbf{H}_5}$ represents the input latent feature of CA module, $\mathbf{Y} \in \{\mathbf{L}_{\mathbf{\Phi}}, \mathbf{L}_{\mathbf{T}}, \mathbf{L}_{\mathbf{D}}\}$ represents the resulting latent beamformer feature from CA module and d_k represents the dimensionality of key, which is used for scaling the dot products. Considering the input $\mathbf{L}_{\mathbf{H},5}$ and the output $\mathbf{L}_{\mathbf{\Phi}}$, $\mathbf{L}_{\mathbf{T}}$ and $\mathbf{L}_{\mathbf{D}}$ of CA module can be expressed as follows

$$\mathbf{L}_{\mathbf{\Phi}} = \mathrm{CA}(\mathbf{L}_{\mathbf{H},5}), \mathbf{L}_{\mathbf{T}} = \mathrm{CA}(\mathbf{L}_{\mathbf{H},5}), \mathbf{L}_{\mathbf{D}} = \mathrm{CA}(\mathbf{L}_{\mathbf{H},5}), \tag{28a}$$

where each invocation of the $\mathrm{CA}(\cdot)$ applies different linear projections tailored to the specific beamformer features it generates. Consequently, the outputs of the CA module are then fed into the multi-linear decoders, denoted as $\mathrm{MLD}(\cdot)$. These linear decoders mirror the convolutional encoder structure, each comprising five linear layers, and the decoding process can be described as follow

$$\begin{aligned}\mathbf{Y} &= \mathrm{MLD}(\mathbf{X}) \\ &= \mathrm{Linear}_{1,1}(\mathrm{Linear}_{2,1}(\cdots \mathrm{Linear}_{5,1}(\mathbf{X}))),\end{aligned} \tag{29}$$

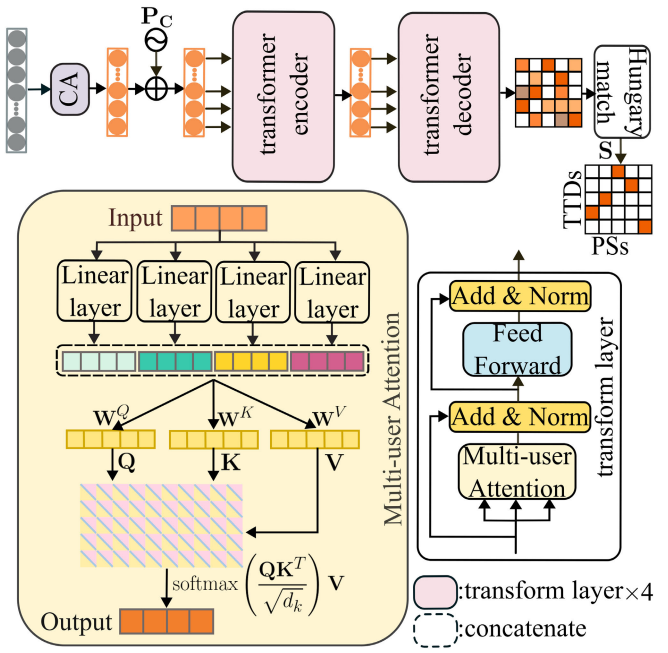


Fig. 3. The network structure of multi-user transformer for adaptive connection between TTDs and PSs.

where $\mathbf{X} \in \{\mathbf{L}_{\mathbf{\Phi}}, \mathbf{L}_{\mathbf{T}}, \mathbf{L}_{\mathbf{D}}\}$ is the input of multi-linear decoders and $\mathbf{Y} = [\Re(\mathbf{\Phi}), \Im(\mathbf{\Phi}), \Re(\mathbf{T}), \Im(\mathbf{T}), \Re(\mathbf{D}), \Im(\mathbf{D})]$ is the output of multi-linear decoders.

B. Switch Multi-User Transformer

Our research focuses on near-field hybrid beamforming, particularly addressing the challenge posed by the diverse range of antenna shapes encountered in practical applications. This diversity highlights the necessity for an adaptive TTD configuration approach to ensure effective beamforming across various scenarios. Motivated by this, we introduce the S-MT, which is designed for beamforming of arbitrary antenna shapes. Specifically, as illustrated in Fig. 1, the S-MT enable the switch network to dynamically control the compensation for time delays through effectively managing the connection between TTDs and PSs. Here, we preallocate the antennas into equal-sized subarrays and construct the correlations between TTD and subarray instead of individual antennas. Thereby, we can reduce the complexity of extremely large antenna modeling and tackle beamforming for arbitrary antenna shapes. As illustrated in Fig. 3, the S-MT comprises a positional encoding, an multi-head attention mechanism, an encoder-decoder transformer, and the Hungarian matching algorithm, which collectively determine the final connection configuration.

To establish the latent feature interrelation between switch network and channel response, we first employ the CA module to generate the latent switch feature, denoted as $\mathbf{L}_{\mathbf{S}} \in \mathbb{R}^{N_{RF}L\times N_{RF}L}$. In the real-time beamforming process, accurately capturing the intrarealtions among various subarrays is essential. Additionally, to leverage the sequence order, we assign a unique positional encoding, denoted as $\mathbf{P}_{\mathbf{C}}$, to each PS group to differentiate and model these subarrays

effectively. After generating the latent switch feature, represented as $\mathbf{L_S}$, we apply a specific positional encoding function, denoted as $\mathrm{E}_{PC}(\cdot)$. Each PS cluster is uniquely characterized through a sinusoidal coding approach [38], which can be expressed as follows:

$$\mathrm{E}_{PC}(p, 2i) = \sin(p/1000^{2i/d}), \tag{30a}$$

$$\mathrm{E}_{PC}(p, 2i+1) = \cos(p/1000^{2i/d}), \tag{30b}$$

$$\mathbf{P_C} = \mathrm{E}_{PC}(p), \forall p \in \mathcal{I}(N_{RF}L) \tag{30c}$$

where $\mathbf{P_C} \in \mathbb{R}^{N_{RF}L \times N_{RF}L}$ represents the unique positional encoding for each PS cluster generated by $\mathbf{E}_{PC}(\cdot)$, and $p = \{1, 2, \ldots, N_{RF}L\}$ represents the index of preallocated subarray, for example $p = 1$ indicating the first PS cluster, and so on. Here, i denotes the coding dimension, and $d = L = 128$ represents the input feature dimension. The position coding is used to embed the spatial information of preallocated PS clusters and model the relationships among different TTD-PS connection scenarios to maximize spectral efficiency. Given that the dimensions of $\mathbf{P_C}$ align with those of the input latent switch feature $\mathbf{L_S}$, we can directly integrate the $\mathbf{P_C}$ with $\mathbf{L_S}$, which can be written as follow

$$\mathbf{L_S} = \mathbf{L_S} + \mathbf{P_C}, \tag{31}$$

where $\mathbf{P_C}$ is calculated as previously described (30).

The encoded latent switch feature $\mathbf{L_S}$ is then fed into a transformer encoder, which comprises I distinct transformer layers. Each transformer layer is structured with a multi-head self-attention block, denoted by $\mathrm{MSA}(\cdot)$; a residual learning block; layer normalization, denoted by $\mathrm{LN}(\cdot)$; and a feed-forward network, denoted by $\mathrm{FFN}(\cdot)$, as depicted in Fig. 3. In order to facilitate the modeling of different user beamforming characteristics across various subspaces, multiple linear layers are employed in the $\mathrm{MSA}(\cdot)$ to generate distinct latent switch feature for different users. The input to the $\mathrm{MSA}(\cdot)$ is formulated as follows:

$$\mathbf{L'_S} = \mathrm{con}\Big(\mathrm{Linear}_{0,5}(\mathbf{L_S}), \mathrm{Linear}_{0,6}(\mathbf{L_S}), \cdots, \mathrm{Linear}_{0,5+K}(\mathbf{L_S})\Big), \tag{32}$$

where $\mathbf{L'_S} \in \mathbb{R}^{KN_{RF}L \times N_{RF}L}$ represents the concatenated latent switch feature maps for different users. The output, $\mathbf{Y} = \tilde{\mathbf{L}}_{\mathbf{S}}$, of a single self-attention layer, denoted by $\mathrm{SA}(\cdot)$, within the $\mathrm{MSA}(\cdot)$ block is expressed as

$$\mathbf{Y} = \mathrm{SA}(X) = \mathbf{AV} = \mathrm{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{33}$$

where

$$\mathbf{Q} = \mathrm{Linear}_{0,K+6}(X), \tag{34a}$$

$$\mathbf{K} = \mathrm{Linear}_{0,K+7}(X), \tag{34b}$$

$$\mathbf{V} = \mathrm{Linear}_{0,K+8}(X). \tag{34c}$$

Moreover, $\mathbf{X} = \mathbf{L'_S}$ represents the input latent feature of $\mathrm{SA}(\cdot)$ and $\mathbf{A}$ represents the calculated self-attention weights, signifying the varying beamforming effects among users. The $\mathrm{MSA}(\cdot)$ further enhance the relative importance among users by aggregating outputs from several self-attention layers, described as

$$\mathrm{MSA}(\mathbf{L'_S}) = \mathrm{Linear}_{0,K+9}\Big(\mathrm{con}\big(\mathrm{SA}(\mathbf{L'_S})_1, \mathrm{SA}(\mathbf{L'_S})_2, \ldots, \mathrm{SA}(\mathbf{L'_S})_J\big)\Big). \tag{35}$$

Here, $J = K$ represents the number of users, and $\mathrm{SA}(\mathbf{L'_S})_j$ denotes the output from the j-th parallel self-attention layer. The latent switch feature maps resulting from the $\mathrm{MSA}(\cdot)$ layer are subsequently passed to the $\mathrm{FFN}(\cdot)$. To clarify the computational process within the i-th transformer layer ($i \in [1, 2, \ldots, I = 8]$), the encoding-decoding process is expressed as

$$\mathrm{FFN}(\mathbf{X}) = \mathrm{Linear}_{i,2}(\mathrm{GELU}(\mathrm{Linear}_{i,3}(\mathbf{X}))), \tag{36}$$

$$\tilde{\mathbf{L}}_{\mathbf{S}_i} = \mathrm{MSA}(\mathrm{LN}(\tilde{\mathbf{L}}_{\mathbf{S}_{i-1}})) + \tilde{\mathbf{L}}_{\mathbf{S}_{i-1}}, \tag{37}$$

$$\tilde{\mathbf{L}}_{\mathbf{S}_i} = \mathrm{FFN}(\tilde{\mathbf{L}}_{\mathbf{S}_i}) + \tilde{\mathbf{L}}_{\mathbf{S}_i}, \tag{38}$$

where $\mathbf{X}$ represents the input feature to the $\mathrm{FFN}(\cdot)$, $\tilde{\mathbf{L}}_{\mathbf{S}_i}$ represents the feature processed by the i-th transformer layer, with $\tilde{\mathbf{L}}_{\mathbf{S}_0} = \mathbf{L_S}$ being the initial input latent switch feature, and $\mathbf{S'} = \tilde{\mathbf{L}}_{\mathbf{S}_8} \in \mathbb{R}^{L \times L}$ indicating the predicted connection matrix of the final output. The $\mathrm{FFN}(\cdot)$ [38] block consists of two linear layers with a Gaussian Error Linear Unit (GELU) activation function between them. The input and output feature dimensions for the $\mathrm{FFN}(\cdot)$ are set to 128, and the dimension of the intermediate layer is expanded to 512 to enhance feature representation ability.

Considering the extremely large number of subarrays, it becomes impractical to directly converts the continuous values in the predicted connection map $\mathbf{S'} \in \mathbb{R}^{N_{RF}L \times N_{RF}L}$ into a binary matrix through classification methods [26]. Additionally, each subarray must be uniquely connected to one TTD, and vice versa. To effectively tackle this challenge, the Hungatian algorithm is employed to optimize the connection selection of $\mathbf{S'}$. We initially define a permutation matrix $\mathbf{P} \in \mathbb{R}^{N_{RF}L \times N_{RF}L}$, which indicates the assignments of subarrays to TTDs. The algorithm treats $\mathbf{S'}$ as the initial cost

Algorithm 1 Hungarian Algorithm for Assignment Problems

Let $\mathbf{C}$ be the initial cost matrix with dimensions $N_{RF}L \times N_{RF}L$
The objective is to maximize the total cost of the assignment:
$\max_{\mathbf{P}} \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} \cdot C_{ij}$
for $i \in \{1, \ldots, n\}$ **do**
  $C_{i,:} \leftarrow C_{i,:} - \min(C_{i,:})$
end for
for $j \in \{1, \ldots, n\}$ **do**
  $C_{:,j} \leftarrow C_{:,j} - \min(C_{:,j})$
end for
while number of lines $P < n$ **do**
  $s \leftarrow \min(\text{uncovered elements of } \mathbf{C})$
  Adjust $\mathbf{C}$ with respect to s
end while
Derive $\mathbf{P}$ from $\mathbf{C}$ as the assignment matrix
return $\mathbf{P}$

matrix, denoted as $\mathbf{C} \in \mathbb{R}^{N_{RF}L \times N_{RF}L}$. Then the problem is reformulated to maximize the sum of products between the assignment decisions and their corresponding costs. The algorithm starts by subtracting the minimum value from each row and column of their respective elements, resulting in a matrix with zero-valued elements indicative of potential assignments. Through iterative refinement, we arrive at an optimal pairing between TTDs and PSs, as illustrated in **Algorithm 1**.

C. Network Architecture for Adaptive TTD Beamforming

Our proposed network architecture for adaptive TTD configuration in near-field beamforming includes the NFC-LM and the S-MT module. However, implementing a hard threshold-based strategy directly in the training of the entire network poses a challenge, particularly because the maximization operation $\max_{\mathbf{P}} \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} \cdot C_{ij}$ outlined in **Algorithm 1** can hinder the backpropagation process within the S-MT. This limitation makes it difficult to optimize the generated beamformers $\mathbf{\Phi}$, $\mathbf{T}$ and $\mathbf{D}$ jointly with connection matrix $\mathbf{S}$. To address this, we introduce a MCA block, denoted by $\mathrm{MCA}(\cdot)$, to enhace the feature connections between the S-MT and NFC-LM modules. As showed in Fig. 1, each latent feature, characterized by beamformers and the switch network, is concatenated with others, serving as the $\mathbf{Q}$. After that, MCA block leverages cross-attention to facilitate the simultaneous optimization of switch selection and beamformer design. The $\mathrm{MCA}(\cdot)$ processes inputs $\mathbf{L}_{\mathbf{\Phi}}, \mathbf{L}_{\mathbf{T}}, \mathbf{L}_{\mathbf{D}}$ and $\mathbf{L}_{\mathbf{S}}$, producing the outputs as follows:

$$\mathbf{Y} = \mathrm{MCA}(\mathbf{X}, \mathbf{X}') = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{39}$$

where

$$\mathbf{Q} = \mathrm{Linear}_{0,K+9}(\mathbf{X}'), \tag{40a}$$
$$\mathbf{K} = \mathrm{Linear}_{0,K+10}(\mathbf{X}), \tag{40b}$$
$$\mathbf{V} = \mathrm{Linear}_{0,K+11}(\mathbf{X}). \tag{40c}$$

$\mathbf{X}' = \mathrm{con}(\mathbf{L}_{\mathbf{\Phi}}, \mathbf{L}_{\mathbf{T}}, \mathbf{L}_{\mathbf{D}}, \mathbf{L}_{\mathbf{S}})$ represents the concatenation of the inputs, and $\mathbf{X} \in \{\mathbf{L}_{\mathbf{\Phi}}, \mathbf{L}_{\mathbf{T}}, \mathbf{L}_{\mathbf{D}}, \mathbf{L}_{\mathbf{S}}\}$ represents each input separately. Furthermore, considering the physical constraint that the TTD beamformer $\mathbf{T}$ must be non-negative real numbers. Additionally, according to [18], the serial structure utilize increased time delays to improve beamforming performance. Thus, guaranteeing that each TTD provides a non-zero time delay can significantly enhance spectral efficiency. To maintain the backpropagation process while ensuring $\mathbf{T}$ remains positive, we utilize the Softplus function, denoted by $\mathrm{Softplus}(\cdot)$, rather than the ReLU function:

$$\mathrm{Softplus}(x) = \ln(1 + e^x), \tag{41}$$

where x is the input to the $\mathrm{Softplus}(\cdot)$. Different from the ReLU function, which outputs x for positive inputs and 0 otherwise, the Softplus function ensures the outputs are always positive and provides a differentiable gradient at every point, including $x = 0$. Therefore, our proposed NFC-LM can adaptively optimize the time delay design of each TTD.

D. Loss Function and Network Training Process

We utilize the NFC-LM to design the $\mathbf{\Phi}$, $\mathbf{T}$, and $\mathbf{D}$ through extracting the latent feature from the input channel response matrix $\mathbf{H}$, and further utilize the S-MT to design the connection matrix $\mathbf{S}$. While numerous DL-based hybrid beamforming methods transform the optimization problem into a convex problem and calculate the optimal result as training labels [24], [25], [26], [27], this approach is impractical in our proposed physical structure. The proposed adaptive TTD configuration structure poses a significant challenge in providing optimal solutions for every possible connection scenarios. Therefore, we adopt an unsupervised, end-to-end training methodology for our hybrid beamforming network, eliminating the need for predetermined optimal solutions for $\mathbf{\Phi}$, $\mathbf{T}$, $\mathbf{D}$ and $\mathbf{S}$.

The proposed loss function consists of two parts: the optimization objective and a set of regularization terms. The primary goal of the proposed adaptive TTD beamforming network is to maximize the spectral efficiency, as expressed in (17a). To facilitate this, we first reshape the $\mathbf{H}$, $\mathbf{\Phi}$, $\mathbf{T}$, $\mathbf{D}$, and $\mathbf{S}$ into the same format $\mathbb{R}^{K \times N_R \times M \times N}$. We then introduce an operation, H_+, which performs $\odot$ across these matrices and sums the results over the N_{RF}, M, and N dimensions. The spectral efficiency optimization term for each k in K is expressed as

$$\mathcal{L}_{\mathrm{Eff}} = -\frac{1}{M+L_{\mathrm{CP}}} \times \log_2\left(1+\frac{\left|\mathrm{H}_+(\mathbf{H}_k, \mathbf{\Phi}_k, \mathbf{S}_k, \mathbf{T}_k, \mathbf{D}_k)\right|^2}{\sum_{i=1, i\neq k}^{K}\left|\mathrm{H}_+(\mathbf{H}_k^H, \mathbf{\Phi}_k, \mathbf{S}_k, \mathbf{T}_k, \mathbf{D}_k)\right|^2+\sigma^2}\right). \tag{42}$$

To enhance the training process and ensure the practicality of our solutions, we introduce three regularization terms:

1) **PS Modulus Constraint**: To maintain the constant modulus nature of the PSs, we employ the Mean Squared Error (MSE) on the magnitude of the predicted PS values $\mathbf{\Phi}$ formulated as

$$\mathcal{L}_{\mathrm{PS}} = \sum_{k=1}^{K}\sum_{N_{RF}=1}^{N_R}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(|\mathbf{\Phi}_{kn_rmn}|^2 - 1\right)^2. \tag{43}$$

2) **TTD Range Constraint**: To ensure that the predicted time delays are within the hardware's feasible range, we apply a conditional MSE that penalizes values outside this range, formulated as

$$\mathcal{L}_{\mathrm{TTD}} = \sum_{k=1}^{K}\sum_{N_{RF}=1}^{N_R}\sum_{m=1}^{M}\sum_{n=1}^{N}\psi(\mathbf{T}_{k,n_r,m,n}) \tag{44}$$

where

$$\psi(x) = \begin{cases} (x - t_{\max})^2, & \text{if } x > t_{\max}, \\ 0, & \text{if } 0 < x < t_{\max}, \\ x^2, & \text{if } x < 0. \end{cases} \tag{45}$$

3) **Power Consumption Constraint:** To encourage energy-efficient beamforming and connection designs, we constrain the total power consumption to not exceed a

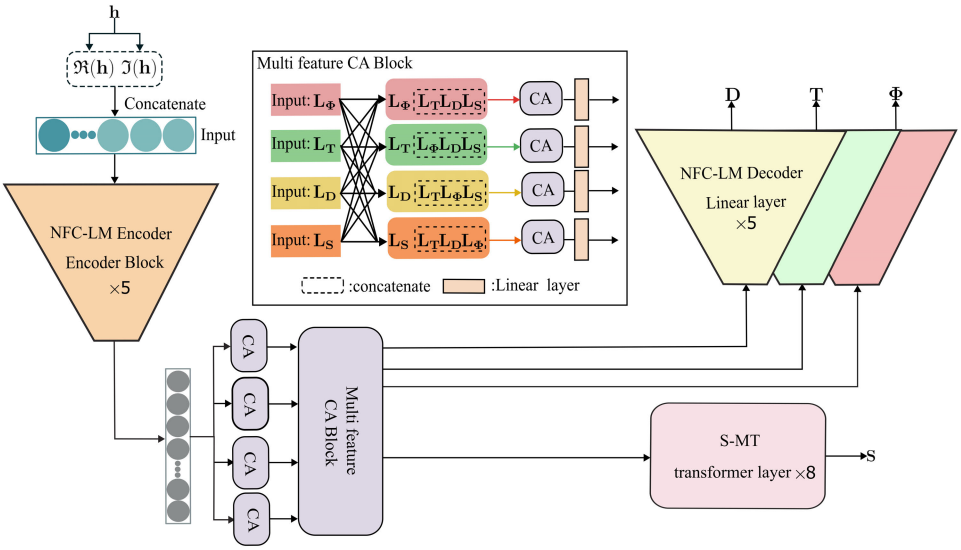


Fig. 4. The network structure of the proposed adaptive TTD configuration beamforming method.

predefined limit P_t expressed as

$$\mathcal{L}_{\mathrm{PC}} = \left(\sum_{k=1}^{K} \sum_{N_{RF}=1}^{N_R} \sum_{m=1}^{M} \sum_{n=1}^{N} \|\mathbf{\Phi S T D}\|_F^2 - P_t \right)^2 . \tag{46}$$

Combining these components, the total loss function for our adaptive beamformer training is formulated as

$$\mathcal{L} = \mathcal{L}_{\mathrm{Eff}} + \mathcal{L}_{\mathrm{PS}} + \mathcal{L}_{\mathrm{TTD}} + \mathcal{L}_{\mathrm{PC}}. \tag{47}$$

Our method consists of training and testing stages. In the training stage, we set the batch size to 64 and use Adam optimizer to train our proposed network for 200 epochs. The learning rate is initialized at 1e-5 and decays by a factor of 0.5 every 10 epochs. To prevent overfitting, we implement early stopping if the validation loss does not decrease over 50 epochs. Training the entire network takes approximately 30 minutes, with a parameter size of 1 million. The testing stage takes an average of 0.4 seconds under different transmit powers. To facilitate the training of whole network, we begin by training the NFC-LM. This module extracts latent features of the near field channel and constructs relationships among the channel response, analog beamformers and digital beamformers using $\mathcal{L}_{\mathrm{Eff}}$. Afterward, the whole network is trained by combining the $\mathcal{L}_{\mathrm{PS}}$, $\mathcal{L}_{\mathrm{TTD}}$ and $\mathcal{L}_{\mathrm{PC}}$ to enforce hardware constraints and improve the final prediction. The Switch Multi-User Transformer (S-MT) is designed with Hungarian algorithm for adaptive TTD configuration. In this way, the proposed network can not only learn the latent correlations among the near field channel response, analog beamformers and digital beamformers, but also effectively model the adaptive connection between antennas and TTDs for hybrid near field beamforming.

.

IV. Numerical Results

In this section, we conduct experiments to evaluate the effectiveness of our proposed deep unsupervised learning-based Adaptive TTD configuration for near-field beamforming.

A. Simulation Setup

To demonstrate the robustness of our proposed adaptive TTD configuration beamforming method across various antenna structures, we conducted experiments using two prevalent antenna configurations: the Uniform Linear Array (ULA) and the Uniform Circular Array (UCA). For the ULA scenario, the scattering effects of the signal paths between the BS and users are numerically represented by L_k. As expressed in (3) and (4), the user positions are set by r_k and θ_k, representing the distance and direction relative to the center of the ULA at the BS, respectively. The position for the l-th scatter are denoted by $\tilde{r}_{k,l}$ and $\tilde{\theta}_{k,l}$. Here, C represents the speed of light, and $\mathbf{r}(r,\theta) \in \mathbb{C}^{N\times 1}$ represents the distances wave travel from the BS antennas to the user. According to the near-field spherical wave model [39], the n-th component of $\mathbf{r}(r,\theta)$, $[\mathbf{r}(r,\theta)]_n$ is defined as $\sqrt{r^2+\delta_n^2 d^2 - 2r\delta_n d cos\theta}$, with $d = c/(2f_c)$ indicating the antenna spacing and $\delta_n \triangleq n - 1 - \frac{N-1}{2}$. Similarly, for the UCA scenario, the positions of k-th user and l-th scatter are also represented by (r_k, θ_k) and $(\tilde{r}_{k,l}, \tilde{\theta}_{k,l})$, respectively. Our model presumes the presence of user is confined to the two-dimensional plane coinciding with the UCA. Additionally, different from the ULA, the n-th component of $\mathbf{r}(r,\theta)$ for UCA is formulated as $\sqrt{r^2+R^2-2rRcos(\theta-\psi_n)}$, where R is the UCA radius, and $\psi_n = \frac{2\pi n}{N}$. To maintain consistent experimental conditions between the ULA and UCA, we calculate the radius such that $2R = N\frac{C}{f_c}$.

For data generation, we assume users and scatters are randomly distributed in a semi-circular area spanning 0 to 180 degrees around the BS, with distances ranging from 5 to 15 meters. User positions are precisely sampled at 0.1-meter intervals and 0.5-degree angles within this domain. Following the (3), we generated 10,000 datasets for both ULA and UCA channels, with each user associated with four scatterers.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Transmit power at the BS P_t	20 dBm
Noise power density	-174 dBm/Hz
Number of antennas at the BS N	512
System bandwidth B	10 GHz
Central OFDM frequency f_c	100 GHz
Number of OFDM subcarriers M	10
Length of OFDM cyclic prefix L_{CP}	4
Number of TTDs for each RF chain L	32
Maximum time delay of TTDs t_{max}	80 ps
Number of channel paths L_k	4
Scattering loss Λ_ℓ	-15 dB
Radius of the UCA R	0.768 m
Transmit and receive antenna gain G_t, G_r	15 dB, 5 dB

Consequently, every channel dataset is a blend of LOS and NLOS channel information. The datasets were divided into training (60%), testing (20%), and validation (20%) sets. In the millimeter-wave and terahertz bands, path loss is mainly affected by propagation loss, but absorption loss can also have an effect. The path loss for a given frequency f and propagation distance r can be expressed as

$$\eta_{pathloss}(f,r) = \left(\frac{4\pi f r}{c}\right)^2 e^{k_{abs}(f)r}, \tag{48}$$

where k_{abs} denotes the frequency-specific medium absorption coefficient, obtainable from the HITRAN database [42]. The LOS channel gain is formulated as $|\beta_{m,k}|^2 = \eta^{-1}_{pathloss}(f_m, r_k)G_rG_t$, with G_r and G_t receiver and transmitter antenna gains, respectively. For the NLoS component, incorporating scattering loss is crucial. The gain from the l-th NLoS component is expressed as $|\beta_{m,k,l}|^2 = \Lambda_l\eta^{-1}_{pathloss}(f_m, r_k)G_rG_t$, where Λ_l denotes the scattering loss and $r_{k,l}$ denotes the propagation distance to user k through the l-th scatterer. Unless specified otherwise, the simulation follows parameters listed in Table I. Our method was implemented in Pytorch and executed on an NVIDIA GeForce A40 GPU. During the training phase, we utilized a batch size of 2 and employed the Adam optimizer across 1000 epochs.

For comparison, we consider the following three benchmark scenarios:

- **Optimal Full-Digital Beamforming (BF)**:In this scenario, each antenna is individually connected to a dedicated Radio Frequency (RF) chain, enabling the creation of a comprehensive baseband digital beamformer for every subcarrier. This setup is designed to establish a theoretical upper bound for performance metrics.
- **Optimal Time Delay Beamforming (TTD-BF)**:This benchmark presumes the availability of TTDs with an infinite range, i.e., $t_{max} = +\infty$, thereby facilitating the highest possible performance for the TTD-BF architecture.
- **Conventional Beamforming (CB)**: This benchmark corresponds to a Phase Shifter (PS)-only hybrid BF architecture. It is constrained to frequency-independent analog beamforming capabilities.

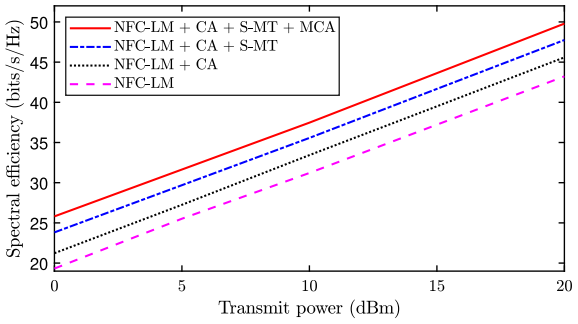


Fig. 5. Spectral efficiency results of different network structures based on the NFC-LM across varying levels of transmit power.

TABLE II
ABLATION STUDY ON THE NUMBER OF TRANSFORMER LAYERS (I) AND EMBEDDING FEATURE DIMENSIONALITY

Layer # (I)	Feature Dim.	Spectral Efficiency (bit/s/Hz)				
		0 dBm	5 dBm	10 dBm	15 dBm	20 dBm
1	512	22.47	29.26	33.89	40.27	46.83
8	512	24.31	30.31	36.00	42.09	48.25
8	512	**25.81**	**31.64**	**37.47**	**43.63**	**49.79**
4	256	23.81	29.69	35.57	41.76	47.79
4	768	25.28	31.29	36.84	43.12	49.19

B. Ablation Studies

In this subsection, we conduct a series of ablation studies to evaluate the effect of critical components and hyperparameter settings of our proposed adaptive TTD configuration beamforming method on spectral efficiency, particularly for short-range TTDs. These essential components include the NFC-LM, CA, S-MT, and MCA. The generated ULA dataset is used in the ablation studies. We treat the NFC-LM as our baseline model which consists of multiple linear decoders and a convolutional encoder but excludes the CA, S-MT, and MCA modules. Moreover, to demonstrate the effectiveness of our proposed S-MT, the NFC-LM employs an additional linear decoder in place of the S-MT which is identical to those decoding the variables $\mathbf{\Phi}$, $\mathbf{T}$ and $\mathbf{D}$. Despite these modifications, we continue to apply the Hungarian algorithm to map the predicted connection matrix $\mathbf{S}'$ to a binary matrix $\mathbf{S}$. The comparisons of spectral efficiency results are shown in Fig. 5 and Table II.

As shown in Fig. 5, we gradually incorporate the CA, S-MT, and MCA modules into the NFC-LM across various transmit power settings, denoted as "NFC-LM", "NFC-LM + CA", "NFC-LM + CA + S-MT", and "NFC-LM + CA + S-MT + MCA", with the last configuration representing our proposed method. The results clearly indicate a progressive improvement in spectral efficiency with the sequential integration of CA, S-MT and MCA. The NFC-LM, operating with short-range TTDs, shows the lower bound of our proposed latent feature learning based adaptive TTD beamforming method. Additionally, our fully enhanced method, 'NFC-LM + CA + S-MT + MCA', achieves an average improvement of approximately 6.5 bit/s/Hz in the results over the NFC-LM across varying levels of transmit power. The CA module, designed to establish the correlation between the near-field channel latent

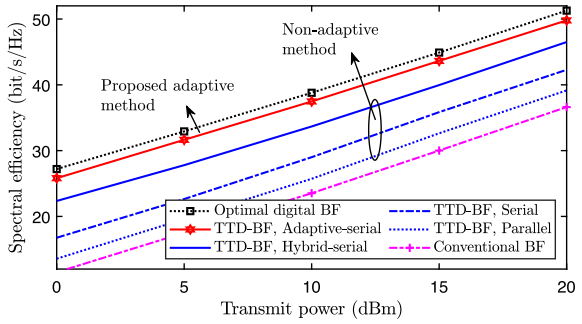


Fig. 6. Average spectral efficiency versus the maximum transmit power in the multi-user ULA system.

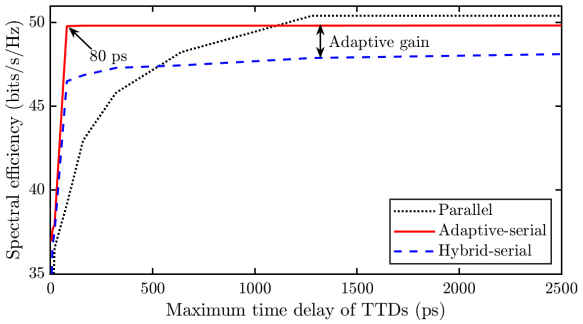


Fig. 7. Average spectral efficiency versus the maximum time delay t_{max} of TTDs in the multi-user ULA system.

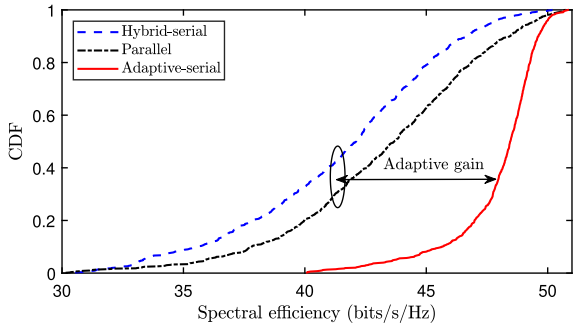


Fig. 8. CDF versus average spectral efficiency for multi-user ULA system.

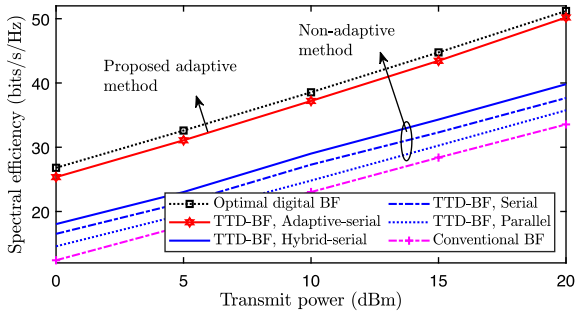


Fig. 9. Average spectral efficiency versus the maximum transmit power in the multi-user UCA system.

feature $\mathbf{H}$ and the latent features of multiple beamformers $\mathbf{\Phi}$, $\mathbf{T}$ and $\mathbf{D}$, augments spectral efficiency by more than 1 bit/s/Hz on average compared to the NFC-LM. Moreover, the S-MT is capable of optimizing the adaptive connection of the TTD through MSA and PC, effectively enhancing the performance of the network. In addition, the MCA module further enhances the joint optimization of $\mathbf{\Phi}$, $\mathbf{T}$, $\mathbf{D}$ and $\mathbf{S}$, leading to an average spectral efficiency increase of 2 bit/s/Hz.

Furthermore, we conduct ablation experiments to evaluate the impact of hyperparameter configurations of S-MT on spectral efficiency. We specifically examine the symmetrical structure of the transformer by modifying the number of layers (I) at 1, 4, 8, and by modifying the feature dimensionality in the linear projection layer of the FFN to 256, 512, 768. The comparative analysis, presented in Table I, clearly shows the spectral efficiency at different transmit power levels. We can find that the optimal performance is obtained with a network configuration of $I = 8$ layers and a feature dimension of 512, yielding spectral efficiency of 28.82, 34.6, 37.97, 42.06 and 47.95 bit/s/Hz at transmit powers of 0, 5, 10, 15, and 20 dBm, respectively.

C. Spectral Efficiency Versus Transmit Power

In our analysis of multi-user scenarios, we explore the spectral efficiency of ULA and UCA systems under conditions of maximum transmit power and diverse TTD configurations. The evaluations, as presented in Fig. 6 and Fig. 9 are based on scenarios where the distance between users and the BS varies randomly from 5 to 15 meters. Our experimental setup involves $K = 4$ users, $N_{RF} = 4$ RF chains, and a TTD constraint of $t_{max} = 80$ ps. Here, "Hybrid" implies that the TTDs for all RF chains are cascaded in serial configuration and hybrid connected with the PSs [33]. And the "parallel" means that each TTD is connected to an individual antenna. As showed in Fig. 6, under the ULA scenario, the CB configuration achieves comparable results to the parallel TTD setup within the given time delay constraints. It is notable that serial configurations surpass their parallel counterparts in performance, attributed to their superior capability in providing more extensive time delay with short-range TTDs. Furthermore, the hybrid configuration, which optimizes the effective coverage area, demonstrates a significant improvement in performance. Therefore, when employing an adaptive TTD configuration which optimizes the connection between TTDs and PSs, our proposed method approaches the theoretical optimum. In addition, thanks to the adaptive TTD configuration, our proposed method can have less performance degradation under different antenna structures. As showed in Fig. 9, our proposed method still achieves comparable results to the theoretical optimum, however, hybrid configuration experience severe performance degradation. Changing the antenna structure disrupts the guaranteed monotonic correspondence between time delays and user positions [18]. Thus, we can find that the parallel configuration outperforms the serial configuration. The results suggest that our proposed adaptive TTD configurations can substantially enhance spectral efficiency in both multi-user ULA and UCA systems.

D. Spectral Efficiency Versus Maximum Time Delay

To further illustrate the effectiveness of our proposed adaptive TTD configurations, we examine the relationship between spectral efficiency and infinite maximum time delay t_{max} fot TTDs, as illustrated in Fig. 7 and Fig. 10. Particularly for maximum time delays below 80 ps, the adaptive

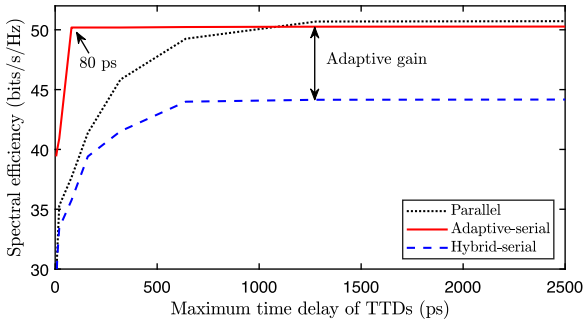


Fig. 10. Average spectral efficiency versus the maximum time delay t_{max} of TTDs in the multi-user UCA system.

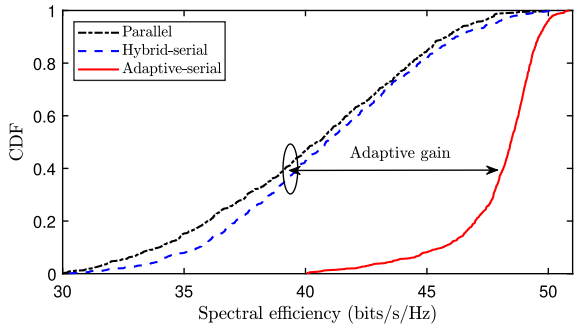


Fig. 11. CDF versus average spectral efficiency for multi-user UCA system.

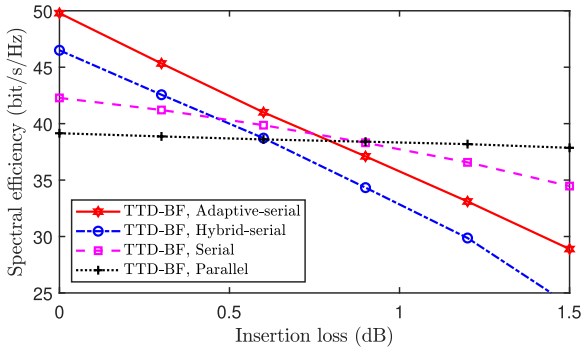


Fig. 12. Average spectral efficiency versus insertion loss in the multi-user ULA system.

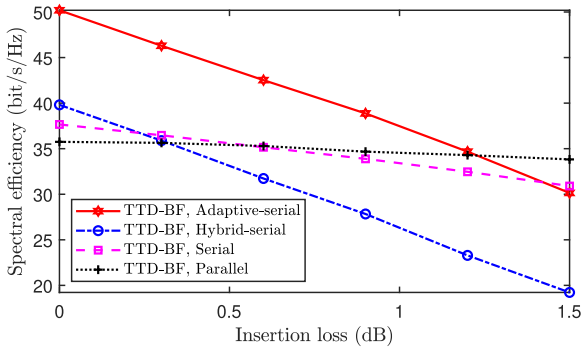


Fig. 13. Average spectral efficiency versus insertion loss in the multi-user UCA system.

configuration shows superior performance compared to the other configurations. This indicates that the adaptive approach effectively leverages dynamic switching to mitigate time delay discrepancies among users, thereby sustaining higher spectral efficiency even with limited time delay capacity. This observation demonstrates the robustness of our proposed adaptive TTD configuration in environments constrained by finite and narrower TTD ranges. As t_{max} increases, all configurations show an clear increase in spectral efficiency, but they exhibit different behaviors. In the ULA system, as shown in Fig. 7, the parallel configuration necessitates a t_{max} of at least 500 ps to match the performance of the serial configuration. However, as shown in Fig. 10, within the UCA system, parallel configuration obviously outperforms the serial configuration in pectral efficiency. For the UCA system, achieving comparable performance with the ULA system through a serial configuration demands a significantly high time delay, surpassing 500 ps. Moreover, we can find that our proposed adaptive configuration achieves comparable results to the parallel configuration when considering an infinite time delay. However, as the maximum time delay t_{max} increases, the parallel configuration begins to show increasing benefits. A primary advantage of the parallel configuration is its direct compensation for each antenna subarray by its respective TTD, simplifying the computational demands. Therefore, for applications accommodating large time delays while aiming to minimize hardware complexity, the parallel configuration may offer enhanced performance and increased adaptability.

E. Cumulative Distribution of Spectral Efficiency

To demonstrate the robustness of our proposed adaptive TTD configuration beamforming in near-field region, we present Fig. 8 and Fig. 11. Here, we fix power consumption at 20 dB and the time delay at 80 ps. We position the users uniformly at a distance of 10 meters from the BS, with their angular distribution ranging from 0 to 180 degrees.

As showed in Fig. 8, we can find that the serial and parallel configuration have similar distribution and serial has higher spectral efficiency. This advantage stems from the serial configuration can provide cumulative time delay compensation, which helps alleviate the spatial-wideband effect with short-range TTDs. Furthermore, the spectral efficiency of our proposed adaptive TTD configuration mainly distributed between 40 and 56, indicating a performance enhancement over the serial configuration. Similarly, under the UCA system, as shown in Fig. 11, our proposed adaptive configuration surpasses both the parallel and serial configurations, with spectral efficiency values distributed between 40 and 53. This is because the adaptive configuration are more robust and can alleviate the performance decrease under different antenna systems. Additionally, it is observed that, under the UCA system, the parallel configuration achieves performance comparable to the serial configuration.

F. Spectral Efficiency Versus Number of Subcarriers

To demonstrate the effectiveness of our proposed method with different numbers of subcarriers, we further analyzed spectral efficiency in ULA and UCA configurations under varying subcarrier settings, all with the same transmit power. As shown in Fig. 14 and Fig. 15, the spectral efficiency increases by 10 bit/s/Hz and 12 bit/s/Hz for UCA and ULA configurations, respectively, when moving from 10 to

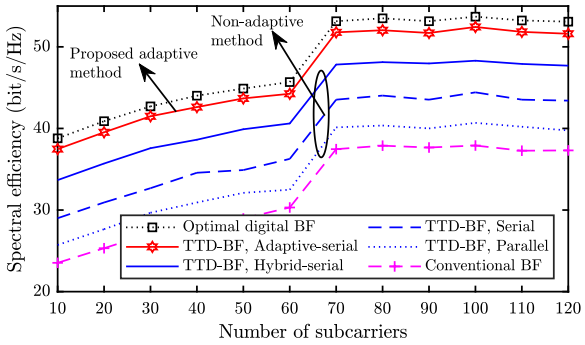


Fig. 14. Average spectral efficiency versus number of subcarriers in the multi-user ULA system.

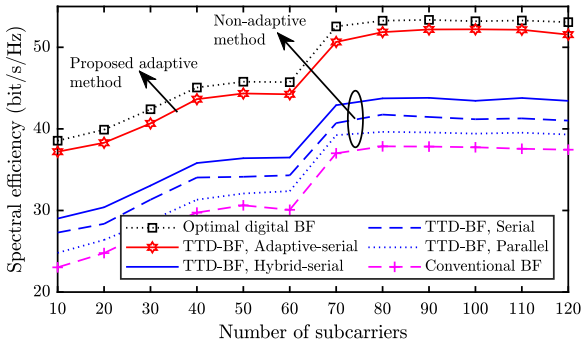


Fig. 15. Average spectral efficiency versus number of subcarriers in the multi-user UCA system.

120 subcarriers. Additionally, Fig. 14 and Fig. 15 indicate that once the number of subcarriers exceeds 70, the spectral efficiency of digital beamforming remains nearly constant. With the encoder-decoder structure, our proposed adaptive TTD hybrid beamforming method is adaptable to arbitrary subcarrier numbers. Therefore, even as the number of subcarriers increases, our method continues to approach the theoretical optimum. Other methods, including hybrid-serial, serial, parallel, and conventional beamforming, exhibit similar trends as the number of subcarriers increases from 10. Under the ULA configuration, hybrid-serial beamforming surpasses serial, parallel, and conventional beamforming, showing significant improvement over the serial method. However, under the UCA configuration, hybrid-serial beamforming experiences severe performance degradation. The loss of the guaranteed monotonic correspondence between time delays and user positions allows the parallel configuration to outperform the serial configuration.

G. Spectral Efficiency Versus Insertion Loss

Both TTDs and power splitters introduce additional insertion loss, which can negatively impact beamforming performance. Therefore, we further extend the analysis to examine the performance oscillations between insertion loss and spectral efficiency in both ULA and UCA configurations. In our previous work [18], we proposed a power equalization approach to address the cumulative insertion loss associated with serial and hybrid TTD configurations. As demonstrated in [18], the superiority of equalized power splitter schemes in serial and hybrid configurations is evident, as they ensure the effective utilization of each individual TTD. This equalization is crucial for optimizing performance, as it allows for more efficient management of the insertion loss, thereby enhancing overall beamforming efficiency. In contrast, the cumulative insertion loss introduced by unequalized power splitter schemes can lead to diminished output power, causing the antenna sub-arrays connected to the TTDs at the end of the serial configuration to operate less effectively. Therefore, our analysis only focuses on scenarios employing equalized power splitter schemes. As shown in Fig. 12 and Fig. 13, the spectral efficiency of our proposed method, as well as the hybrid-serial, serial, and parallel configurations, gradually decreases with increasing insertion loss. Notably, the serial, hybrid-serial, and our proposed method are more sensitive to insertion loss compared to the parallel configuration due to the accumulation of TTDs and power splitters. Specifically, under the ULA configuration, when the insertion loss reaches 0.5 dB and 0.8 dB, the performance of the parallel configuration surpasses that of our proposed method, hybrid-serial, and serial. In contrast, under the UCA configuration, when the insertion loss reaches 0.25 dB and 0.5 dB, the spectral efficiency of the parallel configuration already exceeds that of the hybrid-serial and serial configurations, respectively. Moreover, when the insertion loss increases to 0.75 dB, our proposed method becomes inferior to the parallel configuration. Although low-loss TTDs and power splitters with insertion losses of 0.07 dB [43] and 0.15 dB [44], respectively, are available, they are only applicable to low-frequency bands. The development of low-loss TTDs and power splitters for high-frequency bands remains an open challenge that requires further research. Overcoming the performance oscillation introduced by TTDs and power splitters in these high-frequency bands remains an unresolved issue.

V. Conclusion

In this article, we introduced an adaptive TTD configuration designed for short-range TTDs. Compared to other existing TTD configurations, our proposed method effectively mitigated the spatial-wideband effect across arbitrary user locations and array shapes by dynamically selecting the connections between TTDs and PSs. We proposed a novel end-to-end deep neural network architecture, which consisted of a NFC-LM and a S-MT module. Utilizing an encoding-decoding structure, the NFC-LM explored the latent features of the near-field channel response and employed a CA module to establish relationships among the channel response, analog beamformers, digital beamformer, and switch network. Additionally, the S-MT module guided the connections between TTDs and antenna subarrays through MSA and PC. Furthermore, we introduced a MCA module to enhance the joint optimization of beamformer design and switch network configuration. Simulation results demonstrates that our proposed adaptive TTD configuration outperforms other hybrid TTD configuration beamforming methods in various antenna structures. While our approach has demonstrated significant effectiveness, there are several potential directions for further enhancement. One area of future exploration involves extending the generalization of our method to accommodate varying

transmit power requirements. By incorporating transmit power as a variable input into our adaptive TTD hybrid beamforming method, we can broaden its applicability across different power settings. Additionally, to improve the generalization ability of our proposed method across various bandwidths, a similar strategy could be applied. This would involve integrating time delay restrictions as one of the network inputs, thereby enabling the method to adapt more effectively to different bandwidth conditions. These potential modifications could significantly expand the versatility and robustness of our approach in diverse communication scenarios.

REFERENCES

[1] C.-X. Wang et al., "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart., 2023.

[2] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and terahertz frequency bands," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 102–108, Jun. 2018.

[3] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, Jan. 2023.

[4] Y. Liu, J. Xu, Z. Wang, X. Mu, and L. Hanzo, "Near-field communications: What will be different?" 2023, *arXiv:2303.04003*.

[5] F. Gao, B. Wang, C. Xing, J. An, and G. Y. Li, "Wideband beamforming for hybrid massive MIMO terahertz communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1725–1740, Jun. 2021.

[6] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.

[7] S. Payami, M. Ghoraishi, M. Dianati, and M. Sellathurai, "Hybrid beamforming with a reduced number of phase shifters for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4843–4851, Jun. 2018.

[8] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog–digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[9] L. Yan, C. Han, N. Yang, and J. Yuan, "Dynamic-subarray with fixed phase shifters for energy-efficient terahertz hybrid beamforming under partial CSI," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3231–3245, May 2023.

[10] L. Yan, C. Han, and J. Yuan, "Energy-efficient dynamic-subarray with fixed true-time-delay design for terahertz wideband hybrid beamforming," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2840–2854, Oct. 2022.

[11] M. Longbrake, "True time-delay beamsteering for radar," in *Proc. IEEE Nat. Aerosp. Electron. Conf. (NAECON)*, Jul. 2012, pp. 246–249.

[12] R. Rotman, M. Tur, and L. Yaron, "True time delay in phased arrays," *Proc. IEEE*, vol. 104, no. 3, pp. 504–518, Mar. 2016.

[13] K. Spoof, V. Unnikrishnan, M. Zahra, K. Stadius, M. Kosunen, and J. Ryynänen, "True-time-delay beamforming receiver with RF resampling," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 12, pp. 4457–4469, Dec. 2020.

[14] L. Dai, J. Tan, Z. Chen, and H. V. Poor, "Delay-phase precoding for wideband THz massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7271–7286, Sep. 2022.

[15] M. Cui and L. Dai, "Near-field wideband beamforming for extremely large antenna arrays," 2021, *arXiv:2109.10054*.

[16] J. Xie, Y. Luo, H. Yang, Z. Liu, and C. Luo, "A dynamic group-connected design with true-time-delay for wideband THz beamforming," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 1441–1446, Jan. 2024.

[17] A. Najjar, M. El-Absi, and T. Kaiser, "Hybrid delay-phase precoding in wideband UM-MIMO systems under true time delay and phase shifter hardware limitations," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7246–7262, Jul. 2024.

[18] Z. Wang, X. Mu, Y. Liu, and R. Schober, "TTD configurations for near-field beamforming: Parallel, serial, or hybrid?" *IEEE Trans. Commun.*, vol. 72, no. 6, pp. 3783–3799, Jun. 2024.

[19] P.-H. Chang and T.-D. Chiueh, "Hybrid beamforming for wideband terahertz massive MIMO communications with low-resolution phase shifters and true-time-delay," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 8000–8012, Jul. 2024.

[20] O. El. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Jan. 2014.

[21] B. Zhai, Y. Zhu, A. Tang, and X. Wang, "THzPrism: Frequency-based beam spreading for terahertz communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 897–900, Jun. 2020.

[22] H. He, M. Zhang, S. Jin, C.-K. Wen, and G. Y. Li, "Model-driven deep learning for massive MU-MIMO with finite-alphabet precoding," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2216–2220, Oct. 2020.

[23] C.-J. Wang, C.-K. Wen, S. Jin, and S.-H. Tsai, "Finite-alphabet precoding for massive MU-MIMO with low-resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4706–4720, Jul. 2018.

[24] T. Peken, S. Adiga, R. Tandon, and T. Bose, "Deep learning for SVD and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6621–6642, Oct. 2020.

[25] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 552–563, Jan. 2020.

[26] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1677–1688, Mar. 2020.

[27] T. X. Vu et al., "Machine learning-enabled joint antenna selection and precoding design: From offline complexity to online performance," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3710–3722, Jun. 2021.

[28] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2018.

[29] T. Lin and Y. Zhu, "Beamforming design for large-scale antenna arrays using deep learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 103–107, Jan. 2020.

[30] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Unsupervised deep learning for massive MIMO hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7086–7099, Nov. 2021.

[31] Z. Liu, Y. Yang, F. Gao, T. Zhou, and H. Ma, "Deep unsupervised learning for joint antenna selection and hybrid beamforming," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1697–1710, Mar. 2022.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[33] X. Zhao, U. Shah, O. Glubokov, and J. Oberhammer, "Micromachined subterahertz waveguide-integrated phase shifter utilizing supermode propagation," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 7, pp. 3219–3227, Jul. 2021.

[34] J.-C. Jeong, I.-B. Yom, J.-D. Kim, W.-Y. Lee, and C.-H. Lee, "A 6–18-GHz GaAs multifunction chip with 8-bit true time delay and 7-bit amplitude control," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 5, pp. 2220–2230, May 2018.

[35] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Athens, Greece: Springer, Aug. 2016, pp. 424–432.

[36] O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, and L. Soler, "U-Net transformer: Self and cross attention for medical image segmentation," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Strasbourg, France: Springer, Sep. 2021, pp. 267–276.

[37] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Latent correlation representation learning for brain tumor segmentation with missing MRI modalities," *IEEE Trans. Image Process.*, vol. 30, pp. 4263–4274, 2021.

[38] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Dec. 2017, pp. 5998–6008.

[39] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1999–2049, 2023.

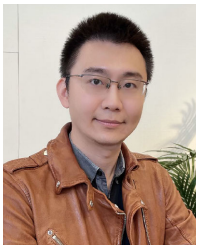
[40] Y. Wu, G. Song, H. Liu, L. Xiao, and T. Jiang, "3-D hybrid beamforming for terahertz broadband communication system with beam squint," *IEEE Trans. Broadcast.*, vol. 69, no. 1, pp. 264–275, Mar. 2023.

[41] R. L. Graham, D. E. Knuth, O. Patashnik, and S. Liu, *Concrete Mathematics*. Reading, MA, USA: Addison-Wesley, 1989.

[42] I. E. Gordon et al., "The HITRAN2016 molecular spectroscopic database," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 203, pp. 3–69, Dec. 2017.
[43] C. Xiang, M. L. Davenport, J. B. Khurgin, P. A. Morton, and J. E. Bowers, "Low-loss continuously tunable optical true time delay based on Si_3N_4 ring resonators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 4, pp. 1–9, Jul. 2018.
[44] Z.-X. Du, X. Y. Zhang, K.-X. Wang, H.-L. Kao, X.-L. Zhao, and X. H. Li, "Unequal Wilkinson power divider with reduced arm length for size miniaturization," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 6, no. 2, pp. 282–289, Feb. 2016.

Hsienchih Ting (Graduate Student Member, IEEE) received the B.Eng. degree from Beijing University of Posts and Telecommunications, China, in 2020, and the M.Sc. degree from Shanghai Jiao Tong University, China, in 2023. He is currently pursuing the Ph.D. degree with the Queen Mary University of London. His research interests include near-field communications, mobile edge generations, and machine learning.

Zhaolin Wang (Member, IEEE) received the first B.Eng. degree from Beijing University of Posts and Telecommunications, China, in 2020, the second B.Eng. degree (Hons.) from the Queen Mary University of London, U.K., in 2020, and the M.Sc. degree (Hons.) from Imperial College London, U.K., in 2021, and the Ph.D. degree from the Queen Mary University of London, in 2024. He is currently a Post-Doctoral Researcher with the Queen Mary University of London. His research interests include near-field communications, integrated sensing and communications, reconfigurable intelligent surfaces, and optimization theory. He was a recipient of the Best Student Paper Award in IEEE VTC2022-Fall and the 2023 IEEE Daniel E. Noble Fellowship Award. More information can be found at https://zhaolin820.github.io.



Yuanwei Liu (Fellow, IEEE) received the Ph.D. degree from the Queen Mary University of London (QMUL), London, U.K., in 2016.

He has been a (tenured) Full Professor with the Department of Electrical and Electronic Engineering (EEE), The University of Hong Kong (HKU), and a Visiting Professor with QMUL since September 2024. Prior to that, he was a Senior Lecturer (an Associate Professor) (2021–2024) and a Lecturer (an Assistant Professor) (2017–2021) with QMUL, and a Post-Doctoral Research Fellow with King's College London (KCL), London, from 2016 to 2017. His research interests include non-orthogonal multiple access, reconfigurable intelligent surface, near-field communications, integrated sensing and communications, and machine learning. He is a fellow of AAIA; a Web of Science Highly Cited Researcher; an IEEE Communication Society Distinguished Lecturer; an IEEE Vehicular Technology Society Distinguished Lecturer; the Rapporteur of ETSI Industry Specification Group on Reconfigurable Intelligent Surfaces on work item of "Multi-functional Reconfigurable Intelligent Surfaces (RIS): Modelling, Optimisation, and Operation;" and the U.K. Representative of the URSI Commission C on "Radio communication Systems and Signal Processing." He was listed as one of 35 Innovators Under 35 China in 2022 by MIT Technology Review. He received the IEEE ComSoc Outstanding Young Researcher Award for EMEA in 2020. He received the 2020 IEEE Signal Processing and Computing for Communications (SPCC) Technical Committee Early Achievement Award and the IEEE Communication Theory Technical Committee (CTTC) 2021 Early Achievement Award. He received the IEEE ComSoc Outstanding Nominee for Best Young Professionals Award in 2021. He was a co-recipient of the 2024 IEEE Communications Society Heinrich Hertz Award, the Best Student Paper Award in IEEE VTC2022-Fall, the Best Paper Award in ISWCS 2022, the 2022 IEEE SPCC-TC Best Paper Award, the 2023 IEEE ICCT Best Paper Award, and the 2023 IEEE ISAP Best Emerging Technologies Paper Award. He serves as the Publicity Co-Chair for IEEE VTC 2019-Fall; the Panel Co-Chair for IEEE WCNC 2024; the Symposium Co-Chair for several flagship conferences, such as IEEE GLOBECOM, ICC, and VTC. He serves the Academic Chair for the Next Generation Multiple Access Emerging Technology Initiative and the Vice Chair of SPCC and Technical Committee on Cognitive Networks (TCCN). He serves as the Co-Editor-in-Chief of the IEEE ComSoc TC Newsletter; an Area Editor of IEEE COMMUNICATIONS LETTERS; an Editor of IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and IEEE TRANSACTIONS ON COMMUNICATIONS from 2018 to 2023. He serves as the (leading) Guest Editor for PROCEEDINGS OF THE IEEE on Next Generation Multiple Access, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Next Generation Multiple Access, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING on Intelligent Signal Processing and Learning for Next Generation Multiple Access, and *IEEE Network* on Next Generation Multiple Access for 6G. For more information visit the link https:// www.eee.hku.hk/yuanwei/.