Intelligent Computing
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Deep Learning–Based Channel Extrapolation and Multiuser Beamforming for RIS-aided Terahertz Massive MIMO Systems over Hybrid-Field Channels

Yang Wang[1], Zhen Gao[1,2,3*], Sheng Chen[4,5], Chun Hu[1], and Dezhi Zheng[1]

[1]MIIT Key Laboratory of Complex-Field Intelligent Sensing, Beijing Institute of Technology, Beijing, China. [2]Yangtze Delta Region Academy of Beijing Institute of Technology, Beijing Institute of Technology, Jiaxing, China. [3]Advanced Research Institute of Multidisciplinary Science, Beijing Institute of Technology, Jinan, China. [4]School of Electronics and Computer Science, University of Southampton, Southampton, UK. [5]Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China.

*Address correspondence to: gaozhen16@bit.edu.cn

Citation: Wang Y, Gao Z, Chen S, Hu C, Zheng D. Deep Learning–Based Channel Extrapolation and Multiuser Beamforming for RIS-aided Terahertz Massive MIMO Systems over Hybrid-Field Channels. *Intell. Comput.* 2024;3:Article 0065. https://doi.org/10.34133/icomputing.0065

Submitted 7 February 2023
Accepted 10 October 2023
Published 13 March 2024

The reconfigurable intelligent surface (RIS) is a promising technology for terahertz (THz) massive multiple-input multiple-output (MIMO) communication systems. However, acquiring high-dimensional channel state information (CSI) and realizing efficient active/passive beamforming for RIS are challenging owing to its cascaded channel structure and lack of signal processing units. To overcome these challenges, this study proposes a deep learning (DL)-based physical signal processing scheme for RIS-aided THz massive MIMO systems over hybrid far-near field channels wherein channel estimation with low pilot overhead and robust beamforming are implemented. Specifically, first, an end-to-end DL-based channel estimation framework that consists of pilot design, CSI feedback, subchannel estimation, and channel extrapolation is introduced. In this framework, only some RIS elements are first activated, a subsampling RIS channel is then estimated, and a DL-based extrapolation network is finally used to reconstruct the full-dimensional CSI. Next, to maximize the sum rate under imperfect CSI, a DL-based scheme is developed to simultaneously design hybrid active beamforming at the base station and passive beamforming at the RIS. Simulation results show that the proposed channel extrapolation scheme achieves better CSI reconstruction performance than conventional schemes while greatly reducing pilot overhead. Moreover, the proposed beamforming scheme outperforms conventional schemes in terms of robustness to imperfect CSI.

Introduction

The surge in demand for wireless data traffic in recent years owing to the exponential growth of Internet-of-Things devices and broadband multimedia applications has spurred the exploration of terahertz (THz) communications as a viable solution [1]. However, extremely high free-space losses and strong atmospheric attenuation in the THz band pose a challenge to the long-range coverage of THz communication systems. To overcome this problem, the massive or ultramassive multiple-input multiple-output (MIMO) technique has been considered to achieve high array gain and mitigate the high propagation loss [2]. Conventional massive MIMO systems require a dedicated radio frequency (RF) chain for each antenna (i.e., a fully digital architecture) and thus suffer from extremely high power consumption and hardware costs. To circumvent this technical issue, hybrid analog–digital massive MIMO architectures have been widely adopted to reduce the number of RF chains while ensuring high array gains [3].

In addition, the reconfigurable intelligent surface (RIS) has garnered attention as a potentially transformative technology for improving communication performance [4–9]. By manipulating the phase and amplitude of RIS phase shifters, a RIS passively reflects incident electromagnetic (EM) signals toward desired directions and provides considerable beamforming gain. More importantly, a RIS does not require power-intensive RF chains, which contributes to a more environmentally friendly and cost-effective communication solution. Therefore, the integration of RIS and massive MIMO techniques holds promise for overcoming the limitations of THz communications and realizing its full potential.

Generally, a simplified planar-wave channel model is appropriate if the user equipment (UE) operates in the far field of the base station (BS). However, given that severe path losses reduce the effective coverage and that enlarging the array in the THz band increases the Rayleigh distance [10], both the far and the near field need to be considered for THz massive MIMO systems. Therefore, the distance from each antenna of the BS to the UE needs to be considered by the spherical-wave channel model under near-field conditions [11]. However, the number of spherical-wave channel parameters is proportional to the number of massive antennas, which makes the direct adoption of the spherical-wave channel model in THz massive MIMO systems unrealistic. To this end, a hybrid-field (hybrid spherical- and planar-wave) channel model characterized by a smaller number of parameters and high accuracy has been proposed for THz massive MIMO systems [12]. In this approach, the EM signal is modeled as a spherical wave for the inter-subarray and as a planar wave for the intra-subarray, based on different subarray architectures. Although the application of RISs has been widely investigated recently [13–17], their use for THz massive MIMO communications over hybrid-field channels is still at an early research stage.

Related work

Acquiring accurate channel state information (CSI) is critical for RIS-aided communication systems [18–21]. However, the accurate estimation of high-dimensional CSI with limited pilot signals remains a formidable challenge [13]. To address this challenge, compressive sensing-based solutions have been proposed to reduce the pilot overhead by leveraging channel sparsity [14,15]. However, these solutions present challenges in terms of computational complexity and storage requirements owing to the need for matrix inversion and iterative operations. Recently, the integration of deep learning (DL) in communication systems has garnered extensive attention. For instance, in [22], the authors proposed an effective pilot reduction technique by gradually pruning less important neurons from dense layers during training. In [16], the authors designed a DL-based channel estimation network to acquire RIS-aided and non-RIS-aided channels. In [17], a semipassive RIS architecture was proposed in which the orthogonal match pursuit (OMP) algorithm and a denoising convolutional neural network (CNN) are applied to reconstruct the CSI. However, the deployment of RF chains negates the key benefits of the RIS, i.e., reduction of hardware cost and power consumption.

In fact, owing to the highly dense arrangement of RIS elements [23], there is a strong correlation between the different elements of the CSI matrix, which makes it possible to extrapolate the complete channel from a partial one, i.e., to perform channel extrapolation [24]. Recently, some initial attempts to utilize channel extrapolation for further reducing the pilot overhead were reported. In [25], the authors proposed a DL-based extrapolation network to extrapolate the complete CSI by exploiting the correlation of the antenna domain; some antennas are activated by a selection network. In [26], the authors utilized a neural network structure modified by ordinary differential equations to improve the performance of extrapolation. In addition, in [27] the authors adopted a grouping strategy to reduce the dimension of the estimated channel and designed a CNN-based network to extrapolate the fully dimensional cascaded channel as well as eliminate the grouping interference. However, the aforementioned extrapolation schemes only consider the extrapolation process from the known subchannels; the estimation of the subchannel is ignored. Moreover, hybrid-field channel modeling of RISs exhibits more complex EM wave propagation characteristics, which hinder subchannel acquisition and the subsequent extrapolation of complete channels.

The proper and effective design of the hybrid beamforming and RIS phase according to the CSI is one of the major engineering challenges in the design of RIS-aided communication systems. Recently, some studies have investigated hybrid beamforming and RIS design problems [28–30]. In [28], simultaneous orthogonal matching pursuit (SOMP)-based hybrid beamforming was proposed for RIS-aided mmWave MIMO systems. In [29], an iteration-based jointly active/passive beamforming algorithm was designed to maximize the sum rate of systems. Furthermore, DL-based beamforming methods have also been studied in RIS-aided wireless communication systems. In [30], a deep neural network-based beamforming approach was developed to jointly optimize the transmit/reflect beamforming vectors for achieving data rate maximization. However, further analysis of the aforementioned schemes in terms of adaptability is necessary, given that the current approach only considers the idealized CSI assumption.

Motivations

Current research on RISs has primarily centered on the development of 2 modes of operation, namely reflective [28,29,31] and transmissive [32–34]. A number of studies have been conducted on RIS-aided communication in reflective mode, which is primarily utilized to address the blind coverage problem. By contrast, the main purpose of transmissive RISs is to improve the spectral efficiency of networks, given that the transmissive mode does not alter the direction of EM waves. Therefore, the deployment of transmissive RIS is suitable when a line-of-sight (LoS) path exists but the propagation attenuation is high, e.g., when an outdoor BS serves indoor UEs, to improve the energy of the received signals. In view of this, transmissive RISs have the potential to enhance indoor signal service.

Considering the hybrid-field channel model, in [35] the authors presented a 2-stage channel estimation mechanism for which a CNN-based network was designed to estimate the channel parameters and the complete channel was reconstructed by channel extrapolation based on the geometric relationships of the channel parameters. However, this parametric-based extrapolation method requires a large number of training labels containing accurate channel parameters. In [36], the authors proposed a sensor-assisted channel estimation and beamforming technique in which a LoS MIMO architecture is considered in the hybrid field. However, the channel estimation in [36] relies heavily on the awareness of sensors, which becomes challenging when it comes to obtaining accurate CSI. Therefore, similar to [25–27], we propose a DL-based channel extrapolation method to address the performance limitations of conventional channel estimation methods for indoor hybrid-field propagation environments. In addition, in this study, we adopted the LoS MIMO architecture under the assumption of the hybrid-field channel model, where the LoS MIMO architecture can support multistream transmission in the pure LoS BS-RIS channel.

Most existing studies in the field of RIS-aided communication systems were based on the assumption that BS-RIS and RIS-UE CSIs are perfect [28–31]. However, this assumption is impractical. Channel estimation error should be considered when designing these systems. Recently, imperfect CSI conditions have been considered in some studies [37,38]. For instance, in [37] the authors utilized a penalty-based alternating algorithm to jointly design the active beamforming and RIS phase under the presence of imperfect CSI. Similarly, in [38] the authors exploited a gradient projection-based alternating optimization algorithm to jointly design the active beamforming, RIS placement, and RIS phase under imperfect CSI. While there are numerous DL-based methods available for RIS-aided communication systems with perfect CSI, there are only a few DL-based methods that consider imperfect CSI [39]. In this context, the present study provides a DL-based hybrid beamforming and

RIS phase design (HBFRPD) solution that incorporates imperfect CSI in RIS-aided communication systems.

Contributions

This paper presents a DL-based spatial-frequency domain channel extrapolation (SFDCExtra) network and DL-based HBFRPD scheme for RIS-aided downlink multiuser THz massive MIMO systems over hybrid-field channels. The main contributions of this study are summarized next.

- We propose the deployment of a transmissive RIS on a window to reduce the penetration loss and thus achieve enhanced indoor communication. In addition, owing to the negligible non-LoS (NLoS) component energy in the THz band, the BS-RIS channel is dominated by the LoS path. To achieve multistream transmission in the LoS case, we consider a LoS MIMO architecture under hybrid-field channel modeling in which the BS and RIS adopt the same subarray structures, and the subarray spacing is optimized to satisfy the LoS MIMO condition.
- Given that the BS and RIS are fixed, and only one LoS path exists, the BS-RIS channel can be considered to be quasi-static and known. In contrast, owing to the mobility of the UE, the RIS-UE channel is time-varying. Therefore, the focus is set on estimating the RIS-UE channel, which considerably reduces the pilot overhead.
- To further reduce the pilot overhead when estimating the RIS-UE channel, we propose a DL-based channel extrapolation scheme in which the RIS only activates some of its elements at the channel estimation stage. Unlike existing extrapolation schemes [25–27], which only focus on the CSI extrapolation process, the complete channel extrapolation framework that we designed includes a pilot design network (PDN), a CSI feedback network, a subchannel estimation network, and a channel extrapolation network. By adopting an end-to-end (E2E) training strategy, the proposed channel estimation scheme can maintain high reconstruction performance with a reduced pilot overhead. Specifically, by using the CSI feedback network, the UE side feeds the quantized pilot information back to the BS, which estimates the subsampled RIS-UE channel and then extrapolates the complete RIS-UE channel using the channel extrapolation network. In addition, for RIS element selection, we analyze the impact of 3 different strategies, namely uniform, random, and learning-based selection, on the final channel estimation performance.
- To solve the multiuser interference problem under imperfect CSI, we propose a DL-based HBFRPD scheme that consists of an analog beamformer design, an DL-based RIS phase design network (RPDN), and an knowledge-data dual-driven digital beamforming network. By maximizing the sum rate with E2E training, the proposed scheme achieves better performance and robustness than the existing state-of-the-art methods.

Notations: In this paper, scalars are denoted as lowercase letters, vectors are denoted as lowercase boldface letters, and matrices are denoted as uppercase boldface letters. The conjugate, transpose, conjugate transpose, inversion, and Moore–Penrose inversion operators are denoted as superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, and $(\cdot)^\dagger$, respectively. The diagonalization, block diagonalization, Kronecker product, and Hadamard product are represented by operators diag$(\cdot)$, blkdiag$(\cdot)$, $\otimes$, and $\odot$, respectively. The Frobenius norm of $\mathbf{A}$ is denoted as $|\mathbf{A}|_F$. The identity matrix with size $n \times n$ is represented by $\mathbf{I}_n$, while the column vector of size n with all elements equal to 1 (0) is represented by $\mathbf{1}_n$ $(\mathbf{0}_n)$. The real and imaginary parts of the corresponding argument are denoted as $\Re\{\cdot\}$ and $\Im\{\cdot\}$, respectively. The m-th row and n-th column element of $\mathbf{A}$ is represented by $\{\mathbf{A}\}_{m,n}$, and the m-th entry of $\mathbf{a}$ is represented by $\{\mathbf{a}\}_m$. The submatrix containing the m-th to n-th columns of $\mathbf{A}$ is represented by $\mathbf{A}_{[:,m:n]}$. The expectation operator is represented by $\mathbb{E}(\cdot)$, and the real (complex) Gaussian distribution with mean μ and variance σ^2 is denoted as $\mathcal{N}(\mu,\sigma^2)$ $(\mathcal{CN}(\mu,\sigma^2))$, where the matrix trace operator is represented by $Tr\{\cdot\}$.

Materials and Methods

System model

System description

As shown in Fig. 1, we consider a downlink RIS-aided MIMO orthogonal frequency division multiplexing (OFDM) transmission system in an indoor environment, where a transparent RIS is attached to a window surface to refract outdoor THz signals from the BS into the room to serve U single-antenna UEs. Thus, the transparent transmissive RIS helps enhance indoor coverage. Let the BS (RIS) have $M^\mathrm{B} = M_y^\mathrm{B} \times M_z^\mathrm{B}$ $(M^\mathrm{R} = M_y^\mathrm{R} \times M_z^\mathrm{R})$ uniformly spaced subarrays, where M_y^B (M_y^R) and M_z^B (M_z^R) are the numbers of BS (RIS)-side subarrays along the horizontal and vertical directions, respectively. Each subarray of the BS (RIS) is a uniform planar array (UPA) with $N_\mathrm{sub}^\mathrm{B} = N_y^\mathrm{B} \times N_z^\mathrm{B}$ $(N_\mathrm{sub}^\mathrm{R} = N_y^\mathrm{R} \times N_z^\mathrm{R})$ isotropically radiating elements, where N_y^B (N_y^R) and N_z^B (N_z^R) are the numbers of BS (RIS)-side subarray antennas along the horizontal and vertical directions, respectively. Therefore, the complete antenna dimension of the BS is $N^\mathrm{B} = M^\mathrm{B} N_\mathrm{sub}^\mathrm{B}$, and the element dimension of the RIS is $N^\mathrm{R} = M^\mathrm{R} N_\mathrm{sub}^\mathrm{R}$. To simplify the analysis, we assume that the normals of the central elements of both the BS and RIS are coaxial, i.e., the BS array and RIS array are parallel to each other with a distance of D, as illustrated in Fig. 1B.

In this study, we considered a BS-side subconnected hybrid analog–digital array architecture. This architecture consists of M^B RF chains capable of supporting $U \leq M^\mathrm{B}$ data streams. Each of these RF chains is connected to a subarray through $N_\mathrm{sub}^\mathrm{B}$ phase shifters. Furthermore, we set the number of subcarriers to K and the sampling frequency (i.e., bandwidth) to f_s. The carrier frequency is denoted as f_c, which corresponds to central wavelength λ.

Channel model

BS-RIS channel model

Owing to the negligible NLoS component energy in the THz band, we only consider the LoS path in the analysis of the BS-RIS channel. Assuming spherical wave propagation, we construct the LoS MIMO link between the BS and RIS with only one single LoS path, although it can support intrapath multiplexing for multistream transmission [40]. The interantenna spacing in each subarray is $d = \lambda/2$. To satisfy the LoS MIMO characteristic, the BS subarray spacing given by d_{sy}^B and d_{sz}^B is set to the following optimal LoS MIMO spacing:

$$d_{sy}^\mathrm{B} = \sqrt{\frac{\lambda D}{M_y^\mathrm{B}}} - \frac{\lambda}{2}\left(N_y^\mathrm{B} - 1\right), d_{sz}^\mathrm{B} = \sqrt{\frac{\lambda D}{M_z^\mathrm{B}}} - \frac{\lambda}{2}\left(N_z^\mathrm{B} - 1\right), \tag{1}$$

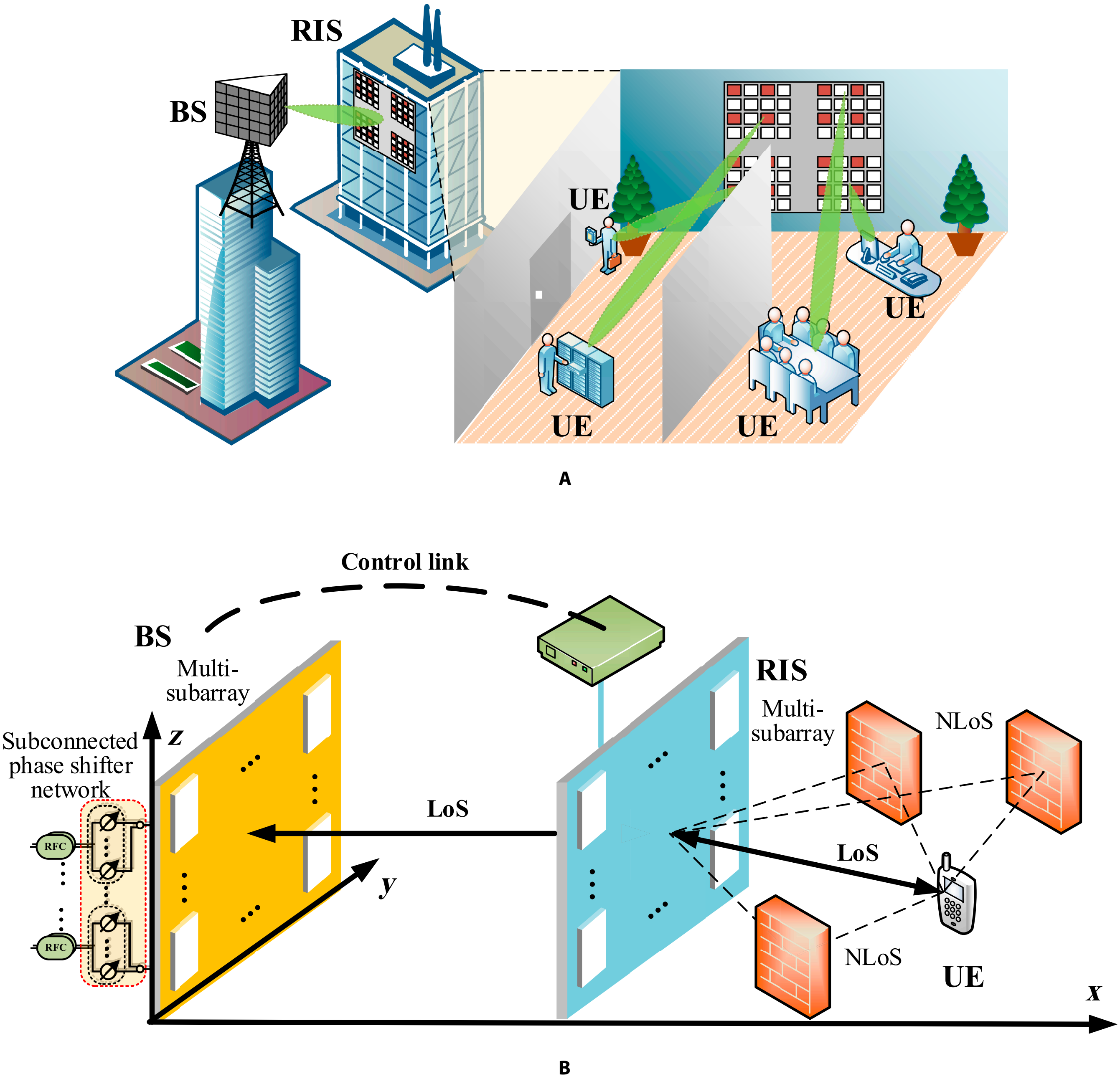

Fig. 1. Schematic diagram of a RIS-aided THz massive MIMO system. (A) Multiple indoor UEs are served by the BS with the help of a transmissive RIS deployed on a window, and (B) hardware architectures at the BS, RIS, and UEs.

i.e., d_{sy}^{B} and d_{sz}^{B} should satisfy the condition $\lambda \ll d_{sy}^{\mathrm{B}}, d_{sz}^{\mathrm{B}} \ll D$. A detailed explanation of Eq. 1 can be found in [40,41]. The RIS subarray spacing expressed by d_{sy}^{R} and d_{sz}^{R} can be obtained using a similar definition. Note that self-orthogonal LoS MIMO not only is obtained from parallel symmetric antenna arrangements but also can be obtained with symmetrical/unsymmetrical arrangements on tilted nonparallel lines/planes [41]. The following proposition was extracted from [42].

Proposition 1. Let the transceiver arrays be placed with a separation distance of D and be working at a carrier wavelength λ ($\lambda \ll D$). If the interantenna spacing and carrier wavelength λ are set in the same order of magnitude, the planar wave model can be applied. Otherwise, the spherical wave model should be employed.

According to Proposition 1, the subarray response vectors $\mathbf{a}(\theta, \phi, f_k) \in \mathbb{C}^{N_{\mathrm{H}} N_{\mathrm{V}} \times 1}$ can be approximated by a planar wave model:

$$\begin{aligned}\mathbf{a}(\theta,\phi,f_k) &= \mathbf{a}_h(\theta,\phi,f_k) \otimes \mathbf{a}_v(\phi,f_k) \\ &= \left[1, \ldots, e^{-j2\pi}\frac{f_k}{c}d\left(n_h \sin\theta\cos\phi + n_v \sin\phi\right), \ldots, e^{-j2\pi}\frac{f_k}{c}d((N_H-1)\sin\theta\cos\phi + (N_V-1)\sin\theta)\right]^{\mathrm{T}},\end{aligned} \tag{2}$$

where $f_k = f_c - \frac{f_s}{2} + \frac{kf_s}{K}$, $1 \leq k \leq K$, is the k-th subcarrier frequency, c is the speed of light, $0 \leq n_h \leq (N_{\mathrm{H}} - 1)$, $0 \leq n_v \leq (N_{\mathrm{V}} - 1)$, N_{H}, and N_{V} are the numbers of horizontal and vertical antennas, respectively, while θ and ϕ are the horizontal and vertical angles of the departure or arrival (AoD or AoA) of the path, respectively.

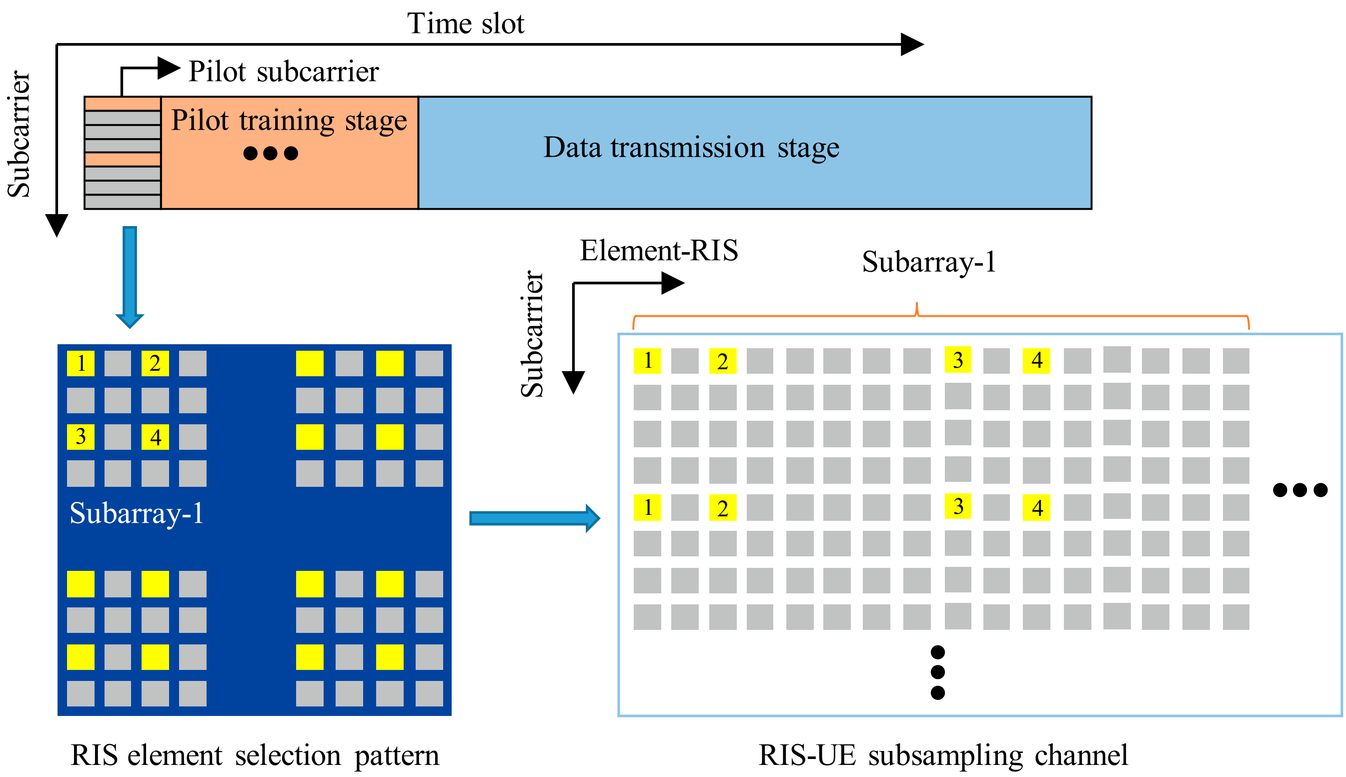

Fig. 2. Block diagram of the frame structure, RIS element selection pattern, and RIS-UE subsampling channel. The selected parts are marked as yellow blocks, and the number in the yellow block is the index of the selected element.

Given that $d_{sy}^{\mathrm{B}}, d_{sz}^{\mathrm{B}}, d_{sy}^{\mathrm{R}}, d_{sz}^{\mathrm{R}} \ll D$, the direction difference of the same path in different subarrays is negligible. Therefore, all subarrays on either the BS or RIS side can be assumed to share identical array response vectors. However, as subarrays are widely spaced, the relative phase differences among subarrays are non-negligible [42]. Motivated by the above analysis, the downlink spatial-frequency BS-RIS channel $\mathbf{G}[k] \in \mathbb{C}^{N^{\mathrm{R}} \times N^{\mathrm{B}}}$ on the k-th subcarrier can be modeled as

$$\mathbf{G}[k] = \alpha[k] G_T \tilde{\mathbf{G}}[k] \otimes \left[\mathbf{a}_{\mathrm{R}}\left(\theta_{\mathrm{R,A}}, \phi_{\mathrm{R,A}}, f_k\right) \mathbf{a}_B^H\left(\theta_{\mathrm{B}}, \phi_{\mathrm{B}}, f_k\right)\right], \quad (3)$$

where $\alpha[k]$ is the channel attenuation coefficient on the k-th subcarrier, and $\left(\theta_{\mathrm{B}}, \phi_{\mathrm{B}}\right)$ and $\left(\theta_{\mathrm{R,A}}, \phi_{\mathrm{R,A}}\right)$ are AoD and AoA of the LoS path, respectively. Without loss of generality, we assume that the LoS angles are fixed and known in advance given that the BS and RIS are fixed. In Eq. 3, the entries of $\tilde{\mathbf{G}}[k] \in \mathbb{C}^{M^{\mathrm{R}} \times M^{\mathrm{B}}}$ are defined according to the spherical wave model as

$$\left\{\tilde{\mathbf{G}}[k]\right\}_{m_r, m_b} = e^{-j2\pi f_k \cdot \frac{D^{(m_r, m_b)}}{c}}, \quad (4)$$

where $D^{(m_r, m_b)}$ represents the distance between the m_r-th RIS-side subarray and m_b-th BS-side subarray. Furthermore, the subarray response vectors $\mathbf{a}_{\mathrm{R}}\left(\theta_{\mathrm{R,A}}, \phi_{\mathrm{R,A}}, f_k\right) \in \mathbb{C}^{N_{\mathrm{sub}}^{\mathrm{R}} \times 1}$ and $\mathbf{a}_{\mathrm{B}}\left(\theta_{\mathrm{B}}, \phi_{\mathrm{B}}, f_k\right) \in \mathbb{C}^{N_{\mathrm{sub}}^{\mathrm{B}} \times 1}$ are defined in Eq. 2. The constant coefficient G_{T} represents the antenna gain at the BS, which is different from the array gain generated by beamforming [43]. The only unknown parameter in Eq. 3 is the channel coefficient $\alpha[k]$, which can be obtained by placing a power detector at the RIS side. Therefore, it is reasonable to assume that the quasi-static BS-RIS channel is known.

RIS-UE channel model

As illustrated in Fig. 1B, we consider a multipath THz channel model for indoor environments [44]. The indoor RIS-UE channel model consists of one LoS path and L_p NLoS paths whose 3-dimensional (3D) distances are represented as d_0 and d_l, for $1 \leq l \leq L_p$, respectively [45]. The total EM wave propagation loss mainly consists of 2 parts: free-space path loss $\beta_{\mathrm{spr}}\left(f_k, d_l\right) = \frac{c}{4\pi f_k d_l}$ and molecular absorption loss $\beta_{\mathrm{abs}}\left(f_k, d_l\right) = e^{-\frac{1}{2}\kappa(f_k) d_l}$, where $\kappa\left(f_k\right)$ denotes the frequency-dependent absorption coefficient [46]. Hence, the spatial-frequency channel $\mathbf{h}[k] \in \mathbb{C}^{1 \times N^{\mathrm{R}}}$ for the RIS-UE link is

$$\mathbf{h}[\mathrm{k}] = \beta[k] \tilde{\mathbf{h}}_{\mathrm{LoS}}[k] \otimes \mathbf{a}_{\mathrm{R}}^H\left(\theta_{\mathrm{R,D}}^{\mathrm{LoS}}, \phi_{\mathrm{R,D}}^{\mathrm{LoS}}, f_k\right) + \frac{1}{\sqrt{L_p}} \sum_{l=1}^{Lp} \beta_l[k] \tilde{\mathbf{h}}^l[k] \otimes \mathbf{a}_{\mathrm{R}}^H\left(\theta_{R,D}^l, \phi_{R,D}^l, f_k\right), \quad (5)$$

where $\beta[k] = \beta_{\mathrm{spr}}\left(f_k, d_0\right) \beta_{\mathrm{abs}}\left(f_k, d_0\right)$ and $\beta_l[k] = \beta_{\mathrm{spr}}\left(f_k, d_l\right) \beta_{\mathrm{abs}}\left(f_k, d_l\right) \beta_{\mathrm{RC}}$ are the channel attenuation coefficients of the LoS and l-th NLoS paths, respectively, and $\left(\theta_{\mathrm{R,D}}^{\mathrm{LoS}}, \phi_{\mathrm{R,D}}^{\mathrm{LoS}}\right)$ and $\left(\theta_{\mathrm{R,D}}^l, \phi_{\mathrm{R,D}}^l\right)$ are the LoS AoD and NLoS AoD of the l-th NLoS path, respectively. Additionally, the reflection coefficient β_{RC} is a Gaussian random variable, i.e., $10 log \beta_{\mathrm{RC}}[\mathrm{dB}] \sim min \left\{\mathcal{N}\left(\mu_{\mathrm{R}} \sigma_{\mathrm{R}}^2\right), 0\right\}$. The entries of $\tilde{\mathbf{h}}_{\mathrm{LoS}}[k] \in \mathbb{C}^{1 \times M^{\mathrm{R}}}$ are given as $\left\{\tilde{\mathbf{h}}_{\mathrm{LoS}}[k]\right\}_{m_r} = e^{-j2\pi f_k \cdot \frac{d^{(m_r)}}{c}}$, where $d^{(m_r)}$ denotes the 3D distance between the UE and m_r-th RIS-side subarray; $\tilde{\mathbf{h}}^l[k]$ has a similar notation and assumptions.

Problem formulation and proposed channel estimation solution

Problem formulation of channel estimation

In this subsection, the downlink channel estimation problem is formulated on the basis of the considered RIS-aided THz massive MIMO communication system over hybrid-field channels. As shown in Fig. 2, we consider a 2-stage frame structure consisting of the pilot training and data transmission stages. At the pilot training stage, the BS transmits M pilot OFDM symbols (i.e., M time slots) dedicated to channel estimation. The m-th received signal at the UE side on the k-th subcarrier is represented by

$$y_m[k] = \sqrt{P_{\mathrm{T}}}\mathbf{h}[k]\mathbf{\Phi}_m\mathbf{G}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\mathbf{s}_m[k] + n_m[k], \quad (6)$$

where $1 \le k \le K, 1 \le m \le M$, P_{T} is the transmit power of the BS, $\mathbf{s}_m[k] \in \mathbb{C}^{U\times 1}$ denotes the transmitted symbol vector with $\mathbb{E}\{\mathbf{s}_m[k]\mathbf{s}_m^H[k]\} = \mathbf{I}_U$, and $n_m[k] \sim \mathcal{CN}(0, \sigma_n^2)$ is the effective complex additive white Gaussian noise at the UE, while $\mathbf{h}[k] \in \mathbb{C}^{1\times N^{\mathrm{R}}}$ and $\mathbf{G}[k] \in \mathbb{C}^{N^{\mathrm{R}}\times N^{\mathrm{B}}}$ are the downlink RIS-UE and BS-RIS channels on the k-th subcarrier, respectively. Note that because each UE can perform channel estimation independently, UE subscripts are omitted. Let us denote the control vector $\mathbf{v}_{m_r,m} \in \mathbb{C}^{1\times N_{\mathrm{sub}}^{\mathrm{R}}}$ for the m_r-th subarray elements of the RIS in the m-th time slot as

$$\mathbf{v}_{m_r,m} = \mathbf{o}_{m_r,m} \odot \tilde{\mathbf{v}}_{m_r,m} = \\ \left[\cdots, \eta_{n_{\mathrm{sub}}^r,m_r,m}, \cdots\right] \odot \left[\cdots, e^{j\phi_{n_{\mathrm{sub}}^r,m_r,m}}, \cdots\right], \quad (7)$$

where $\mathbf{o}_{m_r,m} \in \mathbb{C}^{1\times N_{\mathrm{sub}}^{\mathrm{R}}}$ represents the amplitude control vector, $\tilde{\mathbf{v}}_{m_r,m} \in \mathbb{C}^{1\times N_{\mathrm{sub}}^{\mathrm{R}}}$ represents the phase control vector, and $1 \le n_{\mathrm{sub}}^r \le N_{\mathrm{sub}}^{\mathrm{R}}$, while $\eta_{n_{\mathrm{sub}}^r,m_r,m} \in [0, 1]$and $\phi_{n_{\mathrm{sub}}^r,m_r,m} \in [0, 2\pi]$ are the amplitude and phase control coefficients, respectively. Note that $\eta_{n_{\mathrm{sub}}^r,m_r,m}$ can control the switch of the refraction function for each RIS element. The complete set of RIS elements can be expressed as $\mathbf{v}_m = \mathbf{o}_m \odot \tilde{\mathbf{v}}_m = \left[\mathbf{v}_{1,m}, \cdots, \mathbf{v}_{m_r,m}, \cdots, \mathbf{v}_{M^{\mathrm{R}},m}\right]^T \in \mathbb{C}^{N^{\mathrm{R}}\times 1}$, where $\mathbf{o}_m = \left[\mathbf{o}_{1,m}, \cdots, \mathbf{o}_{M^{\mathrm{R}},m}\right]^T \in \mathbb{C}^{N^{\mathrm{R}}\times 1}$ and $\tilde{\mathbf{v}}_m = \left[\tilde{\mathbf{v}}_{1,m}, \cdots, \tilde{\mathbf{v}}_{M^{\mathrm{R}},m}\right]^T \in \mathbb{C}^{N^{\mathrm{R}}\times 1}$. Therefore, the refraction phase matrix of RIS is defined as $\mathbf{\Phi}_m = \mathrm{diag}(\mathbf{v}_m) = \mathbf{O}_m \odot \tilde{\mathbf{V}}_m \in \mathbb{C}^{N^{\mathrm{R}}\times N^{\mathrm{R}}}$, where $\mathbf{O}_m = \mathrm{diag}(\mathbf{o}_m) \in \mathbb{C}^{N^{\mathrm{R}}\times N^{\mathrm{R}}}$ is the RIS selection matrix and $\tilde{\mathbf{V}}_m = \mathrm{diag}(\tilde{\mathbf{v}}_m) \in \mathbb{C}^{N^{\mathrm{R}}\times N^{\mathrm{R}}}$ is the RIS phase matrix.

$\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N^{\mathrm{B}}\times M^{\mathrm{B}}}$ and $\mathbf{F}_{\mathrm{BB}}[k] \in \mathbb{C}^{M^{\mathrm{B}}\times U}$ are respectively analog and digital beamforming matrices used at the BS to provide array gain and eliminate multistream interference. According to the subconnected architecture, the analog beamformer implemented by phase shifters is expressed as

$$\mathbf{F}_{\mathrm{RF}} = \mathrm{blkdiag}\left(\mathbf{f}_1, \cdots, \mathbf{f}_{m_b}, \cdots, \mathbf{f}_{M^{\mathrm{B}}}\right), \quad (8)$$

where $\mathbf{f}_{m_b} = \left[\mathrm{f}_{m_b,1}, \cdots, \mathrm{f}_{m_b,n_{\mathrm{sub}}^b}, \cdots, \mathrm{f}_{m_b,N_{\mathrm{sub}}^{\mathrm{B}}}\right]^T \in \mathbb{C}^{N_{\mathrm{sub}}^{\mathrm{B}}\times 1}$ with $\left|\mathrm{f}_{m_b,n_{\mathrm{sub}}^b}\right|^2 = 1/N_{\mathrm{sub}}^{\mathrm{B}}$. Given that the BS-RIS channel having only a LoS path is quasi-static and known, each analog beamforming vector can be designed as

$$\mathbf{f}_{m_b} = \mathbf{a}_{\mathrm{B}}(\theta_{\mathrm{B}}, \phi_{\mathrm{B}}, f_k), 1 \le m_b \le M^{\mathrm{B}}, \quad (9)$$

where k can be set to $K/2$ for alleviating the beam squint problem induced by the large bandwidth [47]. The digital beamformer $\mathbf{F}_{\mathrm{BB}}[k]$ is designed according to the zero-forcing (ZF) precoding in order to eliminate the multistream interference between the BS and RIS subarrays, i.e.,

$$\mathbf{F}_{\mathrm{BB}}[k] = \zeta\tilde{\mathbf{G}}_{\mathrm{eq}}^{\dagger}[k] = \zeta\tilde{\mathbf{G}}_{\mathrm{eq}}^H[k]\left(\tilde{\mathbf{G}}_{\mathrm{eq}}[k]\tilde{\mathbf{G}}_{\mathrm{eq}}^H[k]\right)^{-1}, \quad (10)$$

where $\tilde{\mathbf{G}}_{\mathrm{eq}}[k] = \left[\alpha[k]G_{\mathrm{T}}\tilde{\mathbf{G}}[k] \otimes \mathbf{a}_{\mathrm{B}}^H(\theta_{\mathrm{B}}, \phi_{\mathrm{B}}, f_k)\right]\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{M^{\mathrm{R}}\times M^{\mathrm{B}}}$ is the equivalent BS-RIS channel obtained from the perspective of the first element of different subarrays at the RIS, and $\zeta = \sqrt{M^{\mathrm{B}}/Tr\left\{\tilde{\mathbf{G}}_{\mathrm{eq}}^{\dagger}[k]\left(\tilde{\mathbf{G}}_{\mathrm{eq}}^{\dagger}[k]\right)^H\right\}}$ is a constant to meet the total transmit power constraint after beamforming. Thus, the multistream interference between the BS and RIS subarrays can be eliminated, i.e., $\mathbf{G}_{\mathrm{eq}}[k] = \mathbf{G}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k] \in \mathbb{C}^{N^{\mathrm{R}}\times U}, \forall k$, is a block diagonal constant matrix.

Therefore, the equivalent pilot signal $\mathbf{p}_m \in \mathbb{C}^{N^{\mathrm{R}}\times 1}$ can be expressed as

$$\mathbf{p}_m = \underbrace{\left[\mathbf{O}_m \odot \tilde{\mathbf{V}}_m\right]}_{\mathbf{\Phi}_m}\underbrace{\left[\mathbf{G}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\right]}_{\mathbf{G}_{\mathrm{eq}}[k]}\mathbf{s}_m[k], \quad (11)$$

where $\mathbf{p}_m$ is identical for different subcarriers because we set the transmit symbol $\mathbf{s}_m[k]$ to be $\mathbf{1}_U, \forall m, k$, and the ZF digital beamformer in Eq. 10 for $\mathbf{G}[k]$. Under the assumption that the normals of the central elements of both the BS and RIS are coaxial, $\mathbf{G}_{\mathrm{eq}}[k]$ is defined by $\sqrt{N^B}\alpha[k]G_{\mathrm{T}}\mathrm{blkdiag}\left(\mathbf{1}_{N_{\mathrm{sub}}^{\mathrm{R}}}^1, \cdots, \mathbf{1}_{N_{\mathrm{sub}}^{\mathrm{R}}}^u, \cdots, \mathbf{1}_{N_{\mathrm{sub}}^{\mathrm{R}}}^U\right)$. Thus, the effective pilot signals can be further expressed as the RIS element vector given by $\mathbf{p}_m = \sqrt{N^B}\alpha[k]G_{\mathrm{T}}\mathbf{v}_m \approx \sqrt{N^B}\alpha G_{\mathrm{T}}\mathbf{v}_m = A_{\mathrm{T}}\mathbf{v}_m$, where the approximation $\alpha[k] \approx \alpha, \forall k$, is further applied and $A_{\mathrm{T}} = \sqrt{N^B}\alpha G_{\mathrm{T}}$ represents the total attenuation from the BS to the RIS.

After collecting continuous measurements of M time slots, the aggregate received signal vector $\mathbf{y}[k] = \left[y_1[k], \cdots, y_M[k]\right] \in \mathbb{C}^{1\times M}$ is expressed as

$$\mathbf{y}[k] = \sqrt{P_{\mathrm{T}}}\mathbf{h}[k]\mathbf{P} + \mathbf{n}[k], \quad (12)$$

where $\mathbf{P} = \left[\mathbf{p}_1, \cdots, \mathbf{p}_M\right] = A_{\mathrm{T}}\mathbf{V} = A_{\mathrm{T}}\left[\mathbf{v}_1, \cdots, \mathbf{v}_M\right] \in \mathbb{C}^{N^{\mathrm{R}}\times M}$, and $\mathbf{n}[k] = \left[n_1[k], \cdots, n_M[k]\right] \in \mathbb{C}^{1\times M}$. Thus, the received signal matrix $\mathbf{Y} = \left[\mathbf{y}^T[1], \cdots, \mathbf{y}^T[K]\right]^T \in \mathbb{C}^{K\times M}$ can be expressed as

$$\mathbf{Y} = \sqrt{P_{\mathrm{T}}}\mathbf{H}\mathbf{P} + \mathbf{N}, \quad (13)$$

where $\mathbf{H} = \left[\mathbf{h}^T[1], \cdots, \mathbf{h}^T[K]\right]^T \in \mathbb{C}^{K \times N^{\mathrm{R}}}$ represents the downlink spatial-frequency domain RIS-UE channel matrix, and $\mathbf{N} = \left[\mathbf{n}^T[1], \cdots, \mathbf{n}^T[K]\right]^T \in \mathbb{C}^{K \times M}$.

DL-based SFDCExtra

As shown in Fig. 2, we activate only $N_s^{\mathrm{R}} \le N^{\mathrm{R}}$ RIS elements at the pilot training stage and define $\rho \Delta = N^R / N_s^R \ge 1$ as the element compression ratio. Furthermore, only $K_s = \frac{K}{\overline{\rho}}$ uniformly selected subcarriers are used for pilot training, where $\overline{\rho}$ is the frequency compression ratio, and the remaining subcarriers can be used for transmitting control signals. Then, we estimate the subchannels associated with the activated RIS elements and selected subcarriers. We also provide an example of the RIS element pattern selected uniformly and the corresponding RIS-UE side subsampling spatial-frequency channel in Fig. 2, where the yellow blocks indicate the selected elements and selected subcarriers. Thus, the practical received pilot signal $\mathbf{Y}_s \in \mathbb{C}^{K_s \times M}$ is defined as

$$\mathbf{Y}_s = \sqrt{P_{\mathrm{T}}} \mathbf{H}_s \mathbf{P}_s + \mathbf{N}_s, \tag{14}$$

where $\mathbf{H}_s \in \mathbb{C}^{K_s \times N_s^{\mathrm{R}}}$ is the subsampling of the spatial-frequency channel, $\mathbf{P}_s \in \mathbb{C}^{N_s^{\mathrm{R}} \times M}$ is the corresponding equivalent pilot signal, and $\mathbf{N}_s$ is the noise. Our goal is to recover the complete channel $\hat{\mathbf{H}} \in \mathbb{C}^{K \times N^{\mathrm{R}}}$ using limited received pilot signals $\mathbf{Y}_s$, i.e., extrapolating the remaining unknown channels from the acquired partial channels. Based on the nonlinear function fitting capability of DL, a mapping can be learned to represent proximity correlations between different spatial/frequency locations of channels. Thus, we propose a DL-based SFDCExtra network that consists of the element selection strategy (ESS), pilot design, CSI feedback, subchannel estimation, and SFDCExtra modules, as illustrated in Fig. 3. The complete process of the proposed scheme can be expressed as

$$\hat{\mathbf{H}} = f_{\mathrm{SFDE}}\left(f_{\mathrm{SCE}}\left(f_{\mathrm{CsiFd}}\left(\sqrt{P_{\mathrm{T}}} f_{\mathrm{ESS}}(\mathbf{H}) \mathbf{P}_s + \mathbf{N}_s\right)\right)\right), \tag{15}$$

where the mapping $f_{\mathrm{ESS}}(\cdot)$ represents the element selection strategy for deciding the subsampling channel $\mathbf{H}_s$, the equivalent pilot signal P_s can be learned as trainable parameters, and $f_{\mathrm{CsiFd}}(\cdot)$, $f_{\mathrm{SCE}}(\cdot)$, and $f_{\mathrm{SFDE}}(\cdot)$ represent the CSI feedback network, subchannel estimation network, and spatial-frequency domain extrapolation network, respectively. We next detail each component.

Element selection strategy

With only N_s^{R} activated RIS elements, from Eq. 7, the RIS element selection vector $\mathbf{o} = \mathbf{o}_m = \left[\mathbf{o}_{1,m}, \cdots, \mathbf{o}_{m_r,m}, \cdots, \mathbf{o}_{M^{\mathrm{R}},m}\right]^T \in \{0,1\}^{N^{\mathrm{R}} \times 1}$ is an N_s^{R}-hot vector with N_s^{R} elements being "1" and the other elements being "0", where the subscript "m" can be dropped given that $\mathbf{o}$ is fixed at the pilot training stage. Moreover, given that only K_s subcarriers are uniformly selected for pilot training, the frequency selection vector $\boldsymbol{\kappa} \in \{0,1\}^{K \times 1}$ is defined by $\{\boldsymbol{\kappa}\}_{\overline{\rho}k+1} = 1, 0 \le k \le K_s - 1$, and the other elements are set to "0". Thus, the selection operation of the subsampling function $f_{\mathrm{ESS}}(\cdot)$ can be expressed as

$$\mathbf{H}_s = f_{\mathrm{ESS}}(\mathbf{H}) = \mathbf{S} \odot \mathbf{H}, \tag{16}$$

where $\mathbf{S} = \boldsymbol{\kappa} \otimes \mathbf{o}^T \in \{0,1\}^{K \times N^{\mathrm{R}}}$ is the spatial-frequency selection matrix, and the zero rows/columns in $\mathbf{S} \odot \mathbf{H}$ are deleted directly to yield $\mathbf{H}_s$. Note that different RIS element selection vectors can affect the extrapolation performance. Thus, we consider the 3 element selection strategies described next.

1. Uniform selection strategy: Given that each subarray in the RIS is a UPA, its element compression ratio is expressed as $\rho = \rho_y \times \rho_z$, where ρ_y and ρ_z are the compression ratios along the azimuth and elevation directions, respectively. To ensure balanced estimation performance along 2 directions, ρ_y and ρ_z are expected to be as close as possible. However, $\rho_y = \rho_z$ cannot always be guaranteed under all system parameter configurations. In cases where $\rho_y \ne \rho_z$, it is desirable to allocate more activated elements along the azimuth (y-axis) direction rather than the z-axis direction (i.e., $\rho_y \le \rho_z$). This strategic choice aligns with the consideration of indoor UEs, which are more likely to be distributed across a wide azimuth range, as opposed to the elevation range, given that indoor UEs are typically stationary in the vertical dimension. Accordingly, the y-z compression ratio allocation can be solved from the following optimization problem:

$$\begin{aligned} \underset{\{\rho_y, \rho_z\}}{min} \quad & \left|\rho_z - \rho_y\right|, \\ s.t. \quad & \rho_y \times \rho_z = \rho, \\ & 1 \le \rho_y \le \rho_z. \end{aligned} \tag{17}$$

Some allocation examples are $\rho(2,4,8,16) = \rho_y(1,2,2,4) \times \rho_z(2,2,4,4)$. Given ρ_y and ρ_z, the active element index vector $\boldsymbol{\xi}_{m_r} \in \mathbb{C}^{1 \times N_{\mathrm{sub}}^{\mathrm{R}}/\rho}$ of the m_r-th subarray can be expressed as

$$\left\{\boldsymbol{\xi}_{m_r}\right\}_{n_i^y N_z^{\mathrm{R}}/\rho_z + n_i^z + 1} = N_{\mathrm{sub}}^{\mathrm{R}}\left(m_r - 1\right) + N_z^{\mathrm{R}} \rho_y n_i^y + \rho_z n_i^z + 1, \tag{18}$$

where $1 \le m_r \le M^{\mathrm{R}}$, $0 \le n_i^y \le \frac{N_y^{\mathrm{R}}}{\rho_y} - 1$, and $0 \le n_i^z \le \frac{N_z^{\mathrm{R}}}{\rho_z} - 1$. The entire active element index vector or set of the RIS is defined as $\boldsymbol{\xi} = \left[\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{m_r}, \cdots, \boldsymbol{\xi}_{M^{\mathrm{R}}}\right]^T \in \mathbb{C}^{N_s^{\mathrm{R}} \times 1}$. Thus, we set the entries of the RIS element selection vector $\mathbf{o}$ corresponding to the index set $\boldsymbol{\xi}$ to "1", i.e., $\{\mathbf{o}\}_{\xi} = 1$ for $\xi \in \boldsymbol{\xi}$, and the other elements of $\mathbf{o}$ to "0".

2. Random selection strategy: It randomly selects N_s^{R} elements from the RIS as the random pattern and generates the active element index vector $\boldsymbol{\xi}$. If the element compression ratio ρ is not large, then the aperture of a random pattern is usually comparable to that of the RIS.

3. Learning-based selection strategy: In addition to the above 2 fixed selection strategies, the learning-based ESS has also been widely studied. In [25], a differentiable selection network was proposed to learn the element selection vector $\mathbf{o}$. The input of this network is a random initialization vector. By utilizing several fully connected layers and the softmax function, a probability vector $\mathbf{g} = \left[g_1, g_2, \cdots, g_{N^{\mathrm{R}}}\right]^T \in \mathbb{C}^{N^{\mathrm{R}} \times 1}$

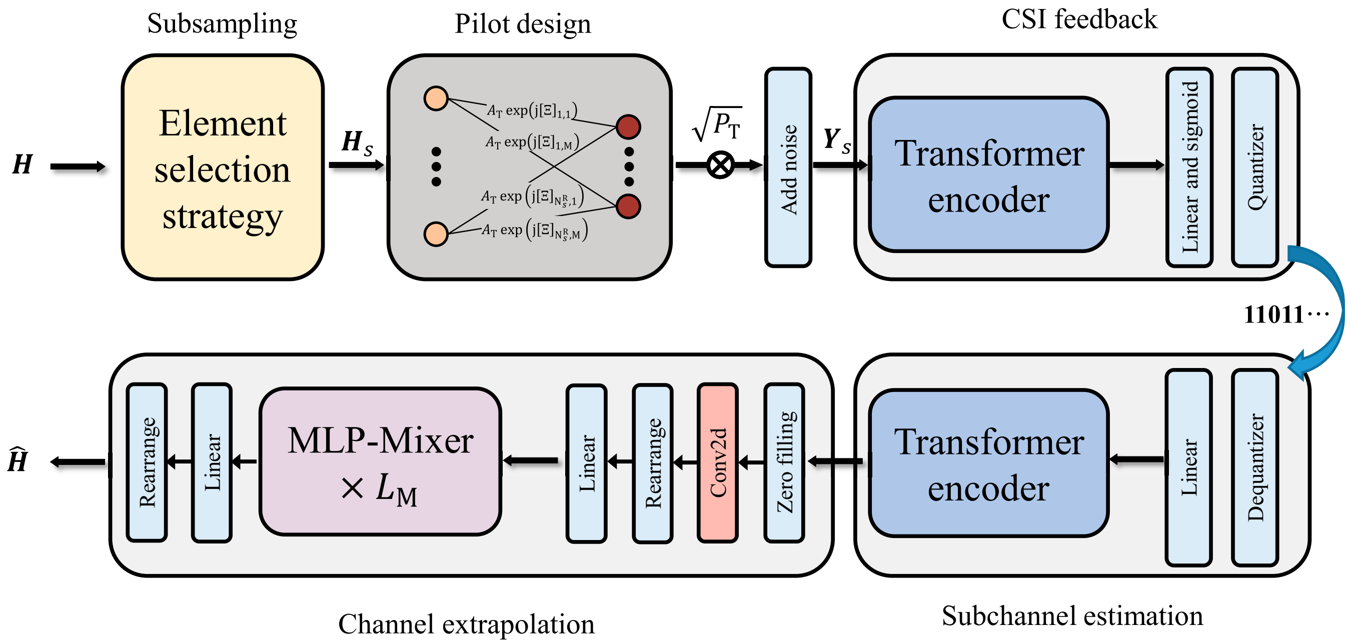
Subsampling
Pilot design
CSI feedback
H
Element selection strategy
H_s
$A_{\mathrm{T}}\exp(\mathrm{j}[\Xi]_{1,1})$
$A_{\mathrm{T}}\exp(\mathrm{j}[\Xi]_{1,\mathrm{M}})$
$A_{\mathrm{T}}\exp(\mathrm{j}[\Xi]_{\mathrm{N}_s^{\mathrm{R}},1})$
$A_{\mathrm{T}}\exp(\mathrm{j}[\Xi]_{\mathrm{N}_s^{\mathrm{R}},\mathrm{M}})$
$\sqrt{P_{\mathrm{T}}}$
Add noise
Y_s
Transformer encoder
Linear and sigmoid
Quantizer
11011⋯
Dequantizer
Linear
Transformer encoder
Zero filling
Conv2d
Rearrange
Linear
MLP-Mixer $\times L_{\mathrm{M}}$
Linear
Rearrange
$\widehat{H}$
Channel extrapolation
Subchannel estimation

Fig. 3. Overall block diagram of the proposed DL-based SFDCExtra scheme.

is generated, where g_i denotes the probability that the i-th element is selected. Thus, the active element index vector $\boldsymbol{\xi}$ can be defined as

$$\boldsymbol{\xi} = \arg\operatorname{top}_{N_s^{\mathrm{R}}}\{\mathbf{g}\}, \tag{19}$$

where $\arg\operatorname{top}_{N_s^{\mathrm{R}}}\{\cdot\}$ is a function that finds the element index set of the first N_s^{R} largest selection probabilities. The details of the selection network can be found in [25].

Pilot design

According to Eq. 11, under the assumption that the BS and RIS are arranged in a parallel symmetric array, the equivalent downlink pilots can be defined as $P_s = A_{\mathrm{T}} V_s$, where $V_s \in \mathbb{C}^{N_s^{\mathrm{R}} \times M}$ denotes the RIS phase matrix of selected elements at the pilot training stage. Thus, the pilot matrix $\mathbf{P}_s$ can be obtained by adjusting the RIS phase at different time slots as follows:

$$\mathbf{P}_s = A_{\mathrm{T}}\exp^{(j\Xi)} = A_{\mathrm{T}}\left(\cos(\Xi) + j\sin(\Xi)\right), \tag{20}$$

where $\Xi \in \mathbb{R}^{N_s^{\mathrm{R}} \times M}$ is the phase control matrix of selected RIS elements. Given that most DL frameworks, such as Tensorflow and Pytorch, have limited support for complex-valued operations, it is challenging to train the complex-valued pilot matrix $\mathbf{P}_s$ directly. To circumvent this issue, we adopt the real-valued RIS phase control matrix Ξ, whose entries take values in $[0, 2\pi)$ as trainable parameters of the PDN, and the pilot matrix $\mathbf{P}_s$ can be obtained from Eq. 20. The structure of the PDN is shown in Fig. 3, where trainable parameters of the PDN, i.e., Ξ, are learned at the DL training stage.

CSI feedback

Recently, DL-based solutions, such as CsiNet [48], have achieved good performance for CSI feedback. Furthermore, an emerging CSI feedback architecture based on a transformer [49] has been demonstrated to further reduce the feedback overhead and obtain more efficient compression performance than the CsiNet framework [50]. Therefore, we utilize a transformer as the backbone of the CSI feedback network $f_{CsiFd}(\cdot)$. The original transformer is divided into an encoder and a decoder. However, given that we are dealing with CSI without time-sequential information, there is no causality constraint. Thus, we only exploit the encoder module of the transformer, which produces outputs in parallel. Because real-valued operations are more effective and the transformer can only extract the correlation between sequences, we convert the received pilot signal into a real-valued 2-dimensional (2D) sequence $\overline{\mathbf{Y}}_s \in \mathbb{R}^{K_s \times 2M}$, which can be expressed as

$$\overline{\mathbf{Y}}_s = \left[\Re\{\mathbf{Y}_s\}, \Im\{\mathbf{Y}_s\}\right], \tag{21}$$

where the number of subcarriers K_s represents the length of the input sequence.

A schematic diagram of the transformer encoder is shown in Fig. 4. Through the fully connected linear embedding layer, the input sequence $\overline{\mathbf{Y}}_s$ can be converted into $\mathbf{X}_s \in \mathbb{R}^{K_s \times d_{\mathrm{T}}}$, which merges the relative position information of the subcarriers using the positional embedding layer. Then, multiple encoder layers are utilized to extract correlations between sequences. Each encoder layer has the same structure and is composed of a multihead self-attention sublayer followed by a position-wise multilayer perceptron (MLP) sublayer. Layer norm is applied before every block and the residual connection is applied after every block. Note that the multihead attention mechanism plays a key role in the performance improvement of the transformer. As shown in Fig. 4, the input sequence $\mathbf{X}_s$ is first projected onto 3 different sequential vectors: queries, keys, and values with different learned linear projections, namely $\{\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i\} \in \mathbb{R}^{K_s \times d_m}, 1 \le i \le h$, where $d_m = d_{\mathrm{T}} / h$ and h is the number of heads. Then, each value $\mathrm{head}_i \in \mathbb{R}^{K_s \times d_m}, 1 \le i \le h$, is outputted by performing the scaled dot-product attention simultaneously, where the weights on values can be obtained from a softmax function, which is expressed as

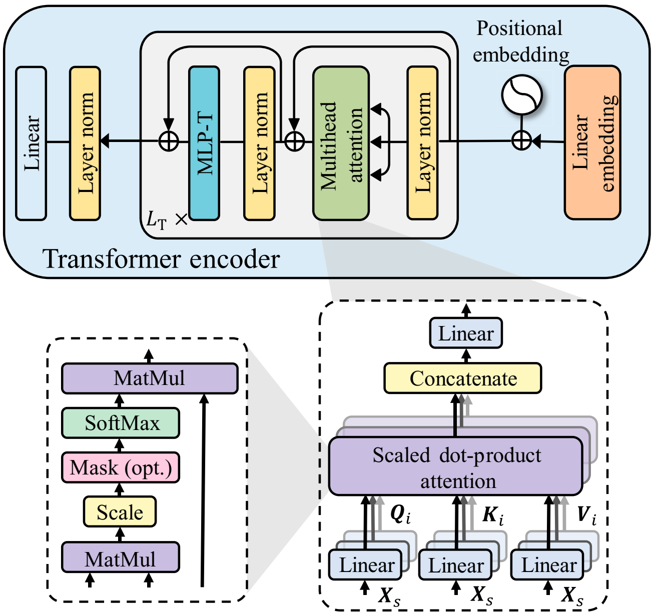

Fig. 4. Structure of the transformer encoder.

$$\text{head}_i = \text{softmax}\left(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{d_m}}\right)\mathbf{V}_i, 1 \le i \le h. \tag{22}$$

These output values are concatenated and projected back to a d_T-dimensional representation using the linear projection matrix d_T as $\mathbf{W}^O \in \mathbb{R}^{K_s \times d_\text{T}}$

$$\text{MultiHead}(\mathbf{X}_s) = \text{Concat}(\text{head}_i, \cdots, \text{head}_h)\mathbf{W}^O. \tag{23}$$

After the transformer encoder, a linear layer followed by a sigmoid function is used to generate a compressed codeword, which is then transformed into B bits as the feedback information through a uniform scalar quantization layer. This feedback process generates the binary vector $\mathbf{q} \in \{0, 1\}^B$ as

$$\mathbf{q} = f_\text{CsiFd}(\overline{\mathbf{Y}}_s; \mathcal{W}_F), \tag{24}$$

where $\mathcal{W}_F$ denotes the trained parameter set of the CSI feedback network.

Subchannel estimation

When the BS receives the feedback bits, the subchannel estimation network is used to reconstruct the subsampling of the complete spatial-frequency channel. As in the CSI feedback subsection, we also consider the transformer encoder as the backbone of this part. As shown in Fig. 3, received CSI feedback bits are initially processed by a dequantization layer, which conducts the inverse operation of the quantizer and outputs a real-valued vector. Then, the initial coarse channel estimate is obtained by a linear layer. Finally, the transformer encoder extracts the spatial-frequency correlation of the channel and further improves the channel estimation performance. The subchannel estimation process can be expressed as

$$\overline{\mathbf{H}}_s = \left[\Re\left\{\hat{\mathbf{H}}_s\right\}, \Im\left\{\hat{\mathbf{H}}_s\right\}\right] = f_\text{SCE}(\mathbf{q}; \mathcal{W}_S), \tag{25}$$

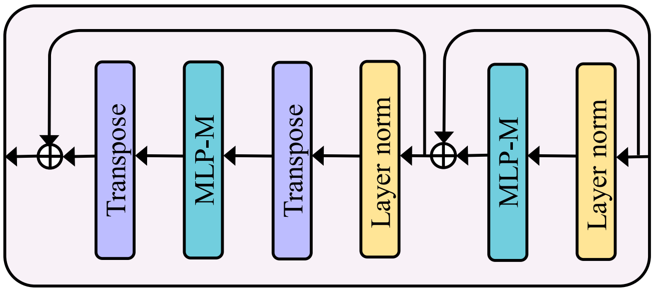

Fig. 5. Structure of the mixer layer.

where $\hat{\mathbf{H}}_s \in \mathbb{C}^{K_s \times N_s^\text{R}}$ is the estimated subsampling channel, $\overline{\mathbf{H}}_s \in \mathbb{R}^{K_s \times N_s^\text{R} \times 2}$ is a real-valued 3D matrix, and $\mathcal{W}_S$ is the trained parameter set of the subchannel estimation network.

Spatial-frequency domain channel extrapolation

First, the initial input $\tilde{\mathbf{H}} \in \mathbb{R}^{K \times N^\text{R} \times 2}$ to the channel extrapolation network is constructed from the estimated subsampling channel $\overline{\mathbf{H}}_s \in \mathbb{R}^{K_s \times N_s^\text{R} \times 2}$ with the known RIS spatial-frequency selection pattern $\mathbf{S}$. Specifically, we copy the entries of $\overline{\mathbf{H}}_s$ to the corresponding positions in $\tilde{\mathbf{H}}$ and fill the other elements of $\tilde{\mathbf{H}}$ with zeros according to the known RIS spatial-frequency selection pattern $\mathbf{S}$. This initial operation is represented by

$$\tilde{\mathbf{H}} = f_\text{zfi}(\overline{\mathbf{H}}_s; \mathbf{S}). \tag{26}$$

The nonzero rows/columns in $\tilde{\mathbf{H}}$ are consistent with $\overline{\mathbf{H}}_s$, and their locations are the same as those of the "1" elements in $\mathbf{S}$. The neighborhood information in the receptive field is then extracted using a convolutional layer for initial interpolation. To guarantee that the output dimensions from the convolution layer remain unchanged, we employ zero padding, i.e., adding zeros around the input feature map.

Subsequently, we consider a competitive yet conceptually and technically simple architecture, called the MLP-Mixer [51], as the backbone of the channel extrapolation network. The architecture of this MLP-Mixer is based entirely on MLPs, which can extract and reconstruct 2D features by repeatedly applying them to either spatial locations or feature channels. Specifically, the input $\tilde{\mathbf{H}} \in \mathbb{R}^{K \times N^\text{R} \times 2}$ is rearranged as a series of flattened 2D patches $\mathbf{X}_p \in \mathbb{R}^{N_p \times (2L^2)}$, where (K, N^R) represents the size of the original input, (L, L) represents the length and width of each path, and $N_p = KN^\text{R} / L^2$ represents the number of patches. Then, all the patches are linearly projected with the same projection matrix. This results in a 2D real-valued matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{N_p \times d_\text{M}}$. Next, the input matrix $\tilde{\mathbf{X}}$ is fed into several mixer layers to extrapolate the complete channel. As illustrated in Fig. 5, each mixer layer consists of 2 MLP blocks. The first acts on the columns of $\tilde{\mathbf{X}}$, maps $\mathbb{R}^{N_p} \mapsto \mathbb{R}^{2N_p} \mapsto \mathbb{R}^{N_p}$, and is shared across all the columns. The second acts on the rows of $\tilde{\mathbf{X}}$, i.e., on the transposed input matrix $\tilde{\mathbf{X}}^T$, maps $\mathbb{R}^{d_\text{M}} \mapsto \mathbb{R}^{2d_\text{M}} \mapsto \mathbb{R}^{d_\text{M}}$, and is shared across all the rows. Each MLP block contains 2 fully connected layers and a nonlinear activation function. The mapping of the t-th mixer layer can be expressed as

$$\begin{aligned} \mathbf{U} &= \tilde{\mathbf{X}}_t + \mathbf{W}_{t,2} f_\sigma(\mathbf{W}_{t,1}\text{LayerNorm}(\tilde{\mathbf{X}}_\text{t})), \\ \tilde{\mathbf{X}}_{t+1} &= \mathbf{U} + (\mathbf{W}_{t,4} f_\sigma(\mathbf{W}_{t,3}\text{LayerNorm}(\mathbf{U})^T))^T, \end{aligned} \tag{27}$$

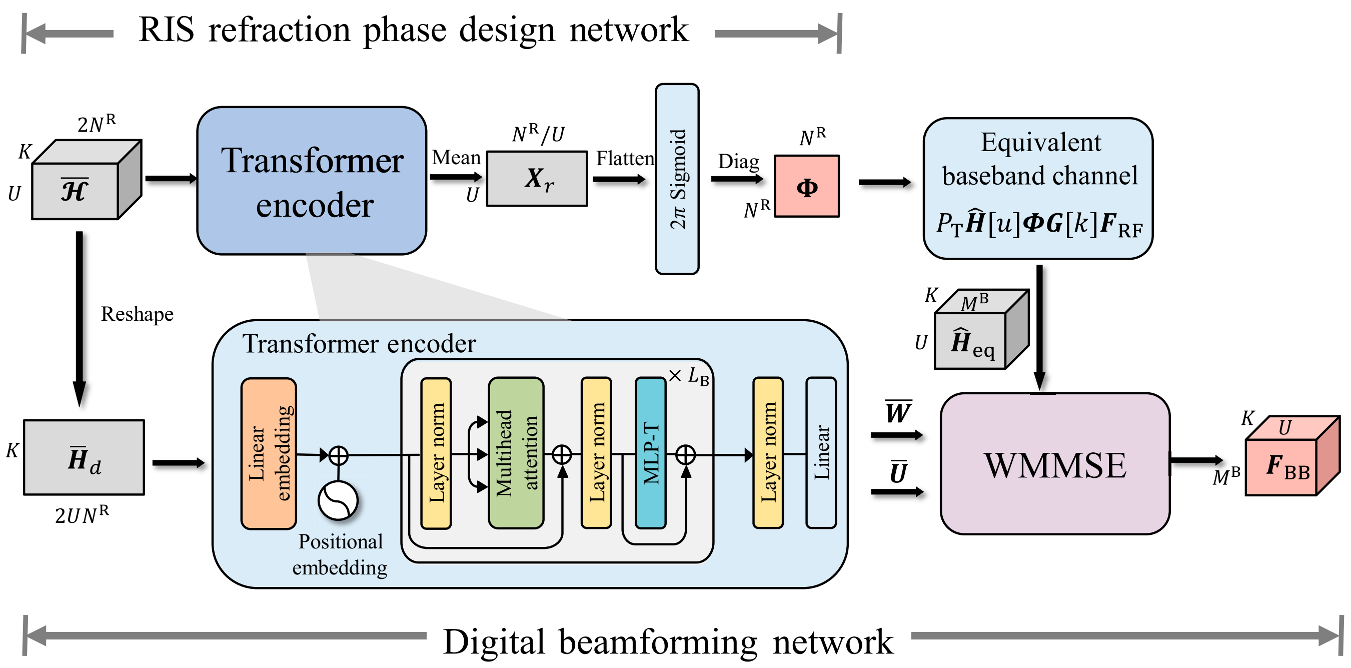

Fig. 6. Overall structure of the proposed DL-based hybrid beamforming and RIS refraction phase design scheme.

where $\tilde{\mathbf{X}}_t$ denotes the input matrix to the t-th mixer layer, $\mathbf{W}_{t,i}$, $1 \le i \le 4$, are the parameter matrices of the fully connected layers in the t-th mixer layer for $1 \le t \le L_{\mathrm{M}}$, L_{M} is the number of mixer layers, and f_σ denotes an activation function.

Finally, the output of the last mixer layer is linearly projected back to the original dimension $\mathbb{R}^{N_p \times d_{\mathrm{M}}} \mapsto \mathbb{R}^{N_p \times (2L^2)}$, and the 2D patches are rearranged back to $\mathbb{R}^{N_p \times (2L^2)} \mapsto \mathbb{R}^{K \times N^{\mathrm{R}} \times 2}$ for obtaining the final extrapolation result $\overline{\mathbf{H}} \in \mathbb{R}^{K \times N^{\mathrm{R}} \times 2}$, which is a real-valued 3D matrix. Thus, the extrapolation process is expressed as

$$\widehat{\mathbf{H}} = \overline{\mathbf{H}}_{[:,:,1]} + j\overline{\mathbf{H}}_{[:,:,2]} = f_{\mathrm{SFDE}}\left(\overline{\mathbf{H}}_s; \mathcal{W}_E\right), \tag{28}$$

where $\widehat{\mathbf{H}} \in \mathbb{C}^{K \times N^{\mathrm{R}}}$ is the estimated complete complex-valued channel, and $\mathcal{W}_E$ is the trained parameter set of the spatial-frequency domain extrapolation network.

Training strategy

The offline training dataset, denoted as $\boldsymbol{\mathcal{H}}$, comprises $|\boldsymbol{\mathcal{H}}| = N_{\mathrm{set}}$ samples. Each sample in $\boldsymbol{\mathcal{H}}$ is an input–label pair denoted as ($\mathbf{H}$, $\mathbf{H}$), where $\mathbf{H}$ serves as both the extrapolation target and the input for the SFDCExtra network. The input will go through the RIS array element and subcarrier subsampling strategy, because the original complete channel must be extrapolated from only the received pilot signal of the subsampling channel.

With the uniform or random ESS $f_{\mathrm{ESS}}(\cdot)$, at the offline training stage, E2E training is conducted for the PDN, CSI feedback network, subchannel estimation network, and channel extrapolation network. Thus, the loss function involves minimizing the normalized mean square error (NMSE) between the output $\widehat{\mathbf{H}}$ and target $\mathbf{H}$, i.e.,

$$\mathcal{L}_c = \frac{1}{B_e} \sum_{i=1}^{Be} \frac{\mathbf{H} - \widehat{\mathbf{H}}_F^2}{\mathbf{H}_F^2}, \tag{29}$$

where B_e is the batch size for offline training.

When the learning-based ESS is adopted, the parameters for the ESS and the above networks are optimized jointly, i.e., the loss function can be expressed as

$$\mathcal{L} = \gamma \mathcal{L}_c + (1 - \gamma)\mathcal{L}_{\mathrm{ESS}}, \tag{30}$$

where $0 < \gamma \le 1$ represents the weight used to balance channel extrapolation and ESS, with $\gamma = 1$ denoting that the nonlearning-based $f_{\mathrm{ESS}}(\cdot)$ is selected, and $\mathcal{L}_{\mathrm{ESS}}$ is the loss function of the learning-based ESS. The details of $\mathcal{L}_{\mathrm{ESS}}$ are available in [25].

Proposed beamforming solution

Problem formulation of RIS-aided multiuser beamforming

The BS can simultaneously support U UEs with the aid of RIS at the data transmission stage, given that the LoS MIMO architecture can support multistream transmission via intrapath multiplexing. Similar to Eq. 6, the received signal at the u-th UE on the k-th subcarrier can be expressed as

$$\begin{aligned} y[u,k] = {} & \sqrt{P_{\mathrm{T}}}\mathbf{h}[u,k]\boldsymbol{\Phi}\mathbf{G}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}[u,k]s[u,k] \\ & + \sum_{i=1, i \ne u}^{U} \sqrt{P_{\mathrm{T}}}\mathbf{h}[u,k]\boldsymbol{\Phi}\mathbf{G}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}[i,k]s[i,k] + n[u,k] \end{aligned} \tag{31}$$

where $\mathbf{h}\left[u,k\right] \in \mathbb{C}^{1 \times N^R}$, $1 \le u \le U, 1 \le k \le K$ denotes the downlink RIS-UE channel of the u-th UE on the k-th subcarrier, and $\mathbf{f}_{\mathrm{BB}}\left[u,k\right] \in \mathbb{C}^{M^{\mathrm{B}} \times 1}$denotes the digital baseband beamforming vector associated with the u-th UE on the k-th subcarrier. Thus, the signal-to-interference-plus-noise ratio (SINR) of the u-th UE on the k-th subcarrier can be expressed as

$$\mathrm{SINR}\left[u,k\right] = \frac{P_{\mathrm{T}}\left|\mathbf{h}\left[u,k\right]\boldsymbol{\Phi}\mathbf{G}\left[k\right]\mathbf{F}_{RF}\mathbf{f}_{BB}\left[u,k\right]\right|^2}{P_T \sum_{i=1, i \ne u}^{U}\left|\mathbf{h}\left[u,k\right]\boldsymbol{\Phi}\mathbf{G}\left[k\right]\mathbf{F}_{RF}\mathbf{f}_{BB}\left[i,k\right]\right|^2 + \sigma_n^2}. \tag{32}$$

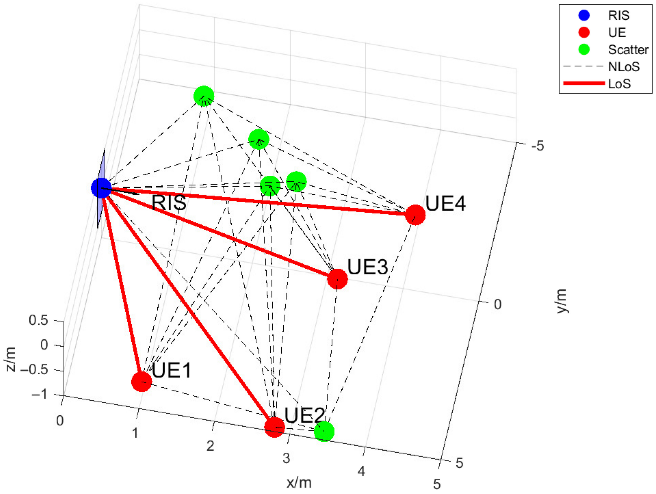

Fig. 7. Schematic diagram of RIS-UE channel model for the indoor environment.

Therefore, the sum rate R for the whole set of UEs is expressed as

$$R = \frac{1}{K}\sum_{u=1}^{U}\sum_{k=1}^{K}\log_2\left(1+\mathrm{SINR}[u,k]\right). \tag{33}$$

By utilizing the estimated RIS-UE channel at the pilot training stage, the BS can design the hybrid beamformer $\{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k], \forall k\}$and RIS refraction phase matrix $\mathbf{\Phi}$ to maximize the sum rate R, where $\mathbf{F}_{\mathrm{BB}}[k] = [\mathbf{f}_{\mathrm{BB}}[1,k], \cdots, \mathbf{f}_{\mathrm{BB}}[U,k]]$. This design process is illustrated as

$$\begin{aligned} &\max_{\mathcal{F}(\cdot)}\ R, \\ &s.t.\ \ \{\mathbf{F}_{RF}, \mathbf{F}_{BB}[k], \forall k, \mathbf{\Phi}\} = \mathcal{F}\left(\widehat{\mathbf{H}}[u], \forall u\right), \\ &\qquad \mathbf{F}_{RF} \in (8), \\ &\qquad \left\|\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\right\|_F^2 = M^{\mathrm{B}}, \forall k, \\ &\qquad \{\mathbf{\Phi}\}_{i,i} = \{\mathbf{v}\}_i = e^{j\phi_i}, \phi_i \in [0, 2\pi), \forall i, \end{aligned} \tag{34}$$

where $\widehat{\mathbf{H}}[u]$ is the estimated spatial-frequency RIS-UE channel of the u-th UE, and $\mathcal{F}(\cdot)$ represents a function that maps the estimated RIS-UE channels onto the hybrid beamformer $\{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k], \forall k\}$ and RIS refraction phase matrix $\mathbf{\Phi}$.

Proposed DL-based HBFRPD scheme

To solve the optimization in Eq. 34, some alternating iterative algorithms [28,29,31] have been proposed to obtain the analog beamformer, digital beamformer, and RIS phase, respectively. Unfortunately, all the aforementioned approaches are based on the idealized case that the CSI is known accurately. However, perfect CSI is usually unavailable, especially for indoor channel cases where the channel characteristics are complex owing to rich scattering. By using the nonlinear function fitting capability of DL, we can learn the complicated and unknown mapping from the estimated channels to the hybrid beamformers and RIS refraction phase. Thus, we propose a DL-based HBFRPD scheme that consists of the analog beamformer design, DL-based RIS refraction phase design, and knowledge-data dual-driven digital beamformer design. A diagram depicting the design of the proposed scheme is presented in Fig. 6.

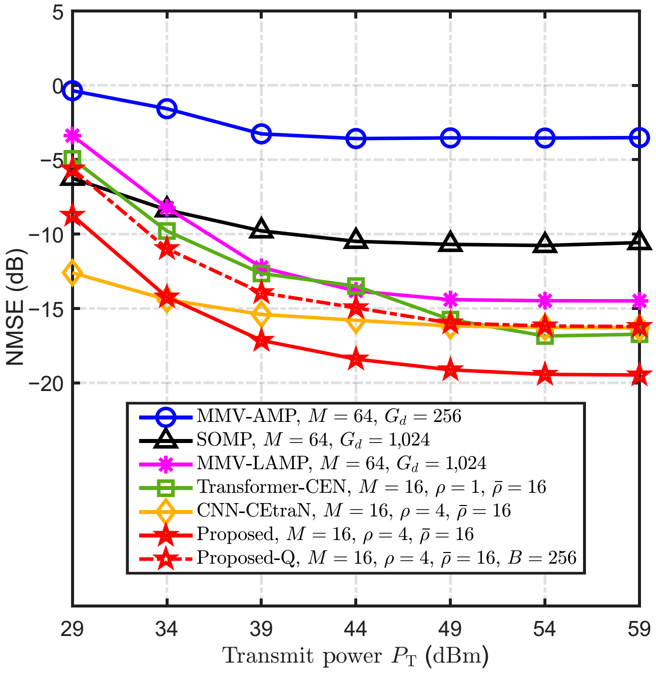

Fig. 8. NMSE performance comparison of different channel estimation schemes versus transmit power $\mathrm{P_T}$. The labels "Transformer-CEN" and "CNN-CEtraN" are short for transformer-based channel estimation network [50] and CNN-based channel extrapolation network [27], respectively.

Analog beamformer design

The integration of active and passive beamforming at the BS and RIS is a nonconvex optimization problem that poses considerable difficulties in finding a global optimum solution. Hence, we separately design the analog and passive beamforming. Specifically, both BS analog and RIS passive beamforming are designed to focus energy for improving the received SINR of UEs. However, given the subconnected structure in the LoS MIMO architecture, the interference among beams from the BS and RIS subarrays cannot be eliminated. Fortunately, this part of interference can be removed by appropriately designing the digital beamforming. Therefore, when designing the analog beamforming on the BS side, it is sufficient to assume that the transmit energy is focused on the RIS.

Given that the BS-RIS channel with only the LoS path is quasi-static and known, we can utilize the angle information of the BS-RIS link to design the analog beamformer. Specifically, the transmit beam of the m_b-th subarray designed for the u-th UE should be aligned with the m_r-th subarray of the RIS, where the u-th UE is assisted by the m_r-th subarray of the RIS. Therefore, the analog beamformer $\mathbf{F}_{\mathrm{RF}} = \mathrm{blkdiag}\left(\mathbf{f}_1, \cdots, \mathbf{f}_{m_b}, \cdots, \mathbf{f}_{M^{\mathrm{B}}}\right)$ can be simply designed for alignment between the BS and RIS subarrays according to Eq. 9.

DL-based RIS refraction phase design

Optimizing a common RIS phase shared by all the subcarriers is a crucial challenge in a RIS-aided OFDM system. In the THz broadband case, there exists a non-negligible beam squint effect

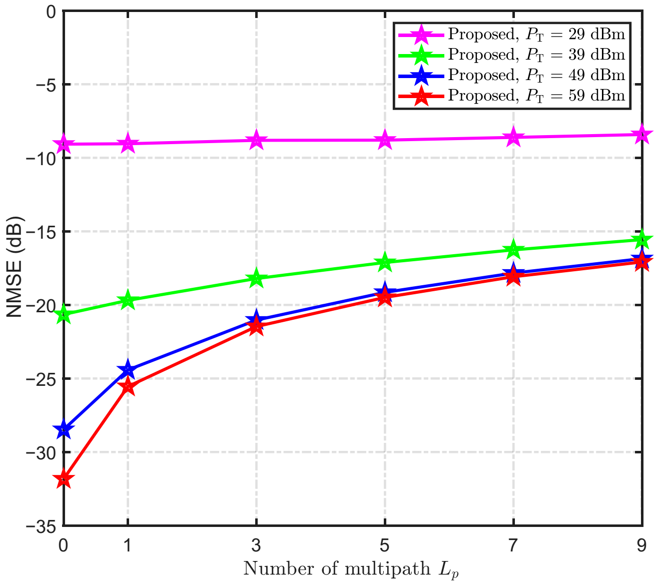

Fig. 9. NMSE performance comparison of the proposed scheme versus the number of multipath L_p, given $\rho = 4$, $\bar{\rho} = 16$ and $M = 16$. Offline training is based on the channel samples with $L_p = 5$ multipath components.

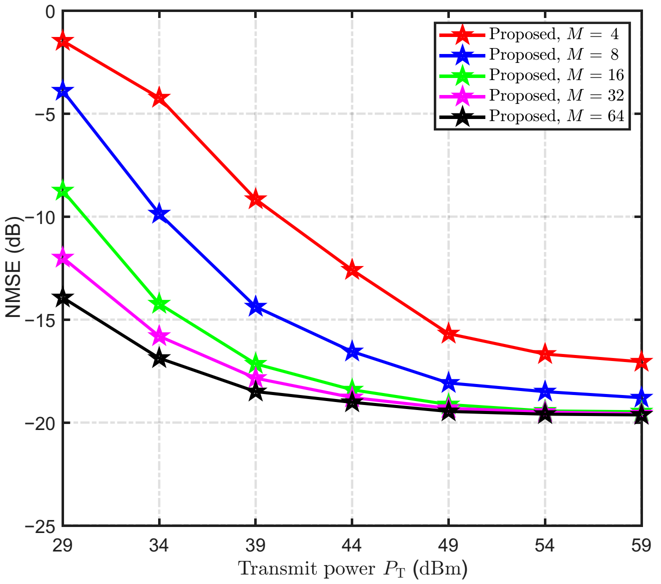

Fig. 10. NMSE performance comparison of the proposed scheme with different pilot numbers versus transmit power P_{T}, given $\rho = 4$, $\bar{\rho} = 16$, $L_p = 5$.

for different subcarriers [47]. Therefore, when designing the common RIS phase, it is necessary to consider this effect on all subcarriers, which makes the RIS phase design for the broadband case much more difficult than the design for the narrowband case. To solve this challenging problem, a transformer-based RPDN is proposed in Fig. 6 to design the RIS refraction phase matrix.

We first convert all the estimated RIS-UE channels $\widehat{\mathbf{H}}[u] \in \mathbb{C}^{K\times N^{\mathrm{R}}}$ for $1 \le u \le U$ into a real-valued 3D matrix $\overline{\mathcal{H}} \in \mathbb{R}^{U\times K\times 2N^{\mathrm{R}}}$, i.e.,

$$\overline{\mathcal{H}} = \left[\overline{\mathbf{H}}[1], \cdots, \overline{\mathbf{H}}[u], \cdots, \overline{\mathbf{H}}[U]\right], \tag{35}$$

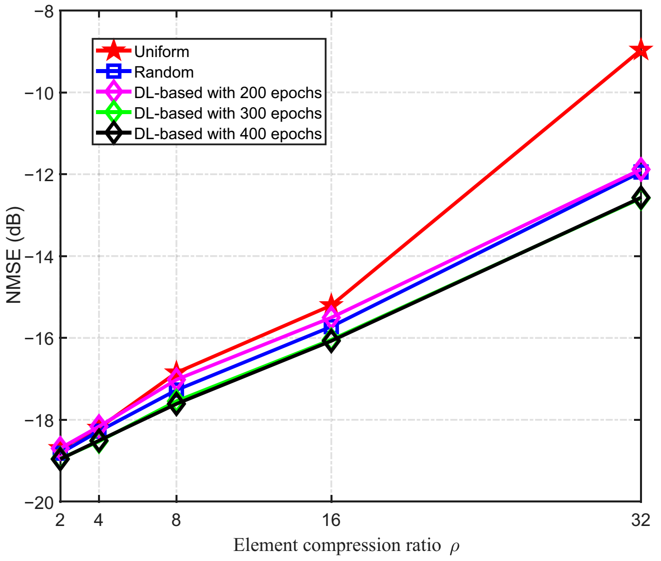

Fig. 11. NMSE performance comparison of the proposed scheme with different element selection strategies versus element compression ratio ρ, for $\bar{\rho} = 16$, $L_p = 5$, $M = 16$, $P_{\mathrm{T}} = 44$ dBm.

where $\overline{\mathbf{H}}[u] = \left[\Re\left\{\widehat{\mathbf{H}}[u]\right\}, \Im\left\{\widehat{\mathbf{H}}[u]\right\}\right] \in \mathbb{R}^{K\times 2N^{\mathrm{R}}}$ and $\widehat{\mathbf{H}}[u]$ is the estimated RIS-UE channel of the u-th UE obtained from the DL-based SFDCExtra network. Note that $\overline{\mathcal{H}}$ is inputted into the transformer encoder, which globally extracts the inter-subcarrier correlation. To consider the beam squint effect for different subcarriers, the 2D matrix $\mathbf{X}_r \in \mathbb{R}^{U\times N^{\mathrm{R}}/U}$ is obtained by averaging over the subcarrier dimension of the output of the transformer encoder. Then, $\mathbf{X}_r$ is flattened as $\mathbf{x}_r \in \mathbb{R}^{N^{\mathrm{R}}\times 1}$ and passes through the activation function to generate the RIS phase vector $\mathbf{v} \in \mathbb{C}^{N^{\mathrm{R}}\times 1}$ that satisfies the constant modulus constraint, i.e.,

$$\mathbf{v} = e^{j2\pi\cdot\mathrm{Sigmoid}(\mathbf{x}_r)}. \tag{36}$$

Finally, the RIS phase matrix $\boldsymbol{\Phi} \in \mathbb{C}^{N^{\mathrm{R}}\times N^{\mathrm{R}}}$ is obtained through diagonalization. The overall process of the RIS refraction phase design, namely the transformer-based RPDN, can be expressed as

$$\boldsymbol{\Phi} = f_{\mathrm{RIS}}\left(\overline{\mathcal{H}}; \mathcal{W}_R\right), \tag{37}$$

where $f_{\mathrm{RIS}}(\cdot)$ denotes the mapping of the RPDN, whose trainable parameter set is $\mathcal{W}_R$.

Knowledge-data dual-driven digital beamformer design

Using the known BS-RIS channel $\mathbf{G}[k]$, designed RIS refraction phase matrix $\boldsymbol{\Phi}$, analog beamforming matrix $\mathbf{F}_{\mathrm{RF}}$, and estimated RIS-UE channel $\widehat{\mathbf{h}}[u,k]$, the estimated equivalent baseband channel $\widehat{\mathbf{h}}_{\mathrm{eq}}[u,k] \in \mathbb{C}^{1\times M^{\mathrm{B}}}$ can be expressed as

$$\widehat{\mathbf{h}}_{\mathrm{eq}}[u,k] = P_{\mathrm{T}}\widehat{\mathbf{h}}[u,k]\boldsymbol{\Phi}\mathbf{G}[k]\mathbf{F}_{RF}. \tag{38}$$

The true equivalent baseband channel $\mathbf{h}_{\mathrm{eq}}[u,k]$ has a form similar to that of Eq. 38, given the designed $\boldsymbol{\Phi}$ and $\mathbf{F}_{\mathrm{RF}}$. Therefore, Eq. 34 can be simplified as

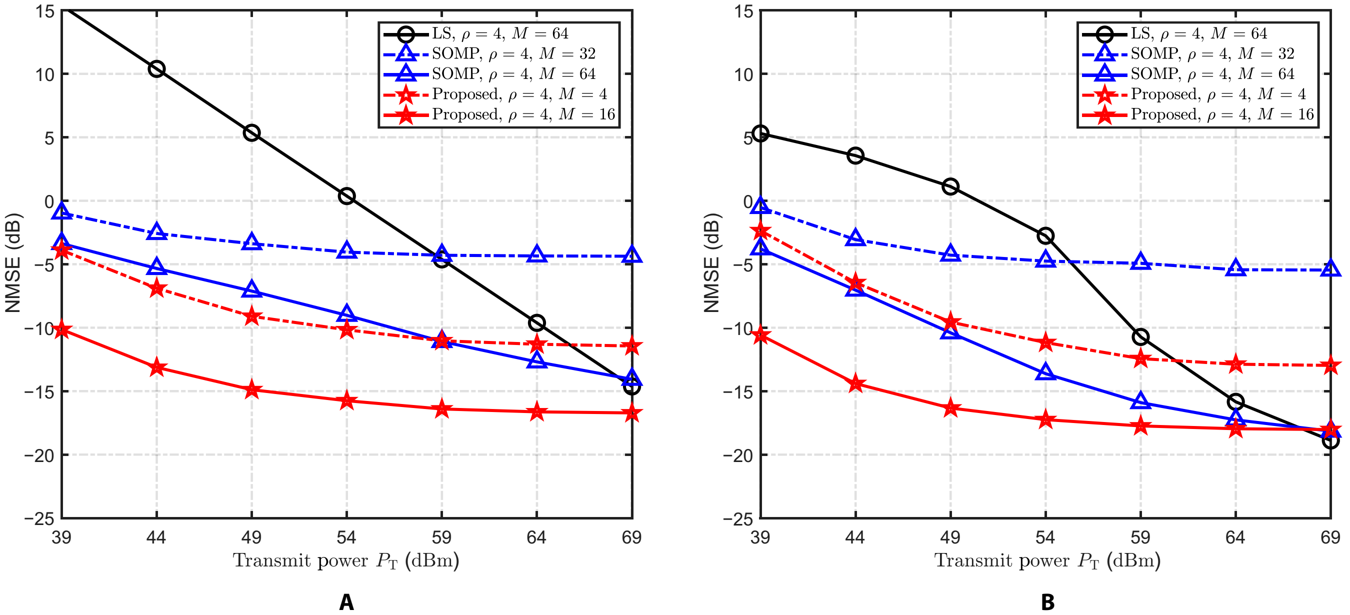

Fig. 12. Effectiveness of the channel extrapolation module with different subchannel estimation schemes. (A) NMSE performance comparison of different subchannel estimation schemes versus transmit power P_{T}. (B) NMSE performance of channel extrapolation versus transmit power P_{T} for different subchannel estimation schemes.

$$\begin{aligned} &\max_{\mathbf{F}_{BB}[k],\forall k} && \frac{1}{K}\sum_{u=1}^{U}\sum_{k=1}^{K}\log_2\left(1+\mathrm{SINR}\left[u,k\right]\right), \\ &s.t. && \mathrm{SINR}\left[u,k\right] = \frac{\left|\mathbf{h}_{\mathrm{eq}}\left[u,k\right]\mathbf{f}_{BB}\left[u,k\right]\right|^2}{\sum_{i=1,i\neq u}^{U}\left|\mathbf{h}_{\mathrm{eq}}\left[u,k\right]\mathbf{f}_{BB}\left[i,k\right]\right|^2+\sigma_n^2} \\ &&& \mathbf{F}_{RF}\mathbf{F}_{BB}\left[k\right]_F^2 = M^{\mathrm{B}},\ \ \forall k. \end{aligned} \tag{39}$$

Note that the design of the analog beamformer and RIS refraction phase have been specifically designed in the preceding subsections. Equation 39 is a classic baseband beamforming problem and can be solved with standard linear beamforming schemes, such as the regularized ZF (RZF) or iterative weighted minimum mean-square error (WMMSE) algorithm. Taking the latter as an example, the iterative WMMSE algorithm is designed to solve the optimization problem in Eq. 39 by addressing the equivalent MMSE problem specified in Eq. 40 below, which has an identical optimal solution $\mathbf{F}_{\mathrm{BB}}[k], \forall\, k$ to the problem in Eq. 39.

$$\begin{aligned} &\max_{\overline{\mathbf{U}},\overline{\mathbf{W}},\mathbf{F}_{\mathrm{BB}}[k],\forall k} && \sum_{u=1}^{U}\sum_{k=1}^{K}\left(\overline{w}_{u,k}e_{u,k}-\log_2\overline{w}_{u,k}\right), \\ &s.t. && \mathbf{F}_{RF}\mathbf{F}_{BB}\left[k\right]_F^2 \leq M^{\mathrm{B}},\ \ \forall k, \end{aligned} \tag{40}$$

where $\overline{w}_{u,k} = \left\{\overline{\mathbf{W}}\right\}_{u,k}$ is the weight of the u-th user on the k-th subcarrier, $e_{u,k} = \mathbb{E}\left\{\left|\hat{s}[u,k]-s[u,k]\right|^2\right\}$ is the MSE between the transceiver symbols under the independence assumption of $s[u,k]$ and $n[u,k]$, $\hat{s}\left[u,k\right] = \overline{u}_{u,k}y\left[u,k\right]$ is the estimated data symbol at the UE side, and $\overline{u}_{u,k} = \left\{\overline{\mathbf{U}}\right\}_{u,k}$ is the receiver gain of the u-th UE on the k-th subcarrier. According to [52], the above problem is convex in each individual optimization variable. This property enables each subproblem to have a closed-form solution, given the other optimization variables. Then, the optimization problem in Eq. 40 can be solved by a block coordinate descent iterative algorithm. The iterative WMMSE beamforming design algorithm is described in Algorithm 1.

However, the iterative WMMSE algorithm typically requires a large number of iterations with a long execution time. Furthermore, the BS can only acquire imperfect estimated CSI $\hat{\mathbf{h}}_{\mathrm{eq}}\left[u,k\right]$, and it is difficult for traditional digital beamforming algorithms, such as Algorithm 1, to overcome the interference induced by imperfect CSI. Thus, we propose a knowledge-data dual-driven digital beamforming network, as shown in Fig. 6, which utilizes the transformer encoder to directly learn the parameters of the iterative WMMSE algorithm from imperfect CSI for better interference elimination and shorter execution time.

Specifically, the real-valued 3D matrix $\overline{\boldsymbol{H}} \in \mathbb{R}^{U\times K\times 2N^{\mathrm{R}}}$ is reshaped into a 2D matrix $\overline{\mathbf{H}}_d \in \mathbb{R}^{K\times 2UN^{\mathrm{R}}}$, which is inputted into the transformer encoder. This encoder outputs $\mathbf{X} \in \mathbb{R}^{K\times 4U}$, which is converted into the weight matrix $\overline{\mathbf{W}}$ and receiver gain matrix $\overline{\mathbf{U}}$, i.e.,

$$\overline{\mathbf{W}} = \mathbf{X}_{[:,:U]}^{T} + j\mathbf{X}_{[:,U:2U]}^{T}, \tag{41}$$

$$\overline{\mathbf{U}} = \mathbf{X}_{[:,2U:3U]}^{T} + j\mathbf{X}_{[:,3U:]}^{T}. \tag{42}$$

Then, we can obtain $\mathbf{F}_{\mathrm{BB}}\left[k\right], \forall k$, based on the learned $\overline{\mathbf{W}}$ and $\overline{\mathbf{U}}$ by the update function of $\mathbf{f}_{\mathrm{BB}}\left[u,k\right]$, i.e., line 5 of Algorithm 1. Compared with the iterative WMMSE beamforming design, the proposed scheme does not involve an iterative process so the execution time can be reduced considerably. To satisfy the transmission power constraint, the normalization operation can be expressed as

$$\mathbf{F}_{\mathrm{BB}}\left[k\right] = \frac{\sqrt{M^{\mathrm{B}}}\mathbf{F}_{\mathrm{BB}}\left[k\right]}{\left\|\mathbf{F}_{RF}\mathbf{F}_{BB}\left[k\right]\right\|_F}, \forall k. \tag{43}$$

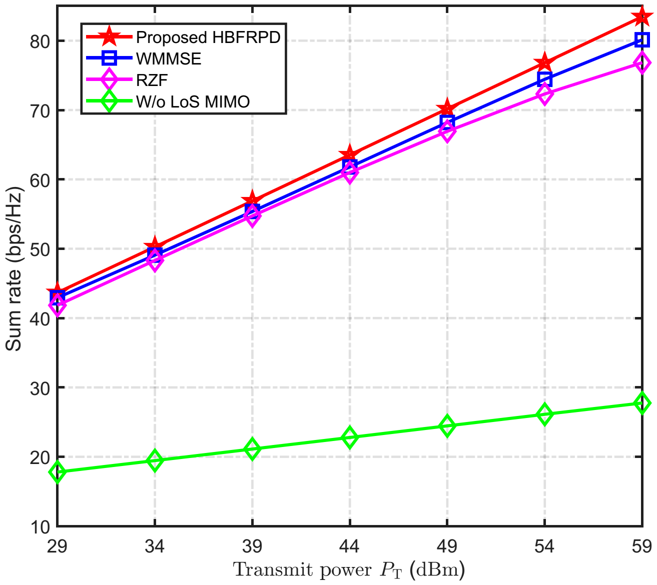

Fig. 13. Sum rates achieved by different schemes versus transmit power $P_{\rm T}$ for perfect CSI. The actual transmit power of the "w/o LoS MIMO" case is $\mathrm{UP_T} = 4\mathrm{P_T}$.

The proposed knowledge-data dual-driven digital beamformer design can be expressed as

$$\{\mathbf{F}_{\rm BB}[k], \forall k\} = f_{\rm DBF}(\overline{\mathbf{H}}_d, \mathcal{W}_D), \tag{44}$$

where $f_{\rm DBF}(\cdot)$ is the map of the digital beamforming network with a trainable parameter set $\mathcal{W}_D$.

```
Algorithm 1 Iterative WMMSE beamforming design algorithm
Initialize $\mathbf{F}_{\rm BB}[k]$ that meets $\|\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\|_F^2 = M^{\rm B}$, set the maximum iteration number $I_{\max}$ and current iteration index $t = 0$;
repeat
    Update $\{\bar{\mathbf{U}}\}_{u,k}$: $\bar{u}_{u,k} = \left(\sum_{i=1}^{U} |\mathbf{h}_{\rm eq}[u,k]\mathbf{f}_{BB}[i,k]|^2 + \sigma_n^2\right)^{-1} \mathbf{h}_{\rm eq}[u,k]\mathbf{f}_{BB}[u,k], \forall u,k;$
    Update $\{\bar{\mathbf{W}}\}_{u,k}$: $\bar{w}_{u,k} = \left(1 - \bar{u}_{u,k}^{*}\mathbf{h}_{\rm eq}[u,k]\mathbf{f}_{BB}[u,k]\right)^{-1}, \forall u,k;$
    Update $\mathbf{f}_{\rm BB}[u,k]$: $\mathbf{f}_{\rm BB}[u,k] = \bar{u}_{u,k}\bar{w}_{u,k}\left(\sum_{i=1}^{U} \bar{w}_{i,k}|\bar{u}_{i,k}|^2\mathbf{h}_{\rm eq}^H[i,k]\mathbf{h}_{\rm eq}[i,k] + \mu_k\mathbf{I}\right)^{-1}\mathbf{h}_{\rm eq}^H[u,k],$
    where $\mu_k = \sum_{j=1}^{U} \frac{\sigma^2}{M^{\rm B}}\bar{w}_{j,k}|\bar{u}_{j,k}|^2, \forall u,k;$
    $t = t + 1;$
    $t \geq I_{\max}$
Scale $\mathbf{F}_{\rm BB}[k]$ to meet the transmission power constraint.
```

Training strategy

We take every U channel sample (i.e., the channels of U UEs) in the training set of the channel estimation stage as a group to form a training set at the beamforming design stage, denoted as $\mathcal{H}_U$. The number of offline training samples is $|\mathcal{H}_U| = N_{\rm set}/U$. A sample in $\mathcal{H}_U$ is a UE set $\{\mathbf{H}[u], 1 \leq u \leq U\}$, where $\mathbf{H}[u]$ is the spatial-frequency RIS-UE channel of the u-th UE.

$\{\mathbf{H}[u], 1 \leq u \leq U\}$ are inputted to the trained SFDCExtra network to obtain the estimated channels $\left\{\hat{\mathbf{H}}[u], 1 \leq u \leq U\right\}$, which form the input to the proposed network. Given that imperfect CSI reduces the sum rate upper bound, to ensure a faster learning process, we apply a teacher forcing technique [53] at an early stage of training by feeding perfect CSI $\{\mathbf{H}[u], \forall u\}$ to the proposed network. At the offline training stage, we consider E2E training to jointly optimize the hybrid beamforming and RIS phase, i.e., the parameters of the entire network are trained by minimizing the negative sum rate. Thus, the loss function is expressed as

$$\mathcal{L}_b = -\frac{1}{B_b}\sum_{i=1}^{B_b} R, \tag{45}$$

where R is the sum rate defined in Eq. 33 and B_b is the batch size for offline training.

Results and Discussion

In this section, we describe how the effectiveness of the proposed SFDCExtra scheme as well as HBFRPD for a RIS-aided THz massive MIMO system was evaluated through numerical simulations.

Simulation settings

Communication scenario setup

In the conducted simulations, the BS was deployed on top of a building of height 30 m, and the RIS was installed on a window surface on one floor of another building. As shown in Fig. 1B, the BS (RIS) was equipped with $M^{\rm B} = M_y^{\rm B}M_z^{\rm B} = 4$ $\left(M^{\rm R} = M_y^{\rm R}M_z^{\rm R} = 4\right)$ subarrays on the yz-plane, where $M_y^{\rm B} = 2$ $\left(M_y^{\rm R} = 2\right)$ and $M_z^{\rm B} = 2$ $\left(M_z^{\rm R} = 2\right)$. Each subarray was a UPA with $N_{\rm sub}^{\rm B} = N_y^{\rm B}N_z^{\rm B} = 64\left(N_{\rm sub}^{\rm R} = N_y^{\rm R}N_z^{\rm R} = 64\right)$ isotropically radiating elements, where $N_y^{\rm B} = 8\left(N_y^{\rm R} = 8\right)$ and $N_z^{\rm B} = 8\left(N_z^{\rm R} = 8\right)$. Therefore, the number of elements of the complete array at the BS (RIS) was $N^{\rm B} = M^{\rm B}N_{\rm sub}^{\rm B} = 256\left(N^{\rm R} = M^{\rm R}N_{\rm sub}^{\rm R} = 256\right)$, and the distance between the BS and RIS was $D = 20$ m. The central frequency was $f_c = 0.3$ THz with a bandwidth $f_s = 1$ GHz. The number of OFDM subcarriers was $K = 128$, and the gain of the BS antenna was $G_T = 10$ dBi. According to these parameter settings, the subarray intervals of both the BS and RIS were calculated from Eq. 1 as $d_{sy}^{\rm B}, d_{sz}^{\rm B}, d_{sy}^{\rm R}, d_{sz}^{\rm R} = 96.5\lambda$ for obtaining the multistream multiplexing gain over the LoS path.

Figure 7 depicts a schematic diagram of the RIS-UE channel model for the indoor environment, where the positions of the RIS, UEs, and scatterers are indicated by blue, red, and green circles, respectively. The RIS-UE LoS path is depicted by a red solid line, and the NLoS link via a scatterer is represented by a black dotted line. We assumed that $U = 4$ UEs were randomly distributed over the xy-plane of the rectangular room ($W_x = 5$ m, $W_y = 10$ m), and the height of UEs was 1 m lower than the RIS. The number of available NLoS paths (scatterers) was set to $L_p = 5$, implying that only a single-bounce scattering mode was considered. The reflection coefficient parameters $\beta_{\rm RC}$ were set to $\mu_{\rm R} = -5, \sigma_{\rm R} = 2$. The noise power spectrum density at the UEs was $\sigma_{\rm NSD}^2 = -174$ dBm/Hz. Thus, the power of the additive white Gaussian noise was $\sigma_n^2 = \sigma_{\rm NSD}^2 f_s/K = -105$ dBm. The RIS-UE channel samples were generated using Eq. 5, and the UEs and scatterers were distributed randomly each time.

SFDCExtra network parameter configuration

In the CSI feedback network, the linear embedding layer of the transformer encoder has $d_{\rm T} = 256$ neurons. In the transformer encoder, the number of encoder layers is $L_{\rm T} = 3$, the number of

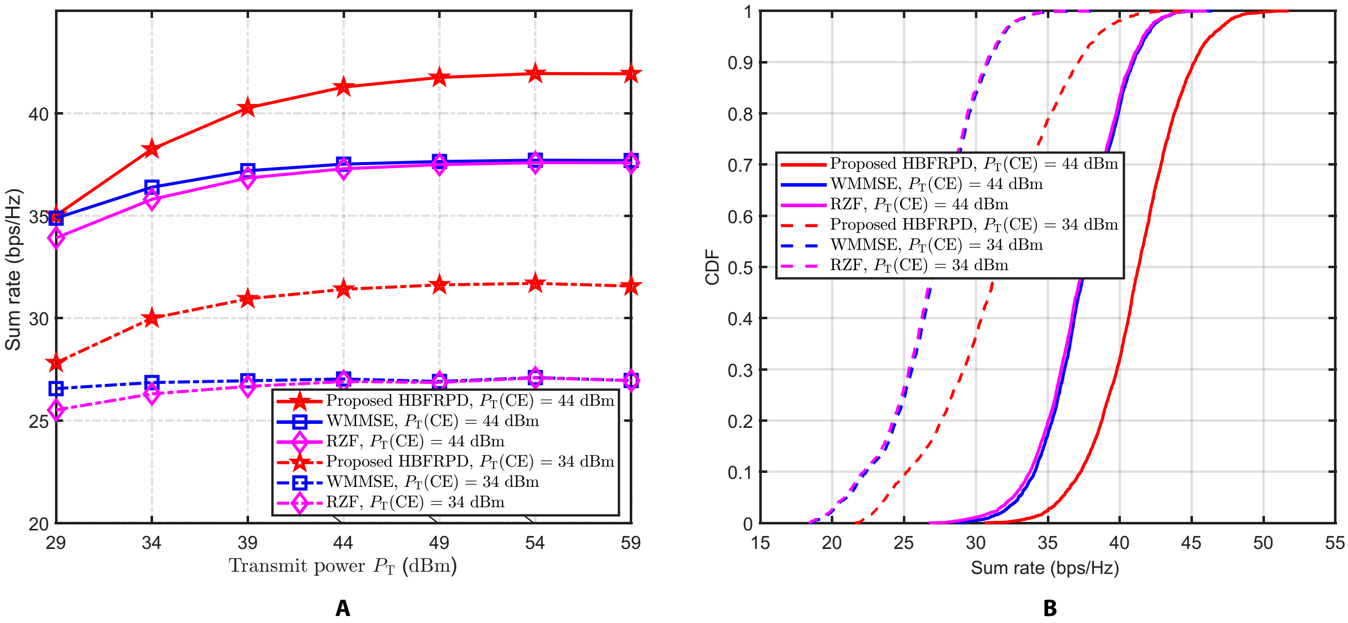

Fig. 14. Effectiveness verification of the proposed HBFRPD network under imperfect CSI. (A) Sum rates achieved by different schemes versus transmit power P_T under imperfect CSI. (B) CDFs of the sum rates achieved by different schemes under imperfect CSI, given $P_T = 44$ dBm. We set $\rho = 4$, ${}^{-}\rho = 16$, $L_p = 5$, and $M = 16$.

heads is $h = 8$, and the position-wise MLP sublayer has 2 fully connected layers with $4d_T$ and d_T neurons, respectively, while the dimension of the output linear layer is $2M$. In the subchannel estimation network, the linear layer is a $2N_s^R$-dimensional fully connected layer and the hyperparameters of the transformer encoder are the same as those of the CSI feedback network. Regarding the channel extrapolation network, the convolutional layer features a 7 × 7 kernel and 2 filters. The parameters of the rearranged operation are $L = 16$ and $N_p = 128$, and the number of neurons in the linear layer is $d_M = 512$. The number of mixer layers is $L_M = 6$, where each mixer layer consists of 2 MLP blocks, and the numbers of neurons in the MLP blocks are set to $2N_p$, N_p, $2d_M$, and d_M, respectively. The above structural parameters of the SFDCExtra network were empirically found to be appropriate.

The dataset was divided into 3 distinct subsets, namely the training, validation, and testing subsets, containing 102,400, 10,240, and 10,240 samples, respectively. Unless otherwise specified, simulations adopted the uniform ESS. When considering the learning-based ESS, the weight factor γ was 0.9. At the network training stage, the Adam optimizer was adopted to update the network weight parameters and the learning rate was varied depending on the *warmup* mechanism [49]. The batch size was set to 512 with 200 epochs.

HBFRPD network parameter configuration

We again determined the appropriate structural parameters of the HBFRPD network empirically. Specifically, in the RPDN, the linear embedding layer of the transformer encoder has $d_B = 128$ neurons. In the transformer encoder, the number of encoder layers is $L_B = 3$, the number of heads is $h = 8$, and the position-wise MLP sublayer has 2 fully connected layers with $4d_B$ and d_B neurons, respectively, while the output linear layer of the transformer encoder has $N^R / U = 64$ neurons. In the digital beamforming network, the hyperparameters of the transformer encoder are the same as those of the RPDN, and the output linear layer of the transformer encoder has $4U$ neurons.

We took each U channel samples as a group to form a dataset composed of 3 parts with sample sizes of 25,600, 2,560, and 2,560 respectively. The batch size was set to 32 with 180 epochs.

DL-based SFDCExtra

Given that the RIS element can only passively receive EM waves, selecting some elements of the RIS array would reduce the signal energy radiated into the room. For a fair comparison between different schemes, we adopted the same transmit power instead of the same SNR as the comparison criterion to avoid ignoring the performance differences induced by the number of activated RIS elements. Specifically, as illustrated in Fig. 8, we assessed the NMSE performance of the different schemes with different transmit power P_T. The number of NLoS paths was $L_p = 5$. We considered 3 model-based channel estimation benchmark algorithms, namely the SOMP algorithm [54], multiple-measurement-vector approximate message passing (MMV-AMP) algorithm [55], and model-driven MMV learned AMP (MMV-LAMP) network [56], and utilized $M = 64$ OFDM symbols on all subcarriers and then directly estimated the complete channel. For the SOMP and MMV-LAMP schemes, a redundant dictionary with an oversampling ratio of 4 was utilized to further improve the performance, i.e., the number of codewords was $G_d = 1{,}024$. However, for the MMV-AMP approach, the requirement of independent and identically distributed elements in the measurement matrix precludes the use of a redundant dictionary (i.e., $G_d = N^R = 256$). Given that data-driven DL algorithms have the potential to achieve better performance, we also compared the proposed DL-based SFDCExtra network with a transformer-based channel estimation network [50] and CNN-based channel extrapolation network [27]. For these methods, we set $M = 16$ OFDM symbols and $\overline{\rho} = 16$ as the subcarrier compression ratio. The transformer-based scheme turns on all RIS elements, i.e., $\rho = 1$, and directly estimates the complete channel. Both the CNN-based and proposed channel extrapolation schemes consider the element compression ratio of $\rho = 4$ to perform partial channel extrapolation.

Table. Computational complexity of different schemes

Channel estimation scheme	Complexity	FLOPs	Execution time (s)
SOMP	$\mathcal{O}\left(G_d KMI + G_d^2 KI\right)$	4.707 G	0.1130
MMV-AMP	$\mathcal{O}\left(MKN^{\mathrm{R}} I\right)$	5.337 G	0.7482
MMV-LAMP	$\mathcal{O}\left(MG_d KI\right)$	0.341 G	0.0689
Transformer-based channel estimation network	$\mathcal{O}\left(L_{\mathrm{T}}\left(Kd_{\mathrm{T}}^2 + K^2 d_{\mathrm{T}}\right)\right)$	0.362 G	0.0103
CNN-based channel extrapolation network	$\mathcal{O}\left(Z_5^2 N^{\mathrm{R}} KC_{32}^2\right)$	10.16 G	0.6039
Proposed	$\mathcal{O}\left(L_{\mathrm{M}}\left(N_p^2 d_{\mathrm{M}} + N_p d_{\mathrm{M}}^2\right)\right)$	1.066 G	0.0248
Beamforming scheme	Complexity	FLOPs	Execution time (s)
RZF	$\mathcal{O}\left(\left(2U\left(M^{\mathrm{B}}\right)^2 + \left(M^{\mathrm{B}}\right)^3\right)K\right)$	24.58 k	0.1389
WMMSE	$\mathcal{O}\left(IK\left(U^2\left(M^{\mathrm{B}}\right)^2 + U\left(M^{\mathrm{B}}\right)^3\right)\right)$	8.192 M	0.9184
Proposed	$\mathcal{O}\left(L_{\mathrm{B}} U\left(Kd_{\mathrm{B}}^2 + K^2 d_{\mathrm{B}}\right)\right)$	0.512 G	0.0616

Note that for the sake of fairness, the quantization of CSI feedback information was not considered for the above model- and data-driven algorithms. Therefore, we additionally considered the proposed scheme with $B = 256$ feedback bits generated via a 2-bit quantizer, denoted as "Proposed-Q".

It can be observed from Fig. 8 that the proposed channel extrapolation scheme notably outperforms the other schemes in terms of NMSE performance while imposing a smaller pilot overhead. This is because exploiting spatial-frequency correlations allows the proposed DL-based channel extrapolation scheme to recover the unobserved channel part from the estimated low-dimensional subchannel, thus reducing the training overhead while improving the NMSE performance. In particular, the proposed extrapolation scheme considerably improves the NMSE performance compared with the state-of-the-art CNN-based channel extrapolation scheme. Unlike local perception in CNNs, the MLP-Mixer is utilized as the backbone of the proposed channel extrapolation module to extract the global features of the channel for enhanced extrapolation accuracy. Considering the actual situation of finite quantized feedback, we can see that the proposed scheme with a 2-bit quantizer, "Proposed-Q", can still achieve very good performance. These results demonstrate that the proposed channel extrapolation scheme can accomplish high reconstruction performance while ensuring low pilot and feedback overheads.

We further investigated the robustness of the proposed DL-based channel extrapolation scheme with respect to the number of multipath components L_p in Fig. 9. The proposed DL-based channel extrapolation scheme was trained offline using channel samples that contained $L_p = 5$ multipath components. As depicted in Fig. 9, at the online estimation stage, the proposed scheme demonstrates its ability to estimate channels with different L_p without the need for retraining the entire network. Thus, the proposed scheme exhibits superior robustness and generalization capabilities in various channel conditions.

In Fig. 10, we evaluate the channel extrapolation NMSE performance of the proposed scheme with different numbers of pilot OFDM symbols, $M = 4, 8, 16, 32$, and 64. As expected, the channel extrapolation performance improves with the increase in the number of pilot OFDM symbols. This is because more pilot OFDM symbols can improve the accuracy of subchannel estimation, thus reducing the error propagation and improving the reconstruction of the extrapolation module. Furthermore, we can see that the proposed scheme can provide more considerable performance gain by increasing the number of pilot OFDM symbols in the case of low transmit power. This is because the increase in the number of observations can improve the received SNR.

Figure 11 depicts the NMSE performance of the proposed DL-based channel extrapolation scheme versus the element compression ratio ρ, with 3 ESEs. Specifically, the curve labeled as "Uniform" corresponds to the uniform selection strategy, the curve labeled as "Random" represents the random selection strategy, while the other 3 curves labeled as "DL-based with 200 epochs", "DL-based with 300 epochs", and "DL-based with 400 epochs" use the DL-based ESS. As expected, the NMSE improves as the element compression ratio ρ decreases. This is largely due to 2 reasons: (a) As the number of selected RIS elements increases, or the element compression ratio ρ decreases, the received signal power increases, thus improving the estimation accuracy of the channel extrapolation input (i.e., subchannel estimate); and (b) the received pilot signal can provide more channel information when more RIS elements are selected. However, this does not imply that we can obtain the best performance by choosing the lowest element compression ratio (or performing complete observations directly without extrapolation). Indeed, the channel extrapolation performance heavily depends on the amount of wireless communication transmission resources, the accuracy of the subchannel estimation, and the number of selected RIS elements (i.e., the dimension of the subchannel). Only when the transmission resources are sufficient can the gain provided by more selected RIS elements stand out. Moreover, we can observe that the performance gap between different element selection strategies is not evident at low compression ratios. However, at a high compression ratio (e.g., $\rho > 8$), the performance rank can be clearly seen to be "Uniform" < "Random" < "DL-based", which demonstrates the effectiveness of the proposed approach. Given that the aperture of the random pattern is statistically larger than that of the fixed uniform pattern, the random selection strategy is better than the uniform selection strategy, especially at a high compression ratio. The performance of the DL-based approach is better than that of the first 2 approaches after reaching a sufficient number of training epochs—specifically 300 epochs in this scenario, as the learning of the selection network requires more epochs to converge.

To fully illustrate the effectiveness of the proposed scheme, its channel extrapolation module was verified separately. To do so, we fixed the compression ratio of RIS elements at 4, i.e., $N_s^{\mathrm{R}} = 64$. First, least squares (LS), SOMP, and the proposed transformer-based algorithm were utilized for subchannel estimation; the results are shown in Fig. 12A. Note that the NMSE of the SOMP-based subchannel estimation with $M = 64$ pilot symbols is considerably better than that of the LS-based subchannel estimation with $M = 64$ pilot symbols, particularly at low transmit power P_{T}. Furthermore, the NMSE of the proposed transformer-based subchannel estimation algorithm with only $M = 16$ pilot symbols is considerably better than that of the SOMP-based subchannel estimation with $M = 64$ pilot symbols. Then, we inputted the subchannels estimated by different algorithms into the trained channel extrapolation network $f_{\mathrm{SFDE}}(\cdot)$, which outputs the estimation of the complete channel. The corresponding results are shown in Fig. 12B. Note that the NMSE performance of the complete channel extrapolated from the proposed channel extrapolation network is even better than the NMSE of the estimated low-dimensional subchannel, without any additional pilot overhead. This demonstrates that the proposed channel extrapolation network can not only be used for DL-based communication architectures but also be combined with traditional algorithms to considerably reduce resource overhead. Therefore, we conclude that the proposed DL-based SFDCExtra scheme can learn a latent mapping among channel elements to considerably reduce the pilot overhead while achieving the same or better channel estimation performance.

DL-based HBFRPD

Figure 13 shows the sum rates of total UEs achieved by different schemes assuming perfect CSI. We considered 2 comparison schemes, both of which adopt the analog beamforming design discussed above and the proposed beam alignment-based RIS phase design. In the latter design, the beam of each subarray is aligned to the corresponding associated UE. For digital beamforming design, these 2 comparison schemes adopt the RZF and iterative WMMSE algorithms, respectively; thus, they are abbreviated as "RZF" and "WMMSE", respectively. Note that the proposed HBFRPD scheme performs better than other schemes and its superiority is more evident as the transmit power increases. In addition, another advantage of the proposed HBFRPD scheme is that it does not require $\mathbf{F}_{\mathrm{BB}}[k], \forall k$, in an iterative manner. Thus, it runs much faster than the iterative WMMSE algorithm. We also analyzed the performance gain provided by the LoS MIMO architecture. Considering the case without LoS MIMO array structure (i.e., both the BS and RIS use conventional UPA arrays), the BS-RIS channel is a single LoS path with rank 1, which only provides single stream data transmission. To ensure a fair comparison, the transmit power in the absence of LoS MIMO was set equal to that with LoS MIMO, i.e., the transmit power in the absence of LoS MIMO was actually $UP_T = 4P_T$. By calculating the sum rate, we obtained the green curve presented in Fig. 13. Note that the sum rate with LoS MIMO is much higher than that without LoS MIMO. This is because the LoS MIMO architecture can increase the sum rate linearly leveraging the extra spatial multiplexing gain, while the conventional array architecture can only provide log-level growth as the SINR increases.

Although most schemes can achieve good sum rate performance under perfect CSI, the sum rate of multiusers is degraded due to interuser interference induced by CSI error. Figure 14A shows the sum rate performance of the different schemes with imperfect CSIs estimated at 2 different transmit powers P_{T}(CE). Compared with the case of perfect CSI, the sum rate degrades considerably with the decrease in CSI estimation accuracy, i.e., with the decrease in transmit power at the channel estimation stage. It can be clearly seen that due to the interuser interference induced by CSI errors, the sum rates of the RZF and iterative WMMSE schemes barely increase with transmit power. Moreover, the proposed HBFRPD scheme exhibits a considerable performance gain over the RZF and iterative WMMSE algorithms in the presence of CSI estimation errors. This result indicates that the proposed scheme can mitigate the interference caused by CSI errors and hence is more robust to inaccurate CSI than the other schemes.

The cumulative distribution functions (CDFs) characterizing the sum rate performance achieved by the different schemes are shown in Fig. 14B. Here, we consider a transmit power of $P_{\mathrm{T}} = 44$ dBm at the data transmission stage. Figure 14B shows that when the transmit power is $P_{\mathrm{T}}(\mathrm{CE}) = 34$ dBm at the channel estimation stage, the proposed HBFRPD network has a probability of approximately 64.6% to achieve a sum rate exceeding 30 bps/Hz, while the other 2 schemes have a probability of only 16.3% to achieve such a rate. When the transmit power is $P_{\mathrm{T}}(\mathrm{CE}) = 44$ dBm at the channel estimation stage, the proposed HBFRPD network has a probability of approximately 68.8% to achieve a sum rate exceeding 40 bps/Hz, which is considerably higher than the other 2 schemes. This result confirms the superior performance of the proposed HBFRPD network over existing conventional schemes.

Computational complexity analysis

A computational complexity analysis of different schemes at the inference stage is presented in the Table. The numerical results were obtained on a PC with Intel Core i9-10980XE CPU @ 3.00GHz and an Nvidia GeForce RTX 3090 GPU. The DL-based methods and existing solutions were implemented on the PyCharm framework. The details are further elaborated next.

1. Channel estimation schemes: In the SOMP algorithm [54], the correlation operation creates considerable computational complexity, where I is the number of iterations. The MMV-AMP algorithm [55] mainly requires matrix multiplication operations, but a large number of iterations I increases its computational complexity. The MMV-LAMP algorithm [56] has a low computational complexity because DL reduces the required number of iterations. The transformer-based channel estimation network [50] also has a low computational complexity, and its main sources of computational complexity come from self-attention and MLP sublayers. In the CNN-based channel extrapolation network [27], convolutional layers introduce considerable computational complexity. By contrast, the MLP-Mixer layers provide the majority of the computational complexity in the proposed SFDCExtra network, and the level of complexity is much lower than that of the CNN-based channel extrapolation network. We further meticulously

counted the number of floating-point operations per second (FLOPs) and measured the execution time per sample on a CPU for different schemes. The results are presented in the Table. Observe that at the inference stage, the FLOPs and execution time per sample of the proposed scheme are lower than those of most benchmarks. Specifically, the SFDCExtra network requires the second-lowest execution time per sample, and only MMV-LAMP and transformer-based channel estimation network have lower FLOPs than the proposed scheme.

2. Beamforming schemes: The matrix inversion required in the RZF algorithm is its main source of computational complexity. In the iterative WMMSE algorithm [52], a large number of iterations increases the computational complexity and execution time per sample. In the proposed DL-based HBFRPD network, self-attention and MLP sublayers create higher computational complexity and FLOPs than the other 2 algorithms. However, the execution time per sample of the proposed scheme is considerably lower than that of the 2 model-based schemes. This is because the DL-based HBFRPD network only needs matrix multiplication operations and does not require an iterative procedure. This is a superior advantage of the proposed DL-based HBFRPD network.

Conclusion

This study proposed a DL-based transmission architecture for RIS-aided THz massive MIMO systems over hybrid-field channels. The contributions of this study are a channel estimation scheme with low pilot overhead and a robust beamforming scheme. More specifically, an E2E DL-based channel estimation framework that consists of a pilot design, a CSI feedback, subchannel estimation, a and channel extrapolation was developed, and to maximize the sum rate of all UEs under imperfect CSI, a DL-based scheme to simultaneously design the hybrid beamforming and RIS phase was formulated. Simulation results show that the proposed channel extrapolation scheme considerably outperformed the existing state-of-the-art schemes in terms of reconstruction performance while imposing a notably reduced pilot overhead. Moreover, the results demonstrate that the proposed beamforming scheme is superior to the existing designs in terms of achievable sum rate performance and robustness to imperfect CSI. Based on these results, potential future research directions include the development of a practical discrete phase shifter, the analysis of complex near-field channels, and the enhancement of sensing-aided communications.

Acknowledgments

We thank Y. Zhang, Z. Wan, and others for their contributions.
Funding: This work was supported by the Natural Science Foundation of China (NSFC) (Grant 62071044), the Shandong Province Natural Science Foundation (Grant ZR2022YQ62), and the Beijing Nova Program.
Author contributions: Y.W., Z.G., and S.C. contributed to the algorithms and analysis. Y.W. and Z.G. prepared the manuscript; all the authors contributed to the drafting of the manuscript. C.H. and D.Z. initiated the presented concept and supervised the research.
Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

Data Availability

Data are available from the corresponding author on reasonable request.

References

1. Elayan H, Amin O, Shihada B, Shubair RM, Alouini M-S. Terahertz band: The last piece of RF spectrum puzzle for communication systems. *IEEE Open J Commun Soc*. 2020;1:1–32.
2. Lin C, Li GY. Indoor Terahertz communications: How many antenna arrays are needed? *IEEE Trans Wirel Commun*. 2015;14(6):3097–3107.
3. Sohrabi F, Yu W. Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE Open J Commun Soc*. 2016;10(3):501–513.
4. Di Renzo M, Zappone A, Debbah M, Alouini M-S, Yuen C, de Rosny J, Tretyakov S. Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead. *IEEE J Sel Areas Commun*. 2020;38(11):2450–2525.
5. Huang C, Hu S, Alexandropoulos GC, Zappone A, Yuen C, Zhang R, Di Renzo M, Debbah M. Holographic mimo surfaces for 6g wireless networks: Opportunities, challenges, and trends. *IEEE Wirel Commun*. 2020;27(5):118–125.
6. Renzo MD, Debbah M, Phan-Huy D-T, Zappone A, Alouini M-S, Yuen C, Sciancalepore V, Alexandropoulos GC, Hoydis J, Gacanin H, et al. Smart radio environments empowered by reconfigurable ai meta-surfaces: An idea whose time has come. *EURASIP J Wirel Commun Netw*. 2019;2019(1):1–20.
7. Huang C, Zappone A, Alexandropoulos GC, Debbah M, Yuen C. Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Trans Wirel Commun*. 2019;18(8):4157–4170.
8. Alexandropoulos GC, Stylianopoulos K, Huang C, Yuen C, Bennis M, Debbah M. Pervasive machine learning for smart radio environments enabled by reconfigurable intelligent surfaces. *Proc IEEE*. 2022;110(9):1494–1525.
9. Wu M, Gao Z, Huang Y, Xiao Z, Ng DWK, Zhang Z. Deep learning-based rate-splitting multiple access for reconfigurable intelligent surface-aided tera-hertz massive mimo. *IEEE J Sel Areas Commun*. 2023;41(5):1431–1451.
10. Selvan KT, Janaswamy R. Fraunhofer and fresnel distances: Unified derivation for aperture antennas. *IEEE Antennas Propag Mag*. 2017;59(4):12–15.
11. Cui M, Dai L. Channel estimation for extremely large-scale MIMO: Far-field or near-field? *IEEE Trans Commun*. 2022;70(4):2663–2677.
12. Yan L, Chen Y, Han C, Yuan J. Joint inter-path and intra-path multiplexing for Terahertz widely-spaced multi-subarray hybrid beamforming systems. *IEEE Trans Commun*. 2022;70(2):1391–1406.
13. Mishra D, Johansson H. Channel estimation and low-complexity beamforming design for passive intelligent surface assisted miso wireless energy transfer. Paper presented at: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019 May 12–17; Brighton, UK.
14. Wang P, Fang J, Duan H, Li H. Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems. *IEEE Signal Process Lett*. 2020;27:905–909.
15. Wei L, Huang C, Alexandropoulos GC, Yuen C, Zhang Z, Debbah M. Channel estimation for RIS-empowered multi-user MISO wireless communications. *IEEE Trans Commun*. 2021;69(6):4144–4157.
16. Elbir AM, Papazafeiropoulos A, Kourtessis P, Chatzinotas S. Deep channel learning for large intelligent surfaces aided

mm-wave massive MIMO systems. *IEEE Wirel Commun Lett*. 2020;9(9):1447–1451.
17. Liu S, Gao Z, Zhang J, Renzo MD, Alouini M-S. Deep denoising neural network assisted compressive channel estimation for mmwave intelligent reflecting surfaces. *IEEE Trans Veh Technol*. 2020;69(8):9223–9228.
18. An J, Xu C, Wu Q, Ng DWK, Di Renzo M, Yuen C, Hanzo L. Codebook-based solutions for reconfigurable intelligent surfaces and their open challenges. *IEEE Wirel Commun*. 2022;1–8.
19. Wei L, Huang C, Guo Q, Yang Z, Zhang Z, Alexandropoulos GC, Debbah M, Yuen C. Joint channel estimation and signal recovery for ris-empowered multiuser communications. *IEEE Trans Commun*. 2022;70(7):4640–4655.
20. Liu M, Li X, Ning B, Huang C, Sun S, Yuen C. Deep learning-based channel estimation for double-ris aided massive mimo system. *IEEE Wirel Commun Lett*. 2022;12(1):70–74.
21. Wan Z, Gao Z, Alouini MS. Broadband channel estimation for intelligent reflecting surface aided mmwave massive mimo systems. Paper presented at: ICC 2020-2020 IEEE International Conference on Communications (ICC); 2020 Jun 7–11; Dublin, Ireland.
22. Mashhadi MB, Gündüz D. Pruning the pilots: Deep learning-based pilot design and channel estimation for mimo-ofdm systems. *IEEE Trans Wirel Commun*. 2021;20(10):6315–6328.
23. Wan Z, Gao Z, Gao F, Di Renzo M, Alouini M-S. Terahertz massive mimo with holographic reconfigurable intelligent surfaces. *IEEE Trans Commun*. 2021;69(7):4732–4750.
24. Zhang S, Liu Y, Gao F, Xing C, An J, Dobre OA. Deep learning based channel extrapolation for large-scale antenna systems: Opportunities, challenges and solutions. *IEEE Wirel Commun*. 2021;28(6):160–167.
25. Lin B, Gao F, Zhang S, Zhou T, Alkhateeb A. Deep learning-based antenna selection and CSI extrapolation in massive MIMO systems. *IEEE Trans Wirel Commun*. 2021;20(11):7669–7681.
26. Xu M, Zhang S, Zhong C, Ma J, Dobre OA. Ordinary differential equation-based CNN for channel extrapolation over RIS-assisted communication. *IEEE Commun Lett*. 2021;25(6):1921–1925.
27. Zhang S, Zhang S, Gao F, Ma J, Dobre OA. Deep learning-based RIS channel extrapolation with element-grouping. *IEEE Wirel Commun Lett*. 2021;10(12):2644–2648.
28. Ying K, Gao Z, Lyu S, Wu Y, Wang H, Alouini MS. GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-Wave massive MIMO. *IEEE Access*. 2020;8:19530–19539.
29. Di B, Zhang H, Song L, Li Y, Han Z, Poor HV. Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts. *IEEE J Sel Areas Commun*. 2020;38(8):1809–1822.
30. Ahn Y, Shim B. Deep learning-based beamforming for intelligent reflecting surface-assisted mmWave systems. Paper presented at: 2021 International Conference on Information and Communication Technology Convergence (ICTC); 2021 Oct 20–22; Jeju Island, Korea.
31. Pradhan C, Li A, Song L, Vucetic B, Li Y. Hybrid precoding design for reconfigurable intelligent surface aided mmWave communication systems. *IEEE Wirel Commun Lett*. 2020;9(7):1041–1045.
32. Zhang S, Zhang H, Di B, Tan Y, Han Z, Song L. Beyond intelligent reflecting surfaces: Reflective-transmissive metasurface aided communications for full-dimensional coverage extension. *IEEE Trans Veh Technol*. 2020;69(11):13905–13909.
33. Youn Y, Lee C, Kim D, Chang S, Hwang M, Jun D, Chae C-B, Hong W. Demo: Transparent intelligent surfaces for sub-6 GHz and mmWave B5G/6G systems. Paper presented at: 2022 IEEE International Conference on Communications Workshops (ICC Workshops); 2022 May 16–20; Seoul, Korea.
34. Kitayama D, Hama Y, Goto K, Miyachi K, Motegi T, Kagaya O. Transparent dynamic metasurface for a visually unaffected reconfigurable intelligent surface: Controlling transmission/reflection and making a window into an RF lens. *Opt Express*. 2021;29(18):29292–29307.
35. Chen Y, Yan L, Han C. Hybrid spherical- and planar-wave modeling and DCNN-powered estimation of Terahertz ultra-massive MIMO channels. *IEEE Trans Commun*. 2021;69(10):7063–7076.
36. Wang X, Lin Z, Lin F, Hanzo L. Joint hybrid 3D beamforming relying on sensor-based training for reconfigurable intelligent surface aided TeraHertz-based multiuser massive MIMO systems. *IEEE Sensors J*. 2022;22(14):14540–14552.
37. Hong S, Pan C, Ren H, Wang K, Chai KK, Nallanathan A. Robust transmission design for intelligent reflecting surface-aided secure communication systems with imperfect cascaded CSI. *IEEE Trans Wirel Commun*. 2021;20(4):2487–2501.
38. Chen Z, Tang J, Zhang XY, Wu Q, Chen G, Wong K-K. Robust hybrid beamforming design for multi-RIS assisted MIMO system with imperfect CSI. *IEEE Trans Wirel Commun*. 2022;22(6):3913–3926.
39. Xu W, Gan L, Huang C. A robust deep learning-based beamforming design for RIS-assisted multiuser MISO communications with practical constraints. *IEEE Trans Cogn Commun Netw*. 2022;8(2):694–706.
40. Larsson P. Lattice array receiver and sender for spatially orthonormal MIMO communication. Paper presented at: 2005 IEEE 61st Vehicular Technology Conference; 2005 May 30–Jun 1; Stockholm, Sweden.
41. Bohagen F, Orten P, Oien GE. Optimal design of uniform planar antenna arrays for strong line-of-sight MIMO channels. Paper presented at: 2006 IEEE 7th Workshop on Signal Processing Advances in Wireless Communications; 2006 Jul 2–5; Cannes, France.
42. Song X, Rave W, Babu N, Majhi S, Fettweis G. Two-level spatial multiplexing using hybrid beamforming for millimeter-wave backhaul. *IEEE Trans Wirel Commun*. 2018;17(7):4830–4844.
43. Yan L, Han C, Yuan J. Energy-efficient dynamic-subarray with fixed true-time-delay design for Terahertz wideband hybrid beamforming. *IEEE J Sel Areas Commun*. 2022;40(10):2840–2854.
44. Han C, Bicen AO, Akyildiz IF. Multi-ray channel modeling and wideband characterization for wireless communications in the Terahertz band. *IEEE Trans Wirel Commun*. 2015;14(5):2402–2412.
45. Wu Y, Kokkoniemi J, Han C, Juntti M. Interference and coverage analysis for Terahertz networks with indoor blockage effects and line-of-sight access point association. *IEEE Trans Wirel Commun*. 2021;20(3):1472–1486.
46. Jornet JM, Akyildiz IF. Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the Terahertz band. *IEEE Trans Wirel Commun*. 2011;10(10):3211–3221.

47. Wang B, Gao F, Jin S, Lin H, Li GY. Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems. *IEEE Trans Signal Process*. 2018;66(13):3393–3406.
48. Wen C-K, Shih W-T, Jin S. Deep learning for massive MIMO CSI feedback. *IEEE Wirel Commun Lett*. 2018;7(5):748–751.
49. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser LU, Polosukhin I. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. *Advances in neural information processing systems*. Curran Associates, Inc.; 2017.
50. Wang Y, Gao Z, Zheng D, Chen S, Gunduz D, Poor HV. Transformer-empowered 6g intelligent networks: From massive MIMO processing to semantic communication. *IEEE Wirel Commun*. 2022;1–9.
51. Tolstikhin IO, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, Yung J, Steiner A, Keysers D, Uszkoreit J, et al. MLP-mixer: An all-MLP architecture for vision. *Advances in neural information processing systems*; 2021. p. 24261–24272.
52. Shi Q, Razaviyayn M, Luo Z-Q, He C. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Trans Signal Process*. 2011;59(9):4331–4340.
53. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput*. 1989;1(2):270–280.
54. Tropp JA, Gilbert AC, Strauss MJ. Algorithms for simultaneous sparse approximation. Part i: Greedy pursuit. *Signal Process*. 2006;86(3):572–588.
55. Ke M, Gao Z, Wu Y, Gao X, Schober R. Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO. *IEEE Trans Signal Process*. 2020;68:764–779.
56. Ma X, Gao Z, Gao F, Di Renzo M. Model-driven deep learning based channel estimation and feedback for millimeter-wave massive hybrid MIMO systems. *IEEE J Sel Areas Commun*. 2021;39(8):2388–2406.