Special Topic: Integration of Large AI Model and 6G

Enhanced channel estimation for near-field IRS-aided multi-user MIMO system via a large deep residual network[†]

Yan WANG[1], Yongqiang LI[1], Minghao CHEN[1], Yu YAO[1],
Feng SHU[1,2,3*] & Jiangzhou WANG[4]

[1]School of Information and Communication Engineering, Hainan University, Haikou 570228, China
[2]School of Information Science and Technology, Xizang University, Lhasa 850000, China
[3]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[4]School of Engineering, University of Kent, Canterbury CT2 7NT, UK

Abstract In this paper, the channel estimation (CE) of intelligent reflecting surface-aided near-field (NF) multi-user communication is investigated. Initially, the least square (LS) estimator and minimum mean square error (MMSE) estimator for the estimated channel are designed, their mean square errors (MSEs) are derived, and the Cramér-Rao lower bound (CRLB) is derived to serve as a benchmark for performance evaluation. Subsequently, in view of the fact that the NF channel model is more sensitive to distance variations compared to the far-field model, this leads to pronounced discrepancies in the user channel characteristics in different regions. To effectively capture and utilize these diverse channel features, users are initially divided into distinct regions predicated on pivotal parameters, such as channel angle and distance. Correspondingly, a user region classifier based on convolutional neural networks is designed. Then, to fully harness the potential of deep residual networks (DRNs) in denoising, the aforementioned CE problem is reconceptualized as a denoising task, and a DRN-driven single region NF CE network, named SR-DRN-NFCE, is proposed. In addition, by integrating SR-DRN-NFCE networks corresponding to different regions and conducting joint training in a federated learning (FL) manner, a new network is obtained, named FL-DRN-NFCE. Simulation results demonstrate that the proposed FL-DRN-NFCE network outperforms LS, MMSE, and no residual connections in terms of MSE, and the proposed FL-DRN-NFCE method has higher CE accuracy than the SR-DRN-NFCE method.

Keywords deep residual network, channel estimation, intelligent reflecting surface, near-field communication, multi-user

1 Introduction

Intelligent reflecting surfaces (IRSs), classified as either passive [1] or active [2], can accurately transmit a signal, thereby improving the coverage and communication efficiency of that signal [3]. However, the realization of this advantage relies on the precise acquisition of channel state information (CSI) [4]. In IRS-assisted wireless scenarios, the introduction of IRSs increases the complexity and uncertainty of signal propagation paths [5], making traditional channel estimation (CE) methods difficult to apply directly. Consequently, how to efficiently and accurately estimate the complete CSI, including the influence of IRS, has become an urgent and challenging core task in the research field of IRS-aided wireless technology [6].

1.1 Prior work

A diversity of new algorithms and techniques have been investigated by scholars for the CE problem of IRS-aided communication networks [7]. For example, three channel estimators were developed by Wang et al. [8] for the uplink CE of IRS-assisted multi-user (MU) massive multiple-input single-output (MISO)

systems. Specifically, the first estimator cleverly utilized the element sparsity of cascaded channels for optimization; the second estimator delved into the sparsity of the common row structure inherent in the IRS to base station (BS) channel; finally, the third estimator further integrated the sparsity of common row and column structures induced by the limited scattering environment around the user, thereby achieving more refined CE. In addition, a Bayesian CE method for IRS-enabled millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) networks was proposed by Kim et al. [9]. Numerical analysis indicated that the introduction of active sensors significantly reduced the needed training cost of the designed CE algorithm. Subsequently, a two-stage strategy was proposed by Chen et al. [10] for estimating the sensing and communication channels in a self-sensing IRS-aided mmWave integrated sensing and communication (ISAC) network. Experimental results verified the availability of the constructed two-stage scheme. In 2022, two novel anchor-assisted CE methods were investigated by Guan et al. [11], and the results showed that scheme I was most suitable when the BS had a significant quantity of antennas, whereas scheme II was more effective otherwise. In addition to focusing on researching high-performance and low-complexity CE algorithms, a productive transmission protocol for CE and beam tracking was developed by Zheng et al. [12], which independently processed satellites as well as users in a distributed way, greatly lessening the complexity of implementation.

In addition, the inherent transceiver hardware impairments (HWIs) inevitably introduce additional noise and distortion, causing the training pilot signal and received signal to deviate from their ideal states, which in turn substantially increases the complexity of CE. The inaccuracy of CE results will further adversely affect the system performance [13]. Therefore, the impact of HWI must be fully accounted for in the design of CE algorithms. To achieve this objective, it may be necessary to integrate HWI models into the estimation process or adopt compensation techniques to effectively mitigate their negative effects [14]. Concurrently, the presence of HWI issues has motivated academic research toward exploring more robust CE approaches. For instance, Li et al. [15] theoretically formulated the ergodic sum-rate of the IRS-assisted non-orthogonal multiple access (NOMA) uplink in the face of CE errors and HWI. Specifically, the IRS phase shift was configured based on the statistical CSI, followed by the linear minimum mean square error (LMMSE) CE of the equivalent channel spanning from the user to the BS.

Technically, traditional CE methods mainly rely on signal processing techniques and statistical models, such as least square (LS) CE and minimum mean square error (MMSE) CE. These methods can estimate the channel state to a certain extent, but they are often limited by the complexity of the model and the limitations in prediction accuracy. In particular, in complex and changeable wireless environments, the prediction accuracy and adaptability of traditional methods are often difficult to comply with the realistic demand.

Recently, deep learning (DL), as one of the core branches in the domain of artificial intelligence, has made extraordinary breakthroughs in two key areas, namely, image recognition and natural language processing, by virtue of its excellent ability in automatic feature extraction and complex pattern recognition. The application of DL in the area of CE is expected to break through the drawbacks of traditional approaches and enhance the correctness and adaptability of CE. For instance, Ye et al. [16] elaborated a conditional generative adversarial network (cGAN) architecture that could effectively estimate the characteristics of cascaded channels exploiting the received signal as a conditional variable. The experimental analysis fully verified that the cGAN approach exhibited excellent robust performance in the context of IRS-assisted communication networks. Subsequently, a graph neural network (GNN)-based CE algorithm was proposed by Ye et al. [17], which could maintain good performance with low pilot overhead. Moreover, existing neural network-based approaches typically involve the manual design of network architectures through a trial-and-error process, demanding extensive domain knowledge and human resources. Therefore, an automated method for constructing a high-performance neural network architecture for CE was designed by Shi et al. [18], which provided the most accurate estimation of CSI over other benchmark CE methods. Specifically, Chu et al. [19] explored an adaptive and robust estimator for time-varying mmWave channels, considering the use of more general arbitrarily shaped IRS, in contrast to classical regular IRS elements arranged on a grid. To reduce the complexity, a DL-based solver using fixed-point iteration and a cascaded DL network framework was studied. The proposed DL-based method exhibits lower complexity when compared to the alternating optimization (AO)-based approach, making it suitable for highly dynamic communication scenarios. In addition, as a critical information hub bridging CSI acquisition and network parameter optimization, the transmission efficiency and reliability of CSI feedback play a pivotal role in determining the overall performance of communication systems. To address this, a decoupled representation learning-enabled neural network was proposed by Xu et al. [20], which

leveraged the uplink CSI characteristics to achieve selective compression feedback.

However, most of the aforementioned CE approaches are based on assumptions of far-field (FF) channel models [21]. This is because traditional wireless communication networks, i.e., the first generation (1G) to the fifth generation (5G), rely heavily on the spectrum below 6 GHz or even below 3 GHz. Due to wavelength limitations, these networks typically employ smaller antenna arrays. The combination of low-dimensional antenna arrays and lower frequencies typically limits the range of wireless near-field (NF) communications to a few meters, or even a few centimeters [22]. Therefore, the depiction of traditional wireless communication networks is commonly based on FF assumptions. The 6G network architectures are expected to introduce larger antenna aperture designs and incorporate high frequency band resources, e.g., centimeter-wave, mmWave, and even terahertz, which will considerably enhance the NF effect characteristics of the network. The fusion application of emerging manners such as IRS [23], extra large-scale MIMO (XL-MIMO), and cell-free MIMO, predicts that NF scenarios will become increasingly prevalent in future wireless communication. Consequently, the traditional assumption based on FF plane waves will no longer be the dominant analytical framework, and new theoretical channel models need to be sought to adapt to this change. As a result, more and more researchers have embarked on relevant studies focusing on NF communication scenarios.

Characterization and estimation algorithms for NF sparse channels have become the focus of extensive academic attention. Specifically, Pisharody et al. [24] innovatively proposed two algorithms aimed at accurately estimating the NF uplink channel in XL-MIMO systems. These algorithms cleverly exploited the spatial non-stationary characteristics unique to XL-MIMO channels, although the implementation was accompanied by relatively high computational complexity. In a follow-up study, a low-complexity sequential angle-distance CE (SADCE) scheme for NF XL-MIMO networks furnished with uniform planar arrays (UPAs) was designed by Huang et al. [25]. Notably, the prominent energy leakage effect in NF XL-MIMO channels constituted one of the key factors affecting the CE performance. To cope with the difficult problem of compressed CE in XL-MIMO models, Guo et al. [26] innovatively proposed a triple parametric decomposition (TPD) structure. The simulation verification showed that this TPD skeleton considerably refined the system performance over the current state-of-the-art technology. In particular, an efficient model-based DL algorithm was investigated by Zhang et al. [27] to estimate NF wireless channels for XL-MIMO communications. This algorithm constructed a sparse dictionary model based on spatial grids, which transformed the NF CE task in XL-MIMO systems into an optimization problem under the compressed sensing (CS) framework. Subsequently, the learning iterative shrinkage and thresholding algorithm (LISTA) was used to efficiently solve the CS problem.

However, the existing NF CE methods are mostly focused on XL-MIMO systems without IRS assistance, whereas corresponding CE methods for IRS-aided NF communication environments have also been explored and investigated in depth. A low complexity CE strategy based on the assumption of NF spherical wavefront was designed by Yang et al. [28]. To decrease the pilot overhead and computational complexity, this strategy simplified the CE process into two stages. The first stage focused on the angle domain variable estimation between the BS and users employing the hybrid beamforming structure, whereas the second stage further estimated the cascaded angle domain and polar domain factors of XL-IRS. However, in this NF channel model, only the line-of-sight (LoS) component was considered, while the non-line-of-sight (NLoS) component was ignored.

1.2 Our contributions

Motivated by the aforementioned discussions, deep residual network (DRN)-based CE algorithms will be investigated for the IRS-aided MU NF communication. To the best of our knowledge, there is relatively limited research on the application of DRN for CE in IRS-assisted MU NF communication systems, especially considering both direct and cascaded channels containing LoS and NLoS components. The core contributions of this paper are condensed into the following succinct summary.

(1) To begin with, an IRS-assisted MU NF communication system is constructed. For the NF channel model, unlike the work by Yang et al. [28], which solely focused on the LoS component, the NLoS component is taken into account in this paper to offer a more holistic portrayal of the NF channel characteristics. Moreover, the CE problem for both direct and cascaded channels is concurrently investigated in this paper, in contrast to the prevalent focus solely on cascaded CE in most studies. Subsequently, the LS estimators for direct and cascaded channels are designed. Correspondingly, the closed-form expression for the mean square error (MSE) of the LS estimator is derived, and the Cramér-Rao lower bound (CRLB)

is derived to serve as a benchmark for performance evaluation.

(2) Considering the heightened sensitivity of NF channel models to distance variations compared to FF models, this leads to significant differences in channel characteristics among users across distinct regions. To capture and utilize these diversified channel characteristics, users are partitioned into distinct regions based on key parameters such as channel angle and distance. Accordingly, a user region classifier (RC) based on convolutional neural networks is designed. Simulation results show that the accuracy of the proposed RC approximates 95% in the high signal-to-noise ratio (SNR) range.

(3) The CE problem in IRS-assisted MU NF communication systems is re-envisioned as a denoising task, and a DRN-driven single region NF CE, namely, SR-DRN-NFCE network, is designed by leveraging the denoising advantages of DRN. Then, we propose an innovative CE framework that integrates SR-DRN-NFCE networks from different regions through collaborative training based on federated learning (FL), ultimately developing a novel network called FL-DRN-NFCE. Simulation results reveal that the proposed FL-DRN-NFCE network exhibits a lower level of MSE over the scheme without residual connections. In addition, compared with traditional methods such as LS and MMSE, the proposed FL-DRN-NFCE algorithm successfully reduces pilot overhead by five-sixths through the powerful feature extraction capabilities of DL technology.

1.3 Organization and notation

The remainder of this paper is structured as follows. In Section 2, the system model of an IRS-aided MU NF communication is established. The CRLB is derived in Section 3. In Section 4, the RC, and the FL-DRN-NFCE network are proposed, and the computational complexity of the aforementioned networks is also given. Finally, the simulation results and conclusion are presented in Sections 5 and 6, respectively.

Notations. Throughout this paper, matrices are denoted by boldface uppercase letter (e.g., $\boldsymbol{A}$), vectors by boldface lowercase letter (e.g., $\boldsymbol{a}$), and scalars by lowercase letter, (e.g., a). Moreover, $\mathbb{C}^{a \times b}$ stands for the space of $a \times b$ matrices with complex entries. The real part is represented by $\mathfrak{Re}\{\cdot\}$, whereas the imaginary part is represented by $\mathfrak{Im}\{\cdot\}$. In addition, the sign $(\cdot)^{\mathrm{T}}$ denotes the transpose operation, whereas $(\cdot)^{\mathrm{H}}$ stands for the conjugate and transpose operation. The signs $\|\cdot\|_2$ and $\|\cdot\|_F$ stand for the 2-norm and F-norm, respectively. The notation $\mathbb{E}\{\cdot\}$ denotes the expectation operation, $\hat{[\ ]}$ represents the estimation operation, and $\mathrm{tr}\{\cdot\}$ is the trace operation. Furthermore, the identity matrix is denoted by $\boldsymbol{I}$, whereas the Kronecker and Hadamard products are respectively denoted by the symbols $\otimes$ and $\odot$. In addition, $\mathrm{diag}\{\boldsymbol{a}\}$ yields a diagonal matrix, where the primary diagonal elements are passively set to the components of $\boldsymbol{a}$. Finally, $\lfloor \cdot \rfloor$ indicates rounding down.

2 System model

As sketched in Figure 1, an MU NF wireless communication system, aided by an IRS, is taken into account, with the IRS being strategically positioned between the BS and single-antenna users for communication augmentation. The BS is furnished with an M-antenna array, while the IRS is composed of N reflecting elements, both of which are set up in a UPA configuration. In MU communication systems, as described in [29], a single user is often limited to information about a particular channel scenario, which leads to the possibility that neural network models trained based on the dataset collected by a single user may fail when the user crosses different regions. In order to overcome this limitation and ensure that the neural network model performs well under a wide range of channel conditions, the whole cell is elaborately divided into T independent regions, where $t = 1, 2, \ldots, T$. For the convenience of subsequent analysis, the count of users in the t-th region is explicitly defined as K_t. In particular, users in the same domain demonstrate highly similar channel characteristics, on the contrary, users distributed in distinct regions exhibit differentiated channel characteristics. Most importantly, when users migrate between regions, their channel characteristics may undergo notable changes, which puts forward higher demands for the design of neural network-based CE algorithms with high robustness.

Let $\boldsymbol{G} \in \mathbb{C}^{N \times M}$, $\boldsymbol{f}_{t,k} \in \mathbb{C}^{1 \times N}$, and $\boldsymbol{h}_{t,k} \in \mathbb{C}^{1 \times M}$ stand for the channels from BS to IRS, IRS to the k-th user in the t-th region, and BS to the k-th user in the t-th region, respectively, where $k = 1, 2, \ldots, K_t$. The received signal at the k-th user positioned in the t-th region can be modeled as follows:

$$y_{t,k} = \boldsymbol{h}_{t,k}\boldsymbol{v}x + \boldsymbol{f}_{t,k}\boldsymbol{\Theta}\boldsymbol{G}\boldsymbol{v}x + w_{t,k}, \tag{1}$$
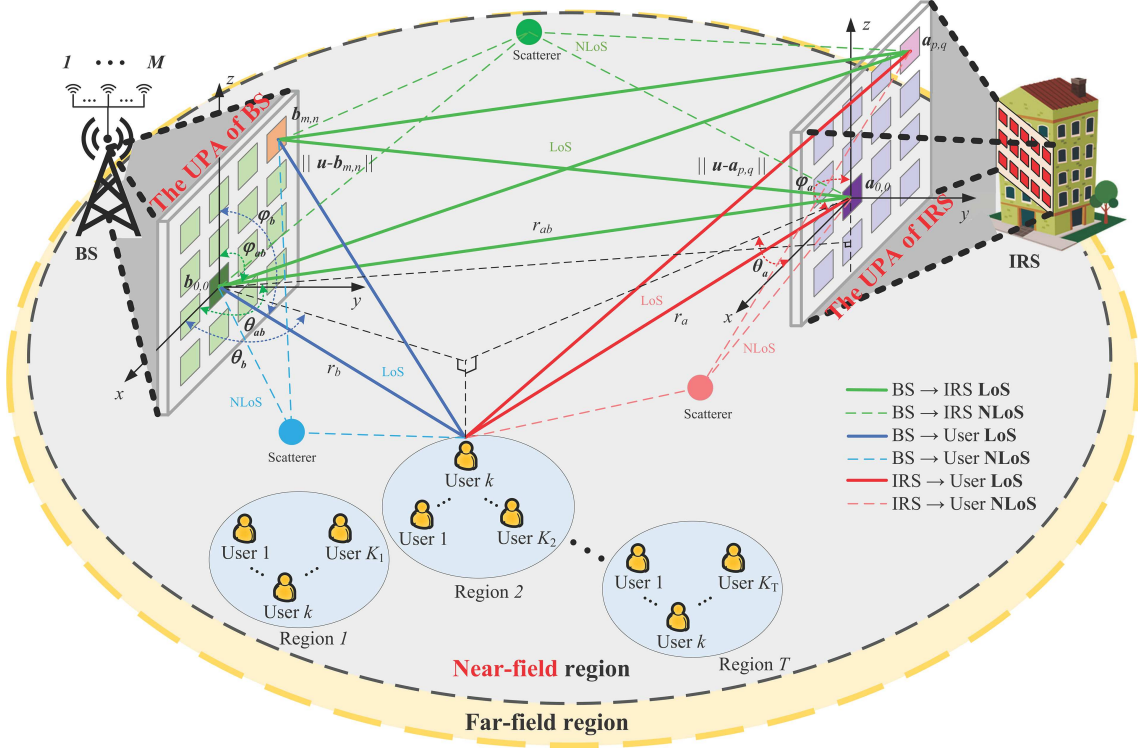
Figure 1   (Color online) System model of an IRS-aided MU NF communication.

where x represents the transmitted signal originating from the BS, and $\boldsymbol{v} \in \mathbb{C}^{M \times 1}$ denotes the transmit beamforming vector at BS. In addition, $\boldsymbol{\Theta} = \mathrm{diag}\{\tilde{\boldsymbol{\theta}}\} \in \mathbb{C}^{N \times N}$ stands for the reflection coefficient matrix of the IRS, and $\tilde{\boldsymbol{\theta}} \in \mathbb{C}^{1 \times N}$. Moreover, $w_{t,k}$ is the additive white Gaussian noise (AWGN) at the corresponding user with $w_{t,k} \sim \mathcal{CN}(0, \sigma_u^2)$.

Since $\boldsymbol{f}_{t,k}\boldsymbol{\Theta} = \tilde{\boldsymbol{\theta}}\mathrm{diag}\{\boldsymbol{f}_{t,k}\}$, it can be intuitively deduced that $\boldsymbol{f}_{t,k}\boldsymbol{\Theta}\boldsymbol{G} = \tilde{\boldsymbol{\theta}}\mathrm{diag}\{\boldsymbol{f}_{t,k}\}\boldsymbol{G}$. Consequently, the core task of CE is to accurately estimate the following:

$$\boldsymbol{H}_{t,k} = \begin{bmatrix} \boldsymbol{h}_{t,k} \\ \mathrm{diag}\{\boldsymbol{f}_{t,k}\}\boldsymbol{G} \end{bmatrix} \in \mathbb{C}^{(N+1) \times M}, \quad \forall\, t, k. \tag{2}$$

Correspondingly, Eq. (1) can be transformed into another form of expression as follows:

$$y_{t,k} = \boldsymbol{\theta}\boldsymbol{H}_{t,k}\boldsymbol{v}x + w_{t,k}, \tag{3}$$

where

$$\boldsymbol{\theta} = \begin{bmatrix} 1 & \tilde{\boldsymbol{\theta}} \end{bmatrix} \in \mathbb{C}^{1 \times (N+1)}. \tag{4}$$

2.1   Channel models

2.1.1   *Near-field channel model*

Drawing inspiration from [30], it is assumed that the M antennas of BS's UPA are arranged in an array of $M_{xb} \times M_{zb}$ on the x-z plane, namely $M = M_{xb} \times M_{zb}$, where $M_{xb} = 2\tilde{M}_{xb} + 1$ and $M_{zb} = 2\tilde{M}_{zb} + 1$, and the antenna spacing of this array is set to d_{xb} and d_{zb} in the two dimensions, respectively. Additionally, θ_b and ϕ_b signify the azimuth angle and elevation angle of the user relative to the x-z plane of BS, respectively. Also, the distance between the user and the central of BS's UPA can be represented as r_b. As a result, the coordinates of the k-th user situated in the t-th region can be given as

$$\boldsymbol{u}_b = (r_b \cos\theta_b \sin\phi_b, \ r_b \sin\theta_b \sin\phi_b, \ r_b \cos\phi_b). \tag{5}$$

Similarly, the coordinates of the (m, n)-th antenna of the BS's UPA are obtained as follows:

$$\boldsymbol{b}_{m,n} = (nd_{xb}, \ 0, \ md_{zb}), \tag{6}$$

where $\forall n \in \{-\tilde{M}_{xb}, \ldots, \tilde{M}_{xb}\}$ and $\forall m \in \{-\tilde{M}_{zb}, \ldots, \tilde{M}_{zb}\}$.

According to [30], assuming $d_{xb}/r_b \ll 1$ and $d_{zb}/r_b \ll 1$, we can derive the propagation distance $\|\boldsymbol{u}_b - \boldsymbol{b}_{m,n}\|$ as follows:

$$
\begin{aligned}
\|\boldsymbol{u}_b - \boldsymbol{b}_{m,n}\| &= \sqrt{(r_b \cos\theta_b \sin\phi_b - nd_{xb})^2 + (r_b \sin\theta_b \sin\phi_b)^2 + (r_b \cos\phi_b - md_{zb})^2} \\
&= r_b \sqrt{1 + \frac{n^2 d_{xb}^2}{r_b^2} + \frac{m^2 d_{zb}^2}{r_b^2} - \frac{2r_b nd_{xb} \cos\theta_b \sin\phi_b}{r_b^2} - \frac{2r_b md_{zb} \cos\phi_b}{r_b^2}}.
\end{aligned}
\tag{7}
$$

Subsequently, by applying the following second-order Taylor expansion to (7):

$$
\sqrt{1+x} \approx 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \mathcal{O}(x^2),
\tag{8}
$$

then, Eq. (7) can be approximated as

$$
\|\boldsymbol{u}_b - \boldsymbol{b}_{m,n}\| \approx r_b \underbrace{-nd_{xb} \cos\theta_b \sin\phi_b + \frac{n^2 d_{xb}^2 (1 - \cos^2\theta_b \sin^2\phi_b)}{2r_b}}_{h_1,\ \text{only depend on } n} \underbrace{-md_{zb} \cos\phi_b + \frac{m^2 d_{zb}^2 \sin^2\phi_b}{2r_b}}_{h_2,\ \text{only depend on } m},
\tag{9}
$$

where h_1 and h_2 solely contingent on n and m, respectively.

Referring to the uniform spherical wave (USW) model of [31, 32], the NF LoS MISO channel can be formulated as follows:

$$
\boldsymbol{h}_{t,k}^{\mathrm{LoS}} = \beta_b \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda} r_b} \left[\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(\|\boldsymbol{u}_b - \boldsymbol{b}_{-\tilde{M}_{xb}, -\tilde{M}_{zb}}\| - r_b)}, \ldots, \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(\|\boldsymbol{u}_b - \boldsymbol{b}_{\tilde{M}_{xb}, \tilde{M}_{zb}}\| - r_b)} \right]^{\mathrm{T}}.
\tag{10}
$$

Once the constant phase $\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda} r_b}$ is taken out, the resulting phase of the array response vector (ARV) is obtained, referred to as h_1 and h_2. More specifically, the ARV of the BS-user channel is as follows:

$$
\boldsymbol{a}(\theta_b, \phi_b, r_b) = \boldsymbol{a}_x(\theta_b, \phi_b, r_b) \otimes \boldsymbol{a}_z(\phi_b, r_b),
\tag{11}
$$

where

$$
[\boldsymbol{a}_x(\theta_b, \phi_b, r_b)]_n = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda} \cdot h_1} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\left(-nd_{xb} \cos\theta_b \sin\phi_b + \frac{n^2 d_{xb}^2 (1 - \cos^2\theta_b \sin^2\phi_b)}{2r_b}\right)},
\tag{12}
$$

and

$$
[\boldsymbol{a}_z(\phi_b, r_b)]_m = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda} \cdot h_2} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\left(-md_{zb} \cos\phi_b + \frac{m^2 d_{zb}^2 \sin^2\phi_b}{2r_b}\right)}.
\tag{13}
$$

As demonstrated in Figure 1, the multipath effect in the NF environment, originating from the refraction and reflection of signals triggered by scatterers, prompts the signals to reach the users via NLoS paths, and this stochasticity endows the channel with dynamic characteristics, which urgently requires a statistical model to capture its changing patterns. In this context, the channel from BS to scatterer can be analogized to the MISO channel model, and thus, a novel expression of the NF multipath channel can be constructed as follows:

$$
\boldsymbol{h}_{t,k} = \underbrace{\beta_b \boldsymbol{a}(\theta_b, \phi_b, r_b)}_{\text{LoS}} + \sum_{\ell=1}^{L_b} \underbrace{\tilde{\beta}_{b,\ell} \boldsymbol{a}(\tilde{\theta}_{b,\ell}, \tilde{\phi}_{b,\ell}, \tilde{r}_{b,\ell})}_{\text{NLoS}},
\tag{14}
$$

where L_b denotes the number of scatterers between the BS and the user, and $\ell = 1, 2, \ldots, L_b$. Moreover, $\tilde{\theta}_{b,\ell}$, $\tilde{\phi}_{b,\ell}$, $\tilde{r}_{b,\ell}$, and $\tilde{\beta}_{b,\ell}$ represent the corresponding azimuth, elevation, distance, and channel gain of the ℓ-th scatterer, respectively.

Likewise, the UPA of the IRS is also placed in the x-z plane, parallel to BS's UPA. For the k-th user positioned in the t-th region, their coordinates relative to the IRS's UPA are articulated by

$$
\boldsymbol{u}_a = (r_a \cos\theta_a \sin\phi_a, r_a \sin\theta_a \sin\phi_a, r_a \cos\phi_a),
\tag{15}
$$

where θ_a, ϕ_a are the user's azimuth angle and elevation angle with respect to the IRS, respectively, and r_a is the user's distance from the IRS center element. Correspondingly, the (p, q)-th element of the IRS's element array is given by

$$\boldsymbol{a}_{p,q} = (qd_{xa}, \ 0, \ pd_{za}), \ \forall q \in \{-\tilde{N}_{za}, \ldots, \tilde{N}_{xa}\}, \ p \in \{-\tilde{N}_{za}, \ldots, \tilde{N}_{za}\}, \tag{16}$$

where $N = N_{xa} \times N_{za}$, $N_{xa} = 2\tilde{N}_{xa} + 1$, and $N_{za} = 2\tilde{N}_{za} + 1$. Correspondingly, the element spacing of IRS's UPA along the x and z axis is represented by d_{xa} and d_{za}, respectively.

Similar to (7), by employing (8), the distance between the k-th user in the t-th region and the (p, q)-th element of the IRS is as follows:

$$\|\boldsymbol{u}_a - \boldsymbol{a}_{p,q}\| \approx r_a \underbrace{-qd_{xa}\cos\theta_a\sin\phi_a + \frac{q^2 d_{xa}^2(1 - \cos^2\theta_a\sin^2\phi_a)}{2r_a}}_{f_1, \text{ only depend on } q} \underbrace{-pd_{za}\cos\phi_a + \frac{p^2 d_{za}^2\sin^2\phi_a}{2r_a}}_{f_2, \text{ only depend on } p}. \tag{17}$$

Thus, referring to [30], the ARV of the IRS-user channel can be displayed as follows:

$$\boldsymbol{a}(\theta_a, \phi_a, r_a) = \boldsymbol{a}_x(\theta_a, \phi_a, r_a) \otimes \boldsymbol{a}_z(\phi_a, r_a), \tag{18}$$

where

$$[\boldsymbol{a}_x(\theta_a, \phi_a, r_a)]_q = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\cdot f_1} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(-qd_{xa}\cos\theta_a\sin\phi_a + \frac{q^2 d_{xa}^2(1-\cos^2\theta_a\sin^2\phi_a)}{2r_a})}, \tag{19}$$

and

$$[\boldsymbol{a}_z(\phi_a, r_a)]_p = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\cdot f_2} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(-pd_{za}\cos\phi_a + \frac{p^2 d_{za}^2\sin^2\phi_a}{2r_a})}. \tag{20}$$

As depicted in Figure 1, the presence of scatterers within the environment can elicit multipath effects in the IRS-user NF channel. Hence, the channel between the k-th user in the t-th region and the IRS can be formulated as

$$\boldsymbol{f}_{t,k} = \underbrace{\beta_a \boldsymbol{a}(\theta_a, \phi_a, r_a)}_{\text{LoS}} + \sum_{\ell=1}^{L_a} \underbrace{\tilde{\beta}_{a,\ell} \boldsymbol{a}(\tilde{\theta}_{a,\ell}, \tilde{\phi}_{a,\ell}, \tilde{r}_{a,\ell})}_{\text{NLoS}}, \tag{21}$$

where L_a is the count of scatterers between IRS and user, and the azimuth angle, elevation angle, and distance of the ℓ-th scatterer are represented as $\tilde{\theta}_{a,\ell}$, $\tilde{\phi}_{a,\ell}$, and $\tilde{r}_{a,\ell}$.

As can be illustrated in Figure 1, the NLoS MIMO path can be conceptualized as the fusion of two distinct MISO channels with respect to the BS and IRS, respectively. As a result, the channel between BS and IRS can be written as follows:

$$\boldsymbol{G} = \boldsymbol{G}_{\text{LoS}} + \sum_{\ell=1}^{L_{ba}} \underbrace{\tilde{\beta}_\ell \boldsymbol{a}(\tilde{\theta}_{ba,\ell}, \tilde{\phi}_{ba,\ell}, \tilde{r}_{ba,\ell}) \boldsymbol{a}^{\mathrm{T}}(\tilde{\theta}_{ab,\ell}, \tilde{\phi}_{ab,\ell}, \tilde{r}_{ab,\ell})}_{\text{NLoS}}, \tag{22}$$

where the quantity of scatterers is designated as L_{ba}. In addition, $\tilde{\theta}_{ba,\ell}$ ($\tilde{\theta}_{ab,\ell}$) is the azimuth angle of the ℓ-th scatterer with respect to the x-z plane of IRS (BS), and the elevation angle of the ℓ-th scatterer with respect to the x-z plane of IRS (BS) is signified as $\tilde{\phi}_{ba,\ell}$ ($\tilde{\phi}_{ab,\ell}$). In particular, $\tilde{r}_{ba,\ell}$ ($\tilde{r}_{ab,\ell}$) denotes the distance between the ℓ-th scatterer and the central element of IRS (BS). More precisely, according to [30], $\boldsymbol{G}_{\text{LoS}}$ can be embodied as follows:

$$\boldsymbol{G}_{\text{LoS}} = \tilde{\beta}\boldsymbol{a}(\theta_{ba}, \phi_{ba}, r_{ba})\Big(\boldsymbol{a}(\theta_{ab}, \phi_{ab}, r_{ab})\Big)^{\mathrm{T}} \odot \big(\boldsymbol{G}_c^x \otimes \boldsymbol{G}_c^z\big), \tag{23}$$

where

$$[\boldsymbol{G}_c^x]_{q,n} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda r_{ab}} n d_{xb} q d_{xa}\left(1 - \cos^2\theta_{ab}\sin^2\phi_{ab}\right)}, \tag{24}$$

and

$$[\boldsymbol{G}_c^z]_{p,m} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda r_{ab}} m d_{zb} p d_{za}\sin^2\phi_{ab}}. \tag{25}$$

In (23)–(25), θ_{ba} (θ_{ab}) and ϕ_{ba} (ϕ_{ab}) represent the azimuth and elevation angles of the center of BS (IRS) relative to the center of IRS (BS), respectively. Besides, $r_{ba} = r_{ab} = \|\boldsymbol{a}_{0,0} - \boldsymbol{b}_{0,0}\|$ denote the distance between the central of IRS and the central of BS. As evident from (23), the NF LoS MIMO path matrix between parallel UPAs incorporates an interleaved variable, namely $\boldsymbol{G}_c^x \otimes \boldsymbol{G}_c^z$, which cannot be simply decomposed into the direct product of the ARVs of BS-side and IRS-side.

2.1.2 *Far-field channel model*

As a comparison, the FF ARV can be specifically described as follows:

$$\boldsymbol{a}_{\mathrm{far}}(\theta_b, \phi_b) = \left[\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\tilde{M}_{xb}d_{xb}\cos\theta_b\sin\phi_b}, \ldots, \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\tilde{M}_{xb}d_{xb}\cos\theta_b\sin\phi_b} \right]^{\mathrm{T}}$$
$$\otimes \left[\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\tilde{M}_{zb}d_{zb}\cos\phi_b}, \ldots, \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\tilde{M}_{zb}d_{zb}\cos\phi_b} \right]^{\mathrm{T}}. \tag{26}$$

Correspondingly, in this case, according to [30], the FF channel model between BS-user, IRS-user, and BS-IRS can be modeled as follows:

$$\boldsymbol{h}_{t,k}^{\mathrm{far}} = \underbrace{\beta_b \boldsymbol{a}_{\mathrm{far}}(\theta_b, \phi_b)}_{\mathrm{LoS}} + \underbrace{\sum_{\ell=1}^{L_b} \tilde{\beta}_{b,\ell} \boldsymbol{a}_{\mathrm{far}}(\tilde{\theta}_{b,\ell}, \tilde{\phi}_{b,\ell})}_{\mathrm{NLoS}}, \tag{27}$$

$$\boldsymbol{f}_{t,k}^{\mathrm{far}} = \underbrace{\beta_a \boldsymbol{a}_{\mathrm{far}}(\theta_a, \phi_a)}_{\mathrm{LoS}} + \underbrace{\sum_{\ell=1}^{L_a} \tilde{\beta}_{a,\ell} \boldsymbol{a}_{\mathrm{far}}(\tilde{\theta}_{a,\ell}, \tilde{\phi}_{a,\ell})}_{\mathrm{NLoS}}, \tag{28}$$

and

$$\boldsymbol{G}^{\mathrm{far}} = \boldsymbol{G}_{\mathrm{LoS}}^{\mathrm{far}} + \underbrace{\sum_{\ell=1}^{L_{ba}} \tilde{\beta}_\ell \boldsymbol{a}_{\mathrm{far}}(\tilde{\theta}_{ba,\ell}, \tilde{\phi}_{ba,\ell}) \boldsymbol{a}_{\mathrm{far}}^{\mathrm{T}}(\tilde{\theta}_{ab,\ell}, \tilde{\phi}_{ab,\ell})}_{\mathrm{NLoS}}, \tag{29}$$

in which, in the FF scenario, the LoS component between the BS and IRS channels can be expressed as

$$\boldsymbol{G}_{\mathrm{LoS}}^{\mathrm{far}} = \tilde{\beta} \boldsymbol{a}_{\mathrm{far}}(\theta_{ba}, \phi_{ba}) \Big(\boldsymbol{a}_{\mathrm{far}}(\theta_{ab}, \phi_{ab}) \Big)^{\mathrm{T}}. \tag{30}$$

2.2 Problem formulation

To ascertain the downlink direct and cascaded channel, i.e., $\boldsymbol{H}_{t,k}$, BS transmits predetermined pilot signals to users, facilitated by the IRS, spanning across Q time slots, where $q = 1, 2, \ldots, Q$. Drawing upon (3), $y_{t,k,q}^p$ for the k-th user in the t-th region, in the q-th time slot, is mathematically formulated as follows:

$$y_{t,k,q}^p = \boldsymbol{\theta}_q \boldsymbol{H}_{t,k} \boldsymbol{v} p_q + w_{t,k,q}, \tag{31}$$

where the pilot signal transmitted by BS is signified as p_q, and $\boldsymbol{\theta}_q$ stands for the IRS phase shift within the current time slot. Also, $w_{t,k,q}$ denotes the noise at the user of the current time slot.

Following the transmission of pilot signals across Q consecutive time slots, we derive the comprehensive $Q \times 1$ received pilot vector as depicted below:

$$\boldsymbol{y}_{t,k}^p = [y_{t,k,1}^p, \ y_{t,k,2}^p, \ldots, \ y_{t,k,Q}^p]^{\mathrm{T}} = \boldsymbol{\Phi} \boldsymbol{H}_{t,k} \boldsymbol{v} p_{t,k,q} + \boldsymbol{w}_{t,k}, \tag{32}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\theta}_1^{\mathrm{T}}, \ \boldsymbol{\theta}_2^{\mathrm{T}}, \ldots, \ \boldsymbol{\theta}_Q^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{Q \times (N+1)}$ and $\boldsymbol{w}_{t,k} = [w_{t,k,1}, \ w_{t,k,2}, \ldots, \ w_{t,k,Q}]^{\mathrm{T}} \in \mathbb{C}^{Q \times 1}$. According to [29], assuming $p_{t,k,q} = 1$ for the purpose of subsequent CE, Eq. (32) can be rephrased as

$$\boldsymbol{y}_{t,k}^p = \boldsymbol{\Phi} \boldsymbol{H}_{t,k} \boldsymbol{v} + \boldsymbol{w}_{t,k}. \tag{33}$$

Due to the fact that $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$, Eq. (33) can be reformulated as

$$\boldsymbol{y}_{t,k}^p = \underbrace{(\boldsymbol{v}^{\mathrm{T}} \otimes \boldsymbol{\Phi})}_{\boldsymbol{A}_{t,k}} \underbrace{\mathrm{vec}(\boldsymbol{H}_{t,k})}_{\boldsymbol{h}_{t,k}'} + \boldsymbol{w}_{t,k} = \boldsymbol{A}_{t,k} \boldsymbol{h}_{t,k}' + \boldsymbol{w}_{t,k}. \tag{34}$$

It follows from (34) that the essence of CE lies in reconstructing $\boldsymbol{h}_{t,k}'$ utilizing the known $\boldsymbol{y}_{t,k}^p$ and $\boldsymbol{A}_{t,k}$. According to [33], the estimated $\boldsymbol{h}_{t,k}'$ using LS estimator is provided by

$$\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}' = \boldsymbol{A}_{t,k}^{\dagger} \boldsymbol{y}_{t,k}^p, \tag{35}$$

where $\boldsymbol{A}_{t,k}^{\dagger} = \boldsymbol{A}_{t,k}^{\mathrm{H}}(\boldsymbol{A}_{t,k}\boldsymbol{A}_{t,k}^{\mathrm{H}})^{-1}$ denotes the pseudoinverse of $\boldsymbol{A}_{t,k}$.

The estimation MSE of (35) is

$$\varepsilon_{\mathrm{LS}} = \mathbb{E}\{\|\boldsymbol{h}_{t,k}' - \widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'\|_F^2\} = \mathbb{E}\{\|\boldsymbol{h}_{t,k}' - \boldsymbol{A}_{t,k}^{\dagger}(\boldsymbol{A}_{t,k}\boldsymbol{h}_{t,k}' + \boldsymbol{w}_{t,k})\|_F^2\} = \sigma_u^2 \mathrm{tr}\big((\boldsymbol{A}_{t,k}\boldsymbol{A}_{t,k}^{\mathrm{H}})^{-\mathrm{H}}\big). \tag{36}$$

It is worth noting that since $\boldsymbol{A}_{t,k} = \boldsymbol{v}^{\mathrm{T}} \otimes \boldsymbol{\Phi}$, $\boldsymbol{v}$ and $\boldsymbol{\Phi}$ need to be pre-determined to be constant values as a prerequisite for CE. Consequently, according to [34], the transmit beamforming vector $\boldsymbol{v}$ randomly selects their components from $\{-\frac{1}{\sqrt{M}}, +\frac{1}{\sqrt{M}}\}$. Furthermore, based on [35], $\boldsymbol{\Phi}$ can be innovatively crafted as a discrete Fourier transform (DFT), namely,

$$\boldsymbol{\Phi}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_Q & \cdots & W_Q^{Q-1} \\ \cdots & \cdots & \ddots & \cdots \\ 1 & W_Q^N & \cdots & W_Q^{N(Q-1)} \end{bmatrix} \in \mathbb{C}^{(N+1) \times Q}, \tag{37}$$

where $W_Q = \mathrm{e}^{\mathrm{j}2\pi/Q}$. It is worth noting that here, $p_{t,k,q}$, $\boldsymbol{v}$, $\boldsymbol{\Phi}$ are pre-designed to be fixed values for channel estimation according to [29, 34, 35], respectively. However, related researchers have demonstrated that proper pilot optimization as well as IRS phase shift configuration will bring additional performance gains. The optimization of the pilot and IRS phase shift in channel estimation for IRS-assisted communication systems will be our future research work.

Subsequently, the MMSE estimator for the channel $\boldsymbol{h}_{t,k}'$ is

$$\widehat{\boldsymbol{h}}_{t,k,\mathrm{MMSE}}' = \boldsymbol{W}_{\mathrm{MMSE}}\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'. \tag{38}$$

Correspondingly, the MMSE estimator obtains an estimate based on $\boldsymbol{W}$ to minimize the MSE in the following given equation:

$$\varepsilon_{\mathrm{MMSE}} = \mathbb{E}\{\|\boldsymbol{h}_{t,k}' - \boldsymbol{W}_{\mathrm{MMSE}}\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'\|_F^2\}. \tag{39}$$

By solving the problem presented in (39), the following outcomes can be derived:

$$\boldsymbol{W}_{\mathrm{MMSE}} = \boldsymbol{R}_{\boldsymbol{h}_{t,k}'\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'} \left(\boldsymbol{R}_{\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'} + \frac{\sigma_u^2}{\sigma_x^2}\boldsymbol{I} \right)^{-1}, \tag{40}$$

where $\boldsymbol{R}_{\boldsymbol{h}_{t,k}'\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'} = \mathbb{E}\{\boldsymbol{h}_{t,k}'\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'^{\mathrm{H}}\}$ represents the cross-correlation matrix between the actual channel and the LS CE, while $\boldsymbol{R}_{\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'} = \mathbb{E}\{\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'\widehat{\boldsymbol{h}}_{t,k,\mathrm{LS}}'^{\mathrm{H}}\}$ denotes the autocorrelation matrix of the LS CE.

3 CRLB

The CRLB is often used to calculate the optimal estimation accuracy that can be achieved theoretically, which can evaluate the effectiveness of proposed CE algorithms. For the CE problem (34), $\boldsymbol{A}_{t,k}$ is a real matrix, $\boldsymbol{y}_{t,k}^p \in \mathbb{C}^{Q \times 1}$ represents the known received signal, $\boldsymbol{h}_{t,k}'$ denotes the channel to be estimated, and $\boldsymbol{w}_{t,k}$ is the whitened noise. Consequently, Eq. (34) can be divided into two parts, where the real part is

$$\boldsymbol{y}_{t,k,u}^p = \boldsymbol{A}_{t,k}\boldsymbol{h}_{t,k,u}' + \boldsymbol{w}_{t,k,u}, \tag{41}$$

and the imaginary part is as follows:

$$\boldsymbol{y}_{t,k,v}^p = \boldsymbol{A}_{t,k}\boldsymbol{h}_{t,k,v}' + \boldsymbol{w}_{t,k,v}, \tag{42}$$

where

$$\boldsymbol{y}_{t,k,u}^p = \mathfrak{Re}(\boldsymbol{y}_{t,k}^p), \ \boldsymbol{y}_{t,k,v}^p = \mathfrak{Im}(\boldsymbol{y}_{t,k}^p),$$
$$\boldsymbol{h}_{t,k,u}' = \mathfrak{Re}(\boldsymbol{h}_{t,k}'), \ \boldsymbol{h}_{t,k,v}' = \mathfrak{Im}(\boldsymbol{h}_{t,k}'),$$

$$w_{t,k,u} = \Re(w_{t,k}), \ w_{t,k,v} = \Im(w_{t,k}). \tag{43}$$

Thus, the CRLB of $\hat{h}'_{t,k}$ can also be divided into two parts, which are shown as

$$\gamma_{t,k} = \gamma_{t,k}^u + \gamma_{t,k}^v = \mathbb{E}\left\{\left\|\widehat{h}'_{t,k,u} - h'_{t,k,u}\right\|^2\right\} + \mathbb{E}\left\{\left\|\widehat{h}'_{t,k,v} - h'_{t,k,v}\right\|^2\right\}. \tag{44}$$

In this case, we first consider the real part (41). Since the $w_{t,k,u}$ follows the distribution of Gaussian distribution with 0 mean and σ_u^2 variance, the conditional probability density function of $y_{t,k,u}^p$ with the given $h'_{t,k,u}$ is

$$p_{y_{t,k,u}^p | h'_{t,k,u}}\left(y_{t,k,u}^p; h'_{t,k,u}\right) = \frac{1}{(2\pi\sigma_u^2)^{Q/2}} \exp\left\{-\frac{1}{2\sigma_u^2}\left\|y_{t,k,u}^p - A_{t,k}h'_{t,k,u}\right\|^2\right\}. \tag{45}$$

The Fisher information matrix of (41) can then be derived as

$$[J]_{i,j} = -\mathbb{E}\left\{\frac{p_{y_{t,k,u}^p|h'_{t,k,u}}(y_{t,k,u}^p; h'_{t,k,u})}{\partial h'_{t,k,u,i}\partial h'_{t,k,u,j}}\right\} = \frac{1}{\sigma_u^2}\left[A_{t,k}^{\mathrm{H}}A_{t,k}\right]_{i,j}, \tag{46}$$

where $h'_{t,k,u,i}$, $h'_{t,k,u,j}$ denote the i-th and j-th entry of $h'_{t,k,u}$. Then, the real part $\gamma_{t,k}^u$ is

$$\gamma_{t,k}^u = \mathbb{E}\left\{\left\|\widehat{h}'_{t,k,u} - h'_{t,k,u}\right\|^2\right\} \geqslant \mathrm{tr}\left\{J_u^{-1}\right\} = \sigma_u^2\mathrm{tr}\left\{(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1}\right\}. \tag{47}$$

Since $A_{t,k} = v^{\mathrm{T}} \otimes \Phi$, $(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1}$ can be presented as

$$(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1} = \left((v^{\mathrm{T}} \otimes \Phi)^{\mathrm{H}}(v^{\mathrm{T}} \otimes \Phi)\right)^{-1} = \left((vv^{\mathrm{H}})^{-1}\right)^{\mathrm{T}} \otimes (\Phi^{\mathrm{H}}\Phi)^{-1}. \tag{48}$$

Since $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$ and $\mathrm{tr}(A^{\mathrm{T}}) = \mathrm{tr}(A)$, thus, $\mathrm{tr}\{(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1}\}$ can be calculated as

$$\mathrm{tr}\left\{(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1}\right\} = \mathrm{tr}\left(\left((vv^{\mathrm{H}})^{-1}\right)^{\mathrm{T}}\right)\mathrm{tr}\left((\Phi^{\mathrm{H}}\Phi)^{-1}\right) = \mathrm{tr}\left((vv^{\mathrm{H}})^{-1}\right)\mathrm{tr}\left((\Phi^{\mathrm{H}}\Phi)^{-1}\right). \tag{49}$$

Assuming $\{\alpha_i\}_{i=1}^M$ is defined as the M eigenvalues of the matrix of vv^{H}, and $\{\beta_j\}_{j=1}^{N+1}$ is defined as the $N+1$ eigenvalues of the matrix of $\Phi^{\mathrm{H}}\Phi$, then

$$\mathrm{tr}\left\{(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1}\right\} = \left(\sum_{i=1}^M \alpha_i^{-1}\right)\left(\sum_{j=1}^{N+1} \beta_j^{-1}\right). \tag{50}$$

According to [36], we have

$$\mathrm{tr}\left\{(A_{t,k}^{\mathrm{H}}A_{t,k})^{-1}\right\} \geqslant M\left(M\bigg/\sum_{i=1}^M \alpha_i\right)(N+1)\left((N+1)\bigg/\sum_{j=1}^{N+1} \beta_j\right) = \frac{M^2}{\mathrm{tr}(vv^{\mathrm{H}})}\frac{(N+1)^2}{\mathrm{tr}(\Phi^{\mathrm{H}}\Phi)}. \tag{51}$$

Finally, the CRLB of the real part of $\hat{h}'_{t,k}$ becomes

$$\gamma_{t,k}^u = \mathbb{E}\left\{\left\|\widehat{h}'_{t,k,u} - h'_{t,k,u}\right\|^2\right\} = \sigma_u^2\frac{M(N+1)}{Q}. \tag{52}$$

Observing (41) and (42), it can be seen that the real part and imaginary part have the same form, namely

$$\gamma_{t,k}^v = \gamma_{t,k}^u = \sigma_u^2\frac{M(N+1)}{Q}. \tag{53}$$

Thus, the CRLB of (34) is as follows:

$$\gamma_{t,k} = \gamma_{t,k}^u + \gamma_{t,k}^v = 2\sigma_u^2\frac{M(N+1)}{Q}. \tag{54}$$
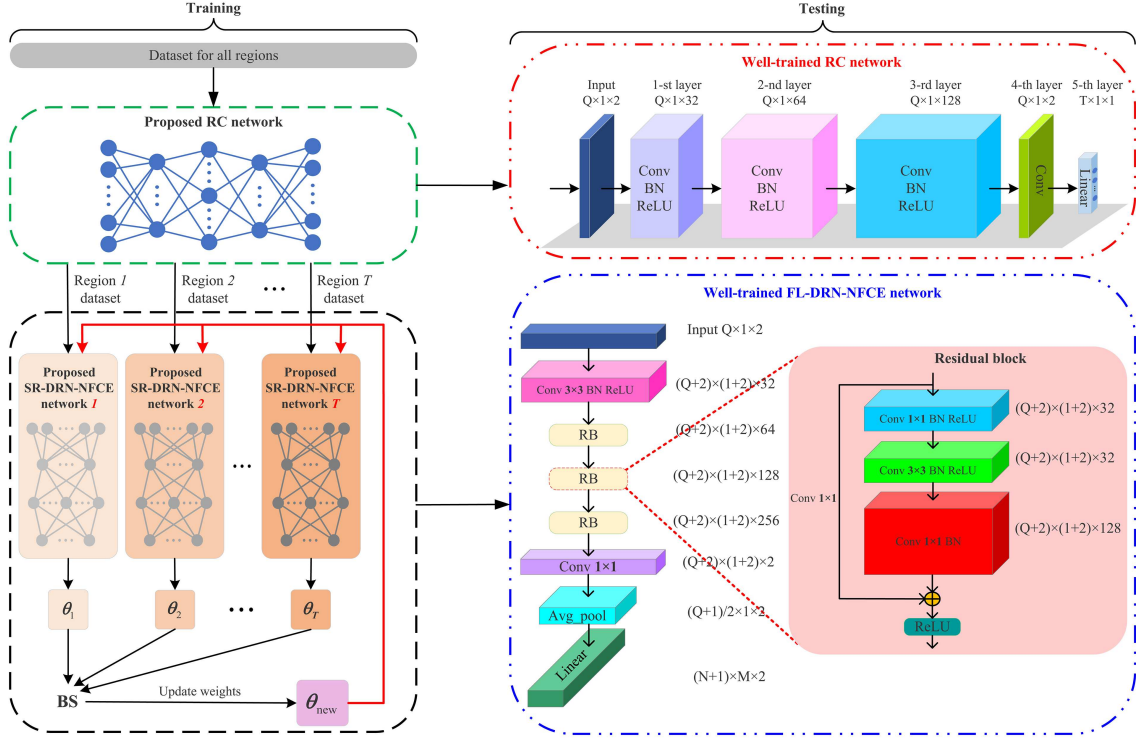
Figure 2   (Color online) Proposed G-DRN-NFCE framework.

4   Proposed G-DRN-NFCE algorithm

In this section, as shown in Figure 2, the proposed global DRN-NFCE (G-DRN-NFCE) framework is introduced. Firstly, the basic principles of the proposed SR-DRN-NFCE network are elaborated in detail in Subsection 4.1. Subsequently, in Subsection 4.2, a CNN-based RC network is designed to precisely match datasets from different user regions with their corresponding SR-DRN-NFCE networks. Then, the training details of the proposed FL-DRN-NFCE network are presented in Subsection 4.3. Finally, in Subsection 4.4, the computational complexity of the above-mentioned networks is given.

4.1   Basic principle of the proposed SR-DRN-NFCE network

Specifically, the proposed SR-DRN-NFCE network introduces a non-linear function $\boldsymbol{f_\omega}$ between $\boldsymbol{y}_{t,k}^p$ and $\widehat{\boldsymbol{h}}'_{t,k,\mathrm{DRN}}$, which can be mathematically expressed as

$$\widehat{\boldsymbol{h}}'_{t,k,\mathrm{DRN}} = \boldsymbol{f_\omega}(\boldsymbol{y}_{t,k}^p), \tag{55}$$

where $\boldsymbol{\omega}$ represents weight. Accordingly, the expression for the loss function is as follows:

$$\mathcal{L}(\boldsymbol{\omega}) = \frac{1}{S_{t,k}} \sum_{s=1}^{S_{t,k}} \parallel \widehat{\boldsymbol{h}}'_{t,k,\mathrm{DRN}} - \widehat{\boldsymbol{h}}'_{t,k,\mathrm{LS}} \parallel_2^2, \tag{56}$$

where $S_{t,k}$ stands for the size of training dataset. The traditional LS CE scheme can be used to obtain the label $\widehat{\boldsymbol{h}}'_{t,k,\mathrm{LS}}$. The objective of exploiting the proposed SR-DRN-NFCE network for CE is to minimize $\mathcal{L}(\boldsymbol{\omega})$ by optimizing $\boldsymbol{\omega}$, which is

$$\min_{\boldsymbol{\omega}} \mathcal{L}(\boldsymbol{\omega}) = \frac{1}{S_{t,k}} \sum_{s=1}^{S_{t,k}} \parallel \boldsymbol{f_\theta}(\boldsymbol{y}_{t,k}^p) - \widehat{\boldsymbol{h}}'_{t,k,\mathrm{LS}} \parallel_2^2 . \tag{57}$$

In each iteration i, $\boldsymbol{\omega}$ is updated through the following way, namely,

$$\boldsymbol{\omega}_{i+1} = \boldsymbol{\omega}_i - \lambda_i \boldsymbol{g}(\boldsymbol{\omega}_i), \tag{58}$$

Table 1 Hyperparameters of the proposed SR-DRN-NFCE network. Input: The received pilot signal $\boldsymbol{y}_{t,k}^p$ with the size of $Q \times 1$. Output: The estimated channel with the size of $(N+1)M \times 1$.

Layer	Operation	Number of parameters
1	Conv+BN+ReLU	$(3 \times 3 \times 2) \times 32$
2	RB	$(1 \times 1 \times 32) \times 16 + (3 \times 3 \times 16) \times 16 + (1 \times 1 \times 16) \times 64 + (1 \times 1 \times 32) \times 64$
3	RB	$(1 \times 1 \times 64) \times 32 + (3 \times 3 \times 32) \times 32 + (1 \times 1 \times 32) \times 128 + (1 \times 1 \times 64) \times 128$
4	RB	$(1 \times 1 \times 128) \times 64 + (3 \times 3 \times 64) \times 64 + (1 \times 1 \times 64) \times 256 + (1 \times 1 \times 128) \times 256$
5	Conv	$(1 \times 1 \times 256) \times 2$
6	Avg-pool	$-$
7	Linear	$(\lfloor \frac{Q+1}{2} \rfloor \times 1 \times 2) \times ((N+1) \times M \times 2) + (N+1) \times M \times 2$

Table 2 Hyperparameters of the proposed RC network.

Layer	Operation	Number of parameters
1	Conv+BN+ReLU	$(3 \times 3 \times 2) \times 32$
2	Conv+BN+ReLU	$(3 \times 3 \times 32) \times 64$
3	Conv+BN+ReLU	$(3 \times 3 \times 64) \times 128$
4	Conv	$(1 \times 1 \times 128) \times 2$
5	Linear	$(Q \times 1 \times 2) \times (3 \times 1 \times 1) + (3 \times 1 \times 1)$

where $\boldsymbol{g}(\boldsymbol{\omega}_i)$ is the gradient vector (GV) for $\boldsymbol{\omega}_i$ and λ_i is the learning rate (LR).

Specifically, as displayed in Table 1, the proposed SR-DRN-NFCE comprises two convolution (Conv) layers, one average pool (Avg-pool) layer, one linear layer, and three residual blocks (RBs). The first layer is configured as a Conv layer with padding set to 2, a kernel size of 3, and a stride of 1. This design expands the input signal's dimensions in height and width from $Q \times 1$ to $(Q+2) \times (1+2)$, aiming to meet the requirements of feature extraction in subsequent Conv layers and dimensionality reduction in Avg-pool layers. Layers 2–4 consist of three consecutive RBs [37, 38], characterized by their unique approach: initially reducing the dimensionality of the feature map using a 1×1 Conv kernel, followed by a standard Conv layer for feature extraction, and finally employing another 1×1 Conv kernel to expand the count of channels in the feature map. This design strategy offers significant advantages in reducing computational load and enhancing model efficiency. Additionally, to maintain dimensional consistency in the residual connections, an extra 1×1 Conv kernel is utilized for dimensionality adjustment. The specific parameter configuration of the second RB is detailed in Figure 2: the input feature map has dimensions of $(Q+2) \times (1+2) \times 64$, and the output feature map has dimensions of $(Q+2) \times (1+2) \times 128$. Within the 2nd RB, the number of channels is first reduced to 32, which halves the number of parameters in the intermediate Conv layer for feature extraction, thereby lowering computational complexity. The final 1×1 Conv layer merely changes the number of channels, with relatively low computational complexity. The fifth layer is equipped with a 1×1 Conv layer that further reduces the count of channels in the input feature map to 2. The sixth layer is an Avg-pool layer with a kernel size of 3 and a stride of 2, which decreases the height and width of the input feature map from $(Q+2) \times (1+2)$ to $\lfloor (Q+1)/2 \rfloor \times 1$. The primary role of the fifth and sixth layers is to decrease the input dimensionality for the seventh linear layer, thereby effectively reducing the parameters and computational complexity of the linear layer. Ultimately, the seventh linear layer is responsible for outputting the predicted signal, with an output dimension of $(N+1) \times M \times 2$.

4.2 Proposed RC network

As diagrammed in Figure 1, considering that the NF model is more sensitive to changes in distance, angle, and other factors than the FF model, which results in significant differences in the channel characteristics of users in different regions. In order to effectively capture and fully utilize these diverse channel characteristics, the entire region is divided into T different sub regions. Therefore, in order to accurately match datasets from different user regions with corresponding SR-DRN-NFCE networks, a CNN-based RC is first designed.

As shown in Table 2, the proposed RC network architecture comprises four Conv layers and a linear layer. To enhance the steadiness of the model and expedite the training efficiency, a batch normalization (BN) is introduced between Conv and the rectified linear unit (ReLU). The first three layers are all "Conv + BN + ReLU" operations, where the Conv layer is responsible for extracting features from the input

---

Algorithm 1 Proposed G-DRN-NFCE algorithm.

Initialization: initialize trainable parameters $i, j = 0$, raw training data $\boldsymbol{y}_{t,k}^p$ and $\boldsymbol{h}_{t,k}'$.

Training:
 1: Input training dataset for all regions;
 2: **while** $j < J$ **do**
 3:     Update $\boldsymbol{\alpha}$ by backpropagation (BP) algorithm to minimize the loss function of the RC;
 4:     $j = j + 1$;
 5: **end while**
 6: Output well-trained RC network;

Testing:
 7: Input testing set;
 8: **do** classification using proposed RC network in Figure 2;
 9: Output classified dataset;

Training:
10: Input training dataset for the current region;
11: **while** $i < I_t$ **do**
12:     Update $\boldsymbol{\omega}$ by BP algorithm to minimize $\mathcal{L}(\boldsymbol{\omega})$;
13:     $i = i + 1$;
14: **end while**
15: Output well-trained SR-DRN-NFCE network;
16: Joint training of SR-DRN-NFCE networks corresponding to different regions in an FL manner;

Testing:
17: Input testing set;
18: **do** channel estimation using proposed FL-DRN-NFCE network in Figure 2;
19: Output estimated channel.

---

signal. Importantly, the count of channels in these layers progressively increases from the first to the third. Specifically, the input of the proposed RC network is the received pilot signal $\boldsymbol{y}_{t,k}^p$, which includes both real and imaginary parts; therefore, the initial number of channels is 2. Upon passing through the first Conv layer, the channel number increases to 32, enabling the network to capture a broader range of features. After the second Conv layer, the channel number escalates further to 64, enhancing the network's capacity to represent complex signal patterns. Finally, after the third Conv layer, the channel number reaches 128, significantly boosting the network's discriminative power and enabling it to better differentiate between similar signals, particularly in regions where signals may overlap or exhibit similar characteristics—this is crucial for capturing intricate patterns and subtle differences within the signal. BN is employed to enhance training stability and accelerate convergence, while the ReLU activation function introduces nonlinearity, thereby bolstering the network's expressive capacity. The fourth layer specifically utilizes a 1×1 Conv kernel for dimensionality reduction of the feature maps received from the preceding layer, reducing the channel number back to 2. This step effectively reduces the parameters and computational complexity of the subsequent linear layer. The entire network is meticulously designed to efficiently extract signal features while mitigating computational load through dimensionality reduction. Ultimately, the linear layer maps the signals to category labels, achieving precise and efficient region classification of signals.

4.3 Proposed FL-DRN-NFCE network

Subsequently, as shown in Figure 2, the user datasets of all regions are input into the well-trained RC network; then, users in each region calculate local GVs based on the current region's dataset. As described in [39, 40], each user utilizes its local dataset to independently train and update local gradients, then uploads the trained local gradients to the BS to update the global model. Once the BS has collected the local GVs of all regions, collaborative training is performed between the BS and the user in an FL manner, i.e., the BS updates the weights via the following manner:

$$\boldsymbol{\omega}_{i+1} = \boldsymbol{\omega}_i - \lambda_i \frac{1}{\sum_{t=1}^{\mathrm{T}} K_t} \sum_{t=1}^{\mathrm{T}} \sum_{k=1}^{K_t} \boldsymbol{g}_{t,k}(\boldsymbol{\omega}_i). \tag{59}$$

Finally, the well-trained FL-DRN-NFCE network is obtained in Figure 2. The well-trained FL-DRN-NFCE network is developed through an FL framework based on SR-DRN-NFCE networks corresponding to different regions. It maintains the same architecture configuration as the SR-DRN-NFCE network, but its weight parameters are specifically distinct.

To sum up, the step-by-step summary of the proposed G-DRN-NFCE algorithm is shown in Algorithm 1.

4.4 Computational complexity analysis

4.4.1 *For the proposed RC network*

The computational complexity of the 1-st Conv [41] layer is

$$C_{r1} = \mathcal{O}\Big(\big(3^2 \cdot 2 \cdot 32\big) \times \big(Q \cdot 1\big)\Big) = \mathcal{O}(576Q). \tag{60}$$

Similarly, the computational complexity of the 2-nd and 3-rd Conv layers is

$$C_{r2} = \mathcal{O}\Big(\big(3^2 \cdot 32 \cdot 64\big) \times \big(Q \cdot 1\big)\Big) = \mathcal{O}(18432Q) \tag{61}$$

and

$$C_{r3} = \mathcal{O}\Big(\big(3^2 \cdot 64 \cdot 128\big) \times \big(Q \cdot 1\big)\Big) = \mathcal{O}(73728Q), \tag{62}$$

respectively. From Figure 2, the computational complexity of the 4-th Conv layer is

$$C_{r4} = \mathcal{O}\Big(\big(1 \cdot 1 \cdot 128 \cdot 2\big) \times \big(Q \cdot 1\big)\Big) = \mathcal{O}(256Q). \tag{63}$$

Under the premise of only considering multiplication operations, the computational complexity of the final linear layer [42] of the proposed RC network is as follows:

$$C_{r5} = \mathcal{O}\Big(\big(Q \cdot 1 \cdot 2\big) \times \big(T \cdot 1 \cdot 1\big)\Big) = \mathcal{O}(2QT). \tag{64}$$

Therefore, the total computational complexity of the proposed RC network is

$$C_r = C_{r1} + C_{r2} + C_{r3} + C_{r4} + C_{r5} = \mathcal{O}(92992Q + 2QT). \tag{65}$$

4.4.2 *For the proposed FL-DRN-NFCE network*

Based on Figure 2 and Table 1, the computational complexity of the 1-st Conv layer in the proposed FL-DRN-NFCE network is

$$C_{f1} = \mathcal{O}(3^2 \cdot 2 \cdot 32 \times (Q+2) \cdot (1+2)) = \mathcal{O}\big(1728(Q+2)\big). \tag{66}$$

In addition, the calculation method for the complexity of the three residual blocks is similar, as follows:

$$C_{f2} = \mathcal{O}\big((1^2 \cdot 32 \cdot 16 + 3^2 \cdot 16 \cdot 16 + 1^2 \cdot 16 \cdot 64 + 1^2 \cdot 32 \cdot 64) \times (Q+2) \cdot (1+2)\big) = \mathcal{O}\big(17664(Q+2)\big),$$
$$C_{f3} = \mathcal{O}\big((1^2 \cdot 64 \cdot 32 + 3^2 \cdot 32 \cdot 32 + 1^2 \cdot 32 \cdot 128 + 1^2 \cdot 64 \cdot 128) \times (Q+2) \cdot (1+2)\big) = \mathcal{O}\big(70656(Q+2)\big),$$
$$C_{f4} = \mathcal{O}\big((1^2 \cdot 128 \cdot 64 + 3^2 \cdot 64 \cdot 64 + 1^2 \cdot 64 \cdot 256 + 1^2 \cdot 128 \cdot 256) \times (Q+2) \cdot (1+2)\big)$$
$$= \mathcal{O}\big(282624(Q+2)\big). \tag{67}$$

The computational complexity of the Conv layer after three residual blocks is as follows:

$$C_{f5} = \mathcal{O}\big(1^2 \cdot 256 \cdot 2 \times (Q+2) \cdot (1+2)\big) = \mathcal{O}\big(1536(Q+2)\big). \tag{68}$$

The computational complexity of the linear layer of the proposed FL-DRN-NFCE network is as follows:

$$C_{f6} = \mathcal{O}\big(\lfloor (Q+1)/2 \rfloor \cdot 1 \cdot 2 \times (N+1) \cdot M \cdot 2\big). \tag{69}$$

Therefore, the total computational complexity of the proposed FL-DRN-NFCE network is

$$C_f = C_{f1} + C_{f2} + C_{f3} + C_{f4} + C_{f5} + C_{f6} = \mathcal{O}\big(374208(Q+2) + 4M(N+1)\big(\lfloor (Q+1)/2 \rfloor\big)\big). \tag{70}$$
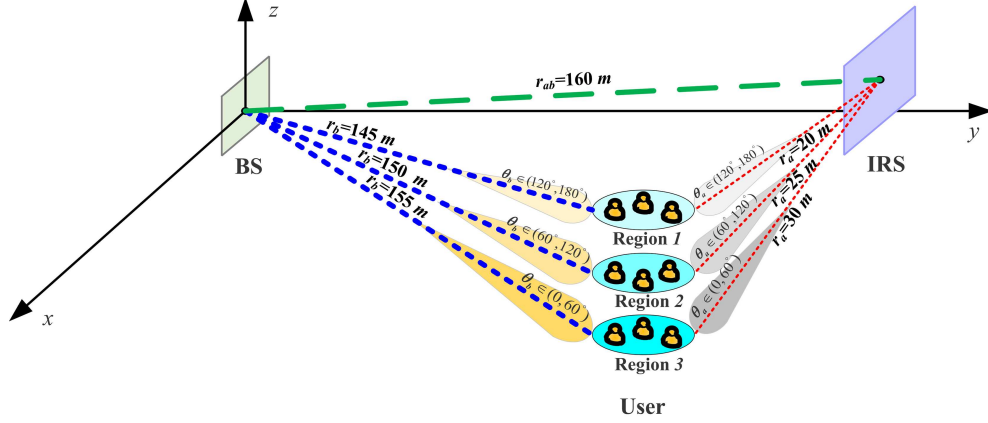
Figure 3   (Color online) Simulation parameter setups for an IRS-aided MU NF communication.

Table 3   Simulation parameter setups for the NF scenario.

Parameter	Value
The azimuth angle of $\boldsymbol{h}_{t,k}$ in three regions	$\theta_b \in \{(0, \frac{\pi}{3}),\ (\frac{\pi}{3}, \frac{2\pi}{3}),\ (\frac{2\pi}{3}, \pi)\}$
The elevation angle of $\boldsymbol{h}_{t,k}$ in three regions	$\phi_b \in (\frac{\pi}{4}, \frac{\pi}{2})$
Distance between the user in three regions and BS	$r_b = \{145, 150, 155\}$ m
The azimuth angle of $\boldsymbol{f}_{t,k}$ in three regions	$\theta_a \in \{(0, \frac{\pi}{3}),\ (\frac{\pi}{3}, \frac{2\pi}{3}),\ (\frac{2\pi}{3}, \pi)\}$
The elevation angle of $\boldsymbol{f}_{t,k}$ in three regions	$\phi_a \in (\frac{\pi}{4}, \frac{\pi}{2})$
Distance between the user in three regions and IRS	$r_a = \{20, 25, 30\}$ m
Distance between BS and IRS	$r_{ab} = r_{ba} = 160$ m
Carrier frequency	$f = 10$ GHz
Wavelength	$\lambda = \frac{c}{f} = \frac{3 \times 10^8}{10 \times 10^9} = 0.03$ m
Element spacing of IRS	$d_{xa} = d_{za} = \frac{\lambda}{2}$
Antenna spacing of BS	$d_{xb} = d_{zb} = \frac{\lambda}{2}$
Width and height of IRS	$L_{xa} = N_{xa} \times d_{xa}$ and $L_{za} = N_{za} \times d_{za}$
Aperture of IRS	$D_{\mathrm{IRS}} = \sqrt{L_{xa}^2 + L_{za}^2}$

5   Simulation results

In this section, the simulation results are given to substantiate the validity of the proposed FL-DRN-NFCE network. Firstly, the configuration of IRS assisted MU NF system is shown in Figure 3, which consists of three regions with three users in each region. Specifically, M and N are considered to be $M = M_{xb} \times M_{zb} = 3 \times 41 = 123$ and $N = N_{xa} \times N_{za} = 3 \times 41 = 123$, respectively. The azimuth angle and elevation angle of the center of BS's UPA relative to the center of IRS's UPA are $\theta_{ba} = \frac{\pi}{4}$ and $\phi_{ba} = \frac{\pi}{4}$. The other specific parameter configurations are detailed in Table 3. Among them, the azimuths of the $\boldsymbol{h}_{t,k}$ and $\boldsymbol{f}_{t,k}$ channels are uniformly distributed in $[0, 2\pi)$ and are equally divided into three parts corresponding to three different user regions in the system. In addition, as depicted in Table 3, the NF distances between users in these three regions and BS and IRS are different. According to [43], r_a and r_{ab} are consistent with the following equation:

$$\frac{r_a \times r_{ab}}{r_a + r_{ab}} < R_{\mathrm{RD}} = \frac{2 \times D_{\mathrm{IRS}}^2}{\lambda}. \tag{71}$$

Then IRS-assisted communication operates in the NF region.

For comparative analysis, in the FF scenario, the distances between users in the three regions and BS and IRS are set as $r_{b,\mathrm{far}} = \{400, 500, 600\}$ m and $r_{a,\mathrm{far}} = \{300, 400, 500\}$ m, respectively. Moreover, the $\boldsymbol{G}$ channel contains $L_{ab} = 15$ NLoS path components, while the NLoS path components of the $\boldsymbol{h}_{t,k}$ and $\boldsymbol{f}_{t,k}$ channels in the three regions are $L_b = \{2, 4, 6\}$ and $L_a = \{3, 5, 7\}$, respectively. Specifically, the distance of the NLoS path components in $\boldsymbol{h}_{t,k}$ is defined to be generated in range of $(R_{\min,b}, R_{\max,b})$ for three regions, namely $R_{\min,b} = \{20, 21, 22\}$ and $R_{\max,b} = \{100, 101, 102\}$. Similarly, for $\boldsymbol{f}_{t,k}$, we have $R_{\min,a} = \{15, 20, 25\}$ and $R_{\max,a} = \{20, 25, 30\}$. Furthermore, the distance of the NLoS path of $\boldsymbol{G}$ is generated within the range of $[30, 120]$ and $[25, 100]$. For the proposed FL-DRN-NFCE network, the
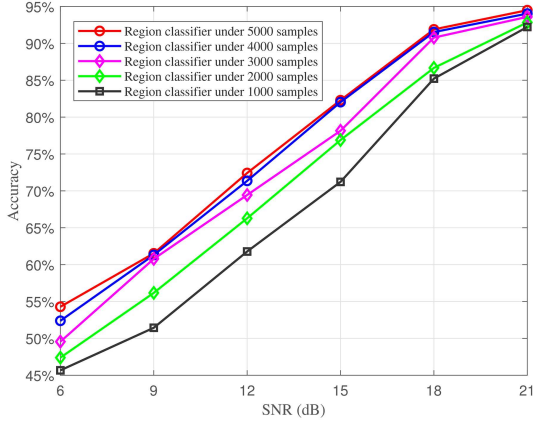
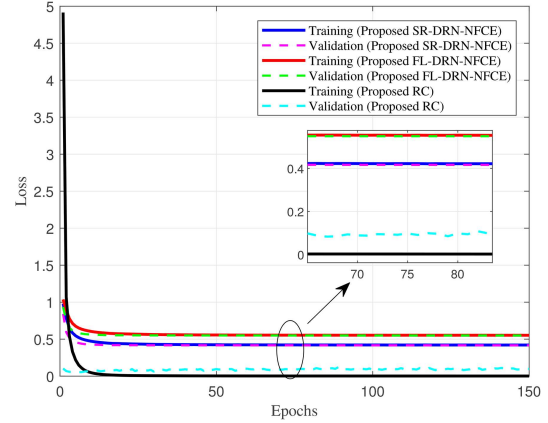**Figure 4** (Color online) Accuracy of the proposed RC network.

Figure 5 (Color online) Training and validation loss of the proposed RC, SR-DRN-NFCE, and FL-DRN-NFCE network versus the training epochs.

required pilot overhead is considered to be set to $Q_1 = \frac{M(N+1)}{6} = 2542$, while the traditional LS and MMSE algorithms require a pilot overhead of $Q_2 = M(N + 1) = 15252$.

As shown in Figure 3, there are a total of 9 users in the 3 regions, and each user collects 5000 training samples. Therefore, for the SR-DRN-NFCE network, $3 \times 5000 = 15000$ training samples are collected, and for the FL-DRN-NFCE network, $3 \times 3 \times 5000 = 45000$ training samples are collected. We allocate 90% of the samples to construct the training dataset, while the remaining 10% is reserved for the validation dataset. The AdamW optimization algorithm is employed to update the model's weight parameters, with an initial learning rate of 1E−3. The training process spans a total of 150 epochs, during which the learning rate is halved every 15 epochs until it reaches the minimum threshold of 1E−6. Upon completion of training, the model's performance is evaluated on the validation dataset. Additionally, the SR-DRN-NFCE network is trained with a batch size of 200, whereas the FL-DRN-NFCE network uses a batch size of 256.

Figure 4 shows the accuracy of the proposed RC network versus different SNRs. From Figure 4, it can be observed that as the number of region samples increases, the classification accuracy gradually improves. When the sample size reaches 4000, continuing to increase the sample size does not improve the accuracy significantly. Thus, each user is set to collect 5000 samples as a training set. Moreover, the proposed RC network can reach 95% accuracy in the high SNR range.

As is clearly evident from Figure 5, the performance of the designed networks has stabilized and converged in terms of loss on both the training and validation sets after 150 epochs, with the losses falling within acceptable ranges. During the training process, to rapidly reduce the networks' losses and effectively mitigate overfitting, we adopted an LR decay strategy. Specifically, at the early stages of training, we initialized the LRs to 1E−3. This relatively high initial value aided the models in quickly capturing the primary features within the data. Following this, we halved the LRs every 15 epochs, progressively decreasing the learning step sizes. The merit of dynamically adjusting the LRs in this fashion is that it not only accelerates the convergence speed of the models, but also significantly boosts their ultimate performance, ensuring stability and accuracy in handling complex tasks.

Figure 6 demonstrates the normalized MSE (NMSE) performance comparison of the proposed SR-DRN-NFCE and FL-DRN-NFCE network versus different SNRs on the testing set from region 1. According to Figure 6, it can be noticed that the proposed SR-DRN-NFCE network trained only on region 1 achieved the lower MSE on the testing set on region 1. However, the proposed SR-DRN-NFCE network trained in regions 2 and 3 did not work well in region 1.

Figure 7 describes the NMSE performance comparison of the proposed SR-DRN-NFCE and FL-DRN-NFCE network versus different SNRs on the testing set from the whole region. It can be clearly remarked that the proposed FL-DRN-NFCE network achieved reliable CE in the entire region's test set, which verifies that the training method based on FL can learn different channel characteristics of more users.

From Figures 6 and 7, it can be seen that both the proposed SR-DRN-NFCE and the proposed FL-DRN-NFCE have smaller errors than their corresponding networks without residual connections, which proves the denoising advantage of residual networks. The LS and MMSE schemes require significant
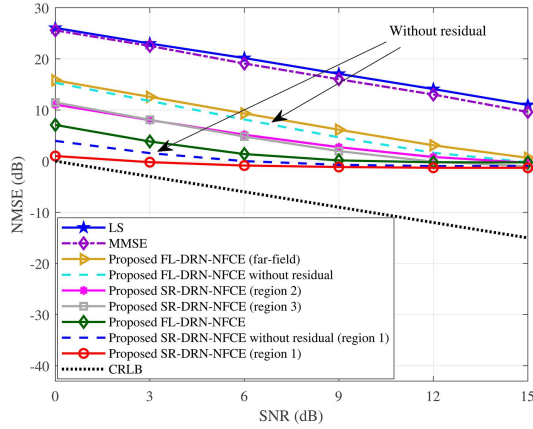
Figure 6 (Color online) NMSE performance comparison for the region 1.
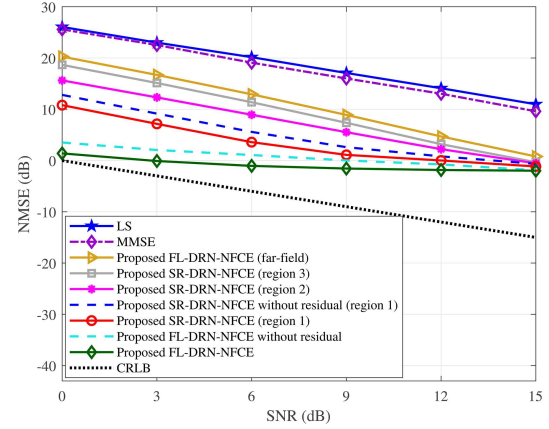
Figure 7 (Color online) NMSE performance comparison for the whole region.

pilot overhead to achieve good estimation accuracy, while the two proposed residual methods can achieve better CE performance while reducing pilot overhead by five-sixths.

6 Conclusion

In this paper, the CE performance of an IRS-aided MU NF MIMO system has been investigated. The LS estimator, MMSE estimator, and CRLB were derived. Subsequently, to reduce the pilot overhead and improve the CE precision, the SR-DRN-NFCE and FL-DRN-NFCE were proposed. Compared with traditional methods such as LS and MMSE, the proposed FL-DRN-NFCE network has successfully reduced the pilot overhead by five-sixths. Simulation results revealed that the proposed FL-DRN-NFCE network presented a lower MSE compared to the scheme without residual connection. The accuracy of the proposed RC approached 95% when the number of user samples was 5000. Finally, the computational complexity of the proposed RC and FL-DRN-NFCE network was delineated. In the whole region, the proposed methods were in ascending order in terms of MSE: FL-DRN-NFCE, SR-DRN-NFCE, MMSE, and LS. In the extreme SNR range, the proposed FL-DRN-NFCE achieves about one-magnitude improvement over SR-DRN-NFCE in terms of MSE performance. For future work, it is necessary to explore potential solutions to address the performance bottleneck of the proposed method and narrow the gap between the proposed method and CRLB, especially at high SNR levels.

References

1 Shu F, Teng Y, Li J, et al. Enhanced secrecy rate maximization for directional modulation networks via IRS. IEEE Trans Commun, 2021, 69: 8388–8401

2 Wang Y, Shu F, Zhuang Z, et al. Asymptotic performance analysis of large-scale active IRS-aided wireless network. IEEE Open J Commun Soc, 2023, 4: 2684–2696

3 Yu G H, Chen X M, Shao X D, et al. Low-cost intelligent reflecting surface aided terahertz multiuser massive MIMO: design and analysis. Sci China Inf Sci, 2021, 64: 200302

4 Hu X L, Wang J W, Zhong C J. Statistical CSI based design for intelligent reflecting surface assisted MISO systems. Sci China Inf Sci, 2020, 63: 222303

5 Tao Q, Zhang S W, Zhong C J, et al. Joint information transmission design for intelligent reflecting surface aided system with discrete phase shifts. Sci China Inf Sci, 2023, 66: 132303

6 Wang X, Shu F, Shi W, et al. Beamforming design for IRS-aided decode-and-forward relay wireless network. IEEE Trans Green Commun Netw, 2022, 6: 198–207

7 Zheng B, You C, Mei W, et al. A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications. IEEE Commun Surv Tut, 2022, 24: 1035–1071

8 Wang S, Li Q, Lin J. One-bit channel estimation for IRS-aided millimeter-wave massive MU-MISO system. IEEE Trans Signal Process, 2023, 71: 3592–3606

9 Kim I S, Bennis M, Oh J, et al. Bayesian channel estimation for intelligent reflecting surface-aided mmWave massive MIMO systems with semi-passive elements. IEEE Trans Wireless Commun, 2023, 22: 9732–9745

10 Chen Z, Zhao M M, Li M, et al. Joint location sensing and channel estimation for IRS-aided mmWave ISAC systems. IEEE Trans Wireless Commun, 2024, 23: 11985–12002

11 Guan X, Wu Q, Zhang R. Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication. IEEE Trans Wireless Commun, 2022, 21: 3764–3778
12 Zheng B, Lin S, Zhang R. Intelligent reflecting surface-aided LEO satellite communication: cooperative passive beamforming and distributed channel estimation. IEEE J Sel Areas Commun, 2022, 40: 3057–3070
13 Li Q, El-Hajjar M, Xu C, et al. Stacked intelligent metasurfaces for holographic MIMO-aided cell-free networks. IEEE Trans Commun, 2024, 72: 7139–7151
14 Li Q, El-Hajjar M, Sun Y, et al. Energy-efficient reconfigurable holographic surfaces operating in the presence of realistic hardware impairments. IEEE Trans Commun, 2024, 72: 5226–5238
15 Li Q, El-Hajjar M, Sun Y, et al. Achievable rate analysis of the STAR-RIS-aided NOMA uplink in the face of imperfect CSI and hardware impairments. IEEE Trans Commun, 2023, 71: 6100–6114
16 Ye M, Zhang H, Wang J B. Channel estimation for intelligent reflecting surface aided wireless communications using conditional GAN. IEEE Commun Lett, 2022, 26: 2340–2344
17 Ye M, Liang X, Pan C, et al. GNN-based channel estimation for intelligent reflecting surface aided multiuser systems relying on user locations. IEEE Wireless Commun Lett, 2024, 13: 2110–2114
18 Shi H, Huang Y, Jin S, et al. Automatic high-performance neural network construction for channel estimation in IRS-aided communications. IEEE Trans Wireless Commun, 2024, 23: 10667–10682
19 Chu H, Pan X, Jiang J, et al. Adaptive and robust channel estimation for IRS-aided millimeter-wave communications. IEEE Trans Veh Technol, 2024, 73: 9411–9423
20 Xu W, Wu J, Jin S, et al. Disentangled representation learning empowered CSI feedback using implicit channel reciprocity in FDD massive MIMO. IEEE Trans Wireless Commun, 2024, 23: 15169–15184
21 Wei X, Dai L. Channel estimation for extremely large-scale massive MIMO: far-field, near-field, or hybrid-field? IEEE Commun Lett, 2022, 26: 177–181
22 Cui M, Dai L. Channel estimation for extremely large-scale MIMO: far-field or near-field? IEEE Trans Commun, 2022, 70: 2663–2677
23 Wang D W, Li X R, He Y X, et al. Intelligent reflecting surface assisted untrusted NOMA transmissions: a secrecy perspective. Sci China Inf Sci, 2023, 66: 192302
24 Pisharody J N, Rajoriya A, Budhiraja R. Near-field channel estimation for XL-MIMO systems using variational Bayesian learning. IEEE Trans Wireless Commun, 2024, 23: 10740–10756
25 Huang C, Xu J, Xu W, et al. Low-complexity channel estimation for extremely large-scale MIMO in near field. IEEE Wireless Commun Lett, 2024, 13: 671–675
26 Guo X, Chen Y, Wang Y. Compressed channel estimation for near-field XL-MIMO using triple parametric decomposition. IEEE Trans Veh Technol, 2023, 72: 15040–15045
27 Zhang X, Wang Z, Zhang H, et al. Near-field channel estimation for extremely large-scale array communications: a model-based deep learning approach. IEEE Commun Lett, 2023, 27: 1155–1159
28 Yang S, Lyu W, Hu Z, et al. Channel estimation for near-field XL-RIS-aided mmWave hybrid beamforming architectures. IEEE Trans Veh Technol, 2023, 72: 11029–11034
29 Dai L, Wei X. Distributed machine learning based downlink channel estimation for RIS assisted wireless communications. IEEE Trans Commun, 2022, 70: 4900–4909
30 Liu Y, Wang Z, Xu J, et al. Near-field communications: a tutorial review. IEEE Open J Commun Soc, 2023, 4: 1999–2049
31 Starer D, Nehorai A. Passive localization of near-field sources by path following. IEEE Trans Signal Process, 1994, 42: 677–680
32 Björnson E, Demir Ö T, Sanguinetti L. A primer on near-field beamforming for arrays and reconfigurable intelligent surfaces. In: Proceedings of the 55th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, 2021. 105–112
33 Biguesh M, Gershman A B. Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals. IEEE Trans Signal Process, 2006, 54: 884–893
34 Rao X B, Lau V K N. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. IEEE Trans Signal Process, 2014, 62: 3261–3271
35 Liu C, Liu X, Ng D W K, et al. Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications. IEEE Trans Wireless Commun, 2022, 21: 898–912
36 Lu Y, Dai L. Near-field channel estimation in mixed LoS/NLoS environments for extremely large-scale MIMO systems. IEEE Trans Commun, 2023, 71: 3694–3707
37 He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 770–778
38 Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. 4510–4520
39 Xu W, Yang Z, Ng D W K, et al. Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing. IEEE J Sel Top Signal Process, 2023, 17: 9–39
40 Yao J, Xu W, Yang Z, et al. Wireless federated learning over resource-constrained networks: digital versus analog transmissions. IEEE Trans Wireless Commun, 2024, 23: 14020–14036
41 Dong P, Zhang H, Li G Y, et al. Deep CNN-based channel estimation for mmWave massive MIMO systems. IEEE J Sel Top Signal Process, 2019, 13: 989–1000
42 Liu J J, Hou Q, Liu Z A, et al. PoolNet+: exploring the potential of pooling for salient object detection. IEEE Trans Pattern Anal Mach Intell, 2023, 45: 887–904
43 Cui M, Wu Z, Lu Y, et al. Near-field MIMO communications for 6G: fundamentals, challenges, potentials, and future directions. IEEE Commun Mag, 2023, 61: 40–46