Machine Learning-Based Channel Estimation for Multi-RIS-Assisted mmWave Massive-MIMO OFDM System in a Dynamic Environment

Thabang C. Rapudu and Olutayo O. Oyerinde, *Senior Member, IEEE*

***Abstract*— Wireless communication in high obscurity environments is not effective especially for millimeter-wave (mmWave) frequencies due to propagation challenges. It is therefore, necessary to deploy multiple reconfigurable intelligent surfaces (multi-RISs) and massive multiple-input multiple-output (massive-MIMO) systems to circumvent the propagation challenges in such environments for effective communication. However, deploying massive-MIMO multi-RIS system increases the dimensionality of the channel, and using conventional channel estimation (CE) methods for estimating this multi-RIS-assisted channel is infeasible. In literature, machine learning (ML)-based CE methods have been proven to yield more accurate CE results for massive-MIMO RIS-assisted systems. Thus, in this paper, the ML-based CE method called the denoising convolutional neural network-gated recurrent unit (DnCNN-GRU) scheme is proposed for estimating the uplink cascaded time-varying massive-MIMO multi-RIS-assisted channel. The proposed scheme was then benchmarked with other CE methods to prove its superiority. It achieved a performance closer to the lower bound oracle least square (LS) estimate with about 1 dB gap in terms of the normalised mean squared error (NMSE).**

***Index Terms*— Channel estimation, RIS, machine learning, GRU, CNN.**

I. Introduction

DUE to its high frequencies,the millimeter-wave (mmWave) band for wireless communication suffers from propagation challenges like high path loss and atmospheric absorption, reducing the effective communication range between transceivers. As a solution to circumvent these challenges, the reconfigurable intelligent surfaces (RISs) and massive multiple-input multiple-output (massive-MIMO) systems are implemented in mmWave wireless communication environments [1]. An RIS is a reflective planar structure consisting of programmable passive elements which can be reconfigured dynamically to suit the wireless communication environment [2]. Specifically, the programmable passive elements of an RIS have the capability of changing the reflection properties of an incident signal and thus, they can control the incident signal towards desired directions. This results in an improved signal coverage. In the case where the line-of-sight (LoS) between the transceivers is blocked, an RIS can play a positive role by creating an alternative pathway of the incident signals around the blockage. In addition, massive-MIMO technology is the use of a very large number of antenna arrays for the improvement of signal energy and spectral efficiency. With massive-MIMO [3], signal interference and fading can be mitigated. Furthermore, massive-MIMO enables beamforming.

However, the acquisition of the channel state information (CSI) by the process of channel estimation (CE) for massive-MIMO RIS-assisted systems is not an easy task. Specifically, deploying RISs in a MIMO communication system involves multiple links/channels (i.e., channels between users and RIS, and between RIS and base station (BS)) which significantly increases the number of channel coefficients to estimate [2]. Furthermore, the RIS elements are typically passive in nature (i.e., lack active radio frequency (RF) chains to directly process signals at the RIS), requiring the CE process to be done at the receiving terminal (i.e., at the BS or user). All these increase the dimensionality of the wireless communication channel and thus, conventional methods like least squares (LS) and minimum mean square error (MMSE) are impractical for estimating the RIS-assisted massive-MIMO channels. In [4], an orthogonal matching pursuit (OMP)-based CE technique was proposed for wideband RIS-assisted mmWave system. By leveraging on the sparsity property of the uplink cascaded RIS-assisted channel, the adaptive block simultaneous orthogonal matching pursuit (ABSOMP) algorithm which is robust to the wideband effect was proposed. However, the ABSOMP algorithm is highly sensitive to noise and does not perform well in low signal-to-soise Ratio (SNR) environments, resulting in inaccurate channel estimates. Furthermore, the ABSOMP algorithm's complexity increases significantly with the number of receiver's/transmitter's antennas and/or RIS's reflecting element. By exploiting the polar-domain sparsity of the wideband RIS-assisted channel, the polar-domain frequency-dependent block orthogonal least squares (PF-BOLS) algorithm was proposed in [5] for CE. However, the core of the PF-BOLS algorithm is a least square (LS) algorithm. Similarly, the LS-based RIS-assisted CE method was proposed in [6]. These LS-based CE techniques cannot optimally exploit the sparsity of the RIS-assisted channel, leading to poor CE performance compared to, for example, OMP-based algorithms which can better exploit the sparsity of the RIS-assisted channels. In [7], a Bayesian framework called

Received 20 August 2024; revised 15 December 2024; accepted 23 February 2025. Date of publication 7 March 2025; date of current version 12 June 2025. The associate editor coordinating the review of this article and approving it for publication was S. Zhou. *(Corresponding author: Thabang C. Rapudu.)*

The authors are with the School of Electrical and Information Engineering, University of the Witwatersrand at Johannesburg, Johannesburg 2000, South Africa (e-mail: 1168108@students.wits.ac.za).

Digital Object Identifier 10.1109/TWC.2025.3546671

a generalized approximating message passing (GAMP) was adopted and assumed Bernoulli-Gaussian (BG) distribution. Thereafter, the conventional minimum mean squared error (MMSE) estimator was used for cascaded RIS-assisted CE. However, the MMSE estimator is computationally expensive especially for large-scale RIS (i.e., with many reflecting elements) systems. Additionally, the MMSE estimator should know the statistical information of the channel for accurate estimates. This statistical information is challenging to obtain in RIS-assisted channels.

In [8], [9], [10], and [11], efforts were made to minimise pilot overhead for RIS-assisted CE. For this purpose, the two-timescale CE scheme was proposed for the reduction of overall pilot overhead. That is, since the BS and the RIS are relatively static than the user equipment (UE), the coherence time of the RIS-BS channel is assumed to be longer than the UE-RIS channel. Therefore, the RIS-BS channel can be estimated only once during the large-timescale. In [8] and [9], the UE-RIS channels were then estimated using the conventional LS-based algorithm. However, if the UE-RIS channel becomes highly dynamic, the LS-based schemes in [8] and [9] will significantly deteriorate. In [10], CE is performed by sequentially transmitting sensing signals for each RIS element at a time. As a consequence, this method is computationally expensive. In [11], a linear minimum mean square error (LMMSE) estimator was proposed to estimate both the UE-RIS and RIS-BS channels. However, the LMMSE estimator is impractical for real-time large scale RIS assisted channel scenarios due to its computational cost.

On the other hand, machine learning (ML) techniques have been used for effectively estimating the single and double RIS-assisted massive-MIMO channels and they are promising techniques for multi-RIS-assisted CE. In [12], a gradient-descent-based deep-iterative-unrolling network (GD-Net) was proposed to estimate the simultaneously transmitting and reflecting RIS (STAR-RIS) assisted channel. Unlike the standard RIS which is fully passive in nature, the STAR-RIS can transmit and reflect signals. However, the STAR-RIS is more expensive since it requires more radio frequency (RF) chains for transmission than the standard RIS. In [13] and [14], deep learning based algorithms named the convolutional neural network based on super-resolution convolutional neural network (SRCNN) and joint CE and CSI feedback Network (JDCNet) for RIS-assisted channels were proposed respectively. However, these schemes require high pilot overhead since the UE-RIS and RIS-BS channels were iteratively estimated together during a single timescale.

In [15], the ML approach was used to estimate the RIS-assisted channel using the autoencoder. The neural network of the autoencoder was pruned by cutting off the insignificant neurons using the average percentage of zeros(APoZ) metric. However, the pruning of the autoencoder increases the complexity of the CE algorithm and it will be significant in high dimensional CE problems like in multi-RIS-assisted systems. Conversely, in [16], the convolutional neural network (CNN) was proposed for RIS-assisted massive-MIMO CE in both frequency-flat and frequency-selective channels to reduce the pilot overhead and complexity of the wireless system. The CNN architecture was used to estimated both the channel between UEs and RIS (i.e., the UEs-RIS channel), and between RIS and the BS. However, the UEs were assumed to be static. For mobile UEs, the performance of the proposed denoising CNN (DnCNN) architecture will deteriorate for the UEs-RIS CE as it was not optimally designed for fast time-varying data. In [17], the ML-based algorithm called the global attention residual network (GARN) was used to estimate the uplink RIS-assisted channel in the outdoor and indoor scenarios with the aim of reducing system's complexity and pilot overhead. The GARN fuses the multi-channel information in order to improve the channel estimation accuracy. Compared to the enhanced deep super-resolution (EDSR) algorithm [18], the GARN performs better in both indoor and outdoor scenarios in terms of normalized mean squared error (NMSE) especially in high SNR scenarios. However, the GARN uses the estimated channel by the conventional LS algorithm as an input which affects the CE accuracy. This LS-based CE will significantly deteriorate for high dimensionality multi-RIS-assisted channel as it does not perform well in non linear data. In [19], [20], and [21], the long short-term memory (LSTM) deep learning based network was used to estimate the RIS-assisted channels. However, in [19], the LSTM was concatenate with the CNN for the complete cascaded CE. The CNN-LSTM can be further improved for accurate CE and system's complexity since the effort was not done to utilise latest normalisation and regularisation methods like batch normalisation (BN) to optimise the CNN-LSTM architecture. The lack of these regularisation methods for the CNN-LSTM for CE results in slower convergence especially in high dimensionality scenarios like in multi-RIS-assisted channels. In [20], the bidirectional-LSTM (Bi-LSTM) was used to estimate the downlink channel. This method used high pilot overhead. Conversely, the LSTM was used to estimate the uplink RIS-assisted channel only between the UEs to the RIS in [21]. However, the method of active-RIS-elements was implemented which requires addition power to activate the RIS elements. Additionally, this method requires extra RF chain which makes the system more costly than a complete passive RIS elements method. Consequently, for multi-RIS-assisted system, even more RF chains will be required, making the system even more costly.

The time-varying channel aspect further adds up to the complexity of the RIS-assisted systems. Therefore, an extra consideration should be done for time-varying RIS-assisted systems. In [22] and [23], a single input single output (SISO) RIS-assisted downlink system was considered with a high mobility UE. Therefore, the MMSE-based interpolation was proposed for CE which can fairly estimate the time-varying cascaded channel. However, the MMSE method can only perform good in moderate time-varying RIS-assisted channels. In [24], a fast time-varying RIS-assisted multiple-user multiple-input single-output (MU-MISO) channel was considered and the two-timescale property was adopted. The channel between the BS and RIS was estimated utilising one specific full-duplex antenna (FDA) with the assumption of channel's reciprocity. Thereafter, the channel between the UEs and the RIS was estimated using a deep neural network. However, this method will not be feasible in cases where the BS-RIS channel is not reciprocal.

Nevertheless, the CE methods in [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], and [24] were implemented in a single-RIS-assisted wireless communication system which is only suitable in the environments where there is low obscurity. Recently, the use of multiple RISs in a communication system has been explored to improve efficiency. In [25] and [26], the stacked intelligent metasurfaces (SIM) technology has been implemented to enhance better MIMO systems' performance, resulting in a variant of MIMO system called the holographic MIMO (HMIMO). In this case, a pair of stacked multiple RISs are deployed at the transmitter and receiver respectively [27]. This setup constructs multiple parallel subchannels, enabling the data streams to be directly radiated from the transmit and recovered at the receiver without imposing any interference [28].

The multi-RIS systems in [25] and [26] focuses on LoS scenarios. That is, the RISs are not utilised to circumvent blockages. In high obscurity environments like in urban areas where there are many buildings, trees and crowded, the methods in [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26] will not be practical. Therefore, there is a need for the implementation of multiple RISs in such environments to deal with high obscurity for effective wireless communication where there is no LoS. Consequently, there is a need for robust CE methods in such cases.

In [29], it was shown that the wireless communication system's performance can be significantly improved by increasing the number of RISs (consequently increasing the number of RIS elements). Inspired by this, the Two-Timescale (large and small-timescale) scheme was proposed in [30] to estimate the double-RIS-assisted uplink channel using the singular value decomposition multiple measurement vector-based compressive sensing (SVD-MMV- CS) framework. Both RISs were assumed to be active-type and to limit hardware cost, one RF chain was connected to all RIS elements as opposed to committing one RF chain per RIS element. However, the UEs were assumed to be in LoS with both the RISs, and the UEs-RISs channels were assumed to be known. Therefore, the task was to estimate the channel between the RISs and the channel between the BS and the second RIS, taking advantage of the known UEs-RISs channels, which simplified the CE of the UE-RIS channels as in single-RIS scenarios. When there is no LoS between the UEs and both RISs, and all the channels were unknown, the proposed method will be more complex and the proposed CE scheme will not be feasible. Similar to [30], a double-RIS-assisted CE scheme was proposed in [31]. The CE was performed using the skip-connection (SC) based CNN. This scheme was proposed for stationary UEs and may not perform well for mobile UEs. In [32], the parallel factor analysis (PARAFAC) decomposition model was used to represent the received signal in a double-RIS-assisted channel and then the conventional alternating least squares (ALS) algorithm was used to estimate the channel. However, the ALS method has a slow convergence rate for large datasets and therefore, it will deteriorate for larger RISs.

This paper is motivated by firstly, the necessity for multi-RIS-assisted systems in urban areas to serve as an alternative pathway around multiple blockages between a transmitter and a receiver. The benefits of multi-RIS-assisted systems in such high obscurity environments were investigated and discussed in [29]. Mainly, it was found that increasing the number of RISs and/or number of RIS elements improves both the ergodic capacity and outage probability. Secondly, as a consequence of implementing a multi-RIS-assisted massive-MIMO system, the dimensionality of the channel between a transmitter and a receiver significantly increases and thus, conventional CE schemes (i.e., CE schemes based on LS,MMSE, OMP, etc.) proposed as in [4], [5], [6], [7], [8], [9], [10], and [11] and [22], [23], [24] will be impractical due to their inability to handle highly non-linear data. Conversely, ML-based techniques are promising solutions for multi-RIS-assisted massive-MIMO CE. However due to the discussed downsides, ML-based schemes proposed as in [12], [13], [14], [15], [16], [17], [18], [19], [20], and [21] are also impractical. Therefore, in this paper, a robust data-driven based CE scheme is proposed for the multi-RIS-assisted mmWave massive-MIMO OFDM system. The main contributions of this paper are as follows:

- Firstly, a generic mmWave massive-MIMO multi-RIS-assisted system model of K RISs is formulated based on the Saleh-Valenzuela (SV) geometric channel model. Assuming that the base station (BS) and all the RISs are at fixed positions, and the UEs are mobile, the two-timescale channel model is adopted as in [30]. That is, since the all RISs and the BS are assumed to be static, then the uplink cascaded channel from the first RIS all the way to the BS is assumed to be quasi-static. Therefore, this quasi-static cascaded channel is estimated only once during a large-timescale coherence time. The uplink channel between the UEs and the first RIS is thereafter estimated during a small-timescale coherence time. This is for the purpose of significantly reducing the pilot overhead.
- One of the major setbacks of dealing with multi-RIS-assisted channels is the multiple channels which are difficult to be estimated individually. In this paper, the RIS reflecting coefficients matrices are formulated as a combined reflecting coefficients matrix of the cascaded channel from UEs to the BS, which simplifies the problem. This is achieved by assuming that the number of RIS elements for all RISs are equal.
- Thereafter, ML-based techniques are proposed for estimating the multi-RIS-assisted cascaded channels from the mobile UEs to the BS. Since the uplink channels from the first RIS to the BS are assumed to be quasi-static, the DnCNN is proposed for estimating these quasi-static channels. The BN is incorporated in the proposed DnCNN for fast learning. Due to the mobility of the UEs, the channel from the UEs to the first RIS is consequently time-varying and therefore, the gated recurrent unit (GRU) architecture is proposed for CE

in this case. To prevent overfitting, the neural dropout is incorporated within every GRU. Furthermore, instead of just learning directly from the DnCNN/GRU outputs as in conventional DnCNN/GRU schemes, the proposed DnCNN-GRU scheme learns the residual for faster and more accurate estimates.

- Finally, comparative performance of the proposed CE via computer simulations is documented. Although the simulation results that are provided are for the double-RIS-assisted scheme, the discussion is provided about how the proposed ML-based CE scheme generalises for a wireless system with K RISs. The double-RIS-assisted ML-based scheme was benchmarked with various double-RIS-assisted CE schemes.

Paper layout: In Section II, the system model and problem formulation for the generic multi-RIS-assisted channel are discussed. Thereafter, in Section III, the ML-based channel estimation scheme is presented. Here, a well detailed explanation of the internal processes of the proposed CE scheme is provided using illustrative diagrams. Section IV presents the simulation results of the convergence of the proposed ML-based scheme. Furthermore, the performance of the proposed CE scheme is benchmarked against previously proposed CNN with skip-connection (CNN-SC [31], conventional linear minimum mean square error (LMMSE) channel estimator [31], and the oracle LS in literature. Finally, Section V concludes the paper.

Notation: The matrices and vectors are represented by the upper case bold and lower case bold letters respectively, while the $||\cdot||_2$ and $||\cdot||_F$ are the $\mathbf{L}_2$ and Frobenius norms respectively. The $\mathbf{I}_X$, $(\cdot)^T$ and $(\cdot)^H$ are the $X \times X$ identity matrix, transpose and Hermitian, respectively. The diag$(\mathbf{x})$ produces a diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal; vec $(\mathbf{X})$ denotes the vectorisation of a matrix $\mathbf{X}$; $\mathbb{C}^{x\times y}$ represents the space of $x \times y$ complex-valued matrices; the $\otimes$ and $\odot$ denote the Kronecker and Hadamard product respectively.

II. System Model and Problem Formulation

In this section, a generic multi-RIS-assisted system model of multiple RISs is firstly formulated and thereafter, the problem formulation is provided. TABLE I shows all major symbols used in this paper.

A. A Generic System Model of K RISs

Consider a wideband multi-RIS-assisted mmWave massive-MIMO OFDM wireless communication system consisting of a BS, N single-antenna UEs and K RISs (i.e., $RIS_1, RIS_2, \ldots, RIS_K$) as shown in Fig. 1. The BS is equipped with $M = M_x \times M_y$ uniform planar array (UPA) of antenna elements while RIS_k has P_k number of reflecting elements where $k \in \{1, 2, \ldots, K\}$. All the RISs and the BS are located at fixed positions while the UEs are mobile, and all elements on the BS and the RISs are assumed to have d_e spacing between them. It is further assumed that the distances between the RISs are very larger than d_e and all the channels are frequency-selective. In this paper, the CE is performed for

TABLE I

Symbols and Their Definations

Symbol	Defination
$\mathbf{A}$	A complex dictionary matrices.
$\mathbf{a}$	A steering vector.
$\mathbf{\Delta}_n$	The sparse matrix associated with the cascaded channel for the n-th UE.
$\mathbf{G}$	The uplink channel between the UEs and the first RIS.
$\mathbf{H}$	The uplink channel between the last RIS and the BS.
K	The total number of RISs.
k	The k-th RIS.
M	The total number of BS antennas.
N	The total number of UEs.
n	The n-th UE.
N_{sc}	The total number of subcarriers.
n_{sc}	The n_{sc}-th subcarrier.
P_k	The total number of reflecting elements of the k-th RIS.
Q	The total number of time slots.
q	The q-th time slot.
$\mathbf{R}_j$	The j-th uplink channel between two adjacent RISs.
$\mathbf{\Theta}_n$	The combined reflecting coefficient matrix of all RISs for the n-th UE
$\mathbf{V}_n$	The uplink cascaded channel between the n-th UE and the BS.
$\mathbf{w}_{n,q}$	The BS's AWGN of the n-th UE at the q-th time slot.
$\mathbf{y}_{n,q}$	The BS received signal of the n-th UE at the q-th time slot.

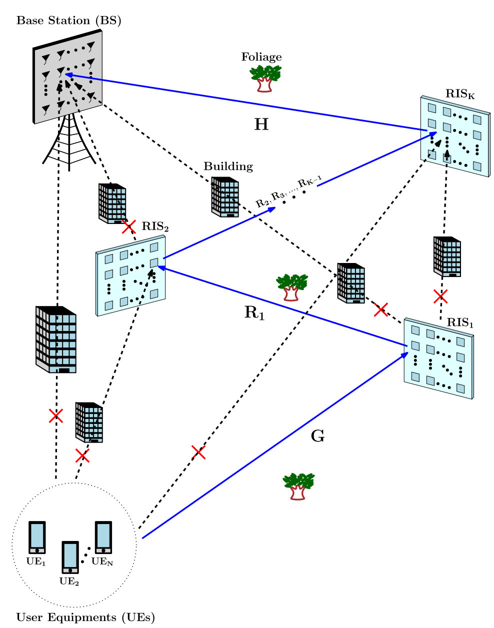


Fig. 1. A generic system model of K RISs.

the uplink channels where the direct LoS between the UEs and the BS is blocked. Moreover, the UEs can only send data through RIS_1 and the data can only be received by the BS through the K-th RIS. This wireless environment is assumed to have high obscurity due to many buildings and foliage and therefore, the k-th RIS is assumed to be in LoS with only two RISs adjecent to it (i.e., the k-th RIS can only receives and reflects signals from the $(k-1)$-th RIS to the $(k+1)$-th RIS respectively).

Suppose that the OFDM block of the system has B transmission bandwidth and N_{sc} subcarriers with $d_{sc} = \frac{B}{N_{sc}}$ as the subcarrier spacing. Furthermore, the Cyclic Prefix (CP) is assumed to be long enough to overcome the discontinuity between the received signals that may happen due to the multi-path signal delays. Let $\mathbf{g}_n \in \mathbb{C}^{P_1 \times 1}$ and $\mathbf{H}_n \in \mathbb{C}^{M \times P_K}$ be the channel from the n-th UE to the RIS_1 and the channel of the n-th UE from the RIS_K to the BS respectively, where $n \in \{1, 2, \ldots, N\}$, and the channels of the n-th user between adjacent RISs from RIS_1 to RIS_K be $\{\mathbf{R}_{1,n} \in \mathbb{C}^{P_2 \times P_1}, \mathbf{R}_{2,n} \in \mathbb{C}^{P_3 \times P_2}, \ldots, \mathbf{R}_{K-1,n} \in \mathbb{C}^{P_K \times P_{K-1}}\}$, where $\mathbf{R}_{k,n}$ is the channel of the n-th user between the (k)-th and the $(k+1)$-th RISs. Since the users share the $\mathbf{H}_n$ and $\mathbf{R}_{:,n}$ channels, the subscript n can be omitted.

Let $L_{G,n}$ be the number of paths between the UE_n and RIS_1, and τ_{l_g} be the time delay of the l_g-th path from the UE_n to RIS_1 where $l_g \in \{1, 2, \ldots, L_{G,n}\}$. Therefore, by Saleh-Valenzuela (SV) geometric channel model [16], the d-th time domain delay tap of the channel $\mathbf{g}_n$ at the n_{sc}-th subcarrier can be represented by

$$\mathbf{g}_{d,n}(n_{sc}) = \mathcal{G} \sum_{l_g=1}^{L_{G,n}} \alpha_{l_g} p_{rc}\left(dT_s - \tau_{l_g}\right) \mathbf{a}^r_{n_{sc}}\left(\vartheta^{r,RIS_1}_{l_g}, \psi^{r,RIS_1}_{l_g}\right) e^{j\frac{v_n f_{sc}}{c} cos\theta_{l_g} dT_s}, \tag{1}$$

where the time domain delay tap length is equal to D_G where $d \in \{0, 1, \ldots, D_G - 1\}$, $n_{sc} \in \{1, 2, \ldots, N_{sc}\}$ and $\mathcal{G} = \sqrt{\frac{P_1}{L_G}}$. The α_{l_g}, T_s and τ_{l_n} are the complex gain of the l_g-th path, sampling period and the delay respectively; the v_n, f_{sc}, c and θ_{l_g} are the speed of the n-th user, subcarrier frequency, speed of light and the angle between the incident direction of the subcarrier frequency and the movement direction of UE for the l_g-th path respectively. Lastly, $p_{rc}(\tau)$ accounts for a low pass filtering and a filter incorporating the pulse-shaping effects evaluated at τ. The parameter $\mathbf{a}^r_{n_{sc}}(\cdot, \cdot)$ is a steering vector. This channel experiences time-selective fading due to UE mobility.

Similarly, the d-th time domain delay tap of channel $\mathbf{R}_j$ (i.e., the channel between the k-th and $(k+1)$-th RISs) at the n_{sc}-th subcarrier, where $j \in \{1, 2, \ldots, K-1\}$ is given by

$$\mathbf{R}_{d,j}(n_{sc}) = \mathcal{R}_{R_j} \sum_{l_{R_j}=1}^{L_{R_j}} \alpha^{R_j}_{l_{R_j}} p_{rc}\left(dT_s - \tau_{l_{R_j}}\right) \mathbf{a}^r_{n_{sc}}\left(\vartheta^{r,RIS_{k+1}}_{l_{R_j}}, \psi^{r,RIS_{k+1}}_{l_{R_j}}\right) \mathbf{a}^{t\;H}_{n_{sc}}\left(\vartheta^{t,RIS_k}_{l_{R_j}}, \psi^{t,RIS_k}_{l_{R_j}}\right), \tag{2}$$

where L_{R_j}, $\alpha^{R_j}_{l_{R_j}}$ and $\tau_{l_{R_j}}$ are the number of paths, complex gain of the l_{R_j}-th path and time delay between the k-th and $(k+1)$-th RISs respectively; D_{R_j} is the time domain tap length of the R_j-th channel where $d \in \{0, 1, \ldots, D_{R_j} - 1\}$, $l_{R_j} \in \{1, 2, \ldots, L_{R_j}\}$ and $\mathcal{R}_{R_j} = \sqrt{\frac{P_{k+1} \times P_k}{L_{R_j}}}$.

Furthermore, the d-th time domain delay tap of the channel between the last RIS (i.e., RIS_K) and the BS is denoted by

$$\mathbf{H}_d(n_{sc}) = \mathcal{H} \sum_{l_H=1}^{L_H} \alpha_{l_H} p_{rc}\left(dT_s - \tau_{l_H}\right) \mathbf{a}^r_{n_{sc}}\left(\vartheta^{r,BS}_{l_H}, \psi^{r,BS}_{l_H}\right) \mathbf{a}^{t\;H}_{n_{sc}}\left(\vartheta^{t,RIS_K}_{l_H}, \psi^{t,RIS_K}_{l_H}\right). \tag{3}$$

In this case, D_H, L_H, α_{l_H} and τ_{l_H} are the time domain tap length, number of paths, complex gain of the l_H-th path and time delay between RIS_K and the BS respectively, $d \in \{0, 1, \ldots, D_H - 1\}$, $l_H \in \{1, 2, \ldots, L_H\}$ and $\mathcal{H} = \sqrt{\frac{M \times P_K}{L_H}}$.

In (1), (2) and (3), $\mathbf{a}^t_{n_{sc}}(\cdot, \cdot)$ and $\mathbf{a}^r_{n_{sc}}(\cdot, \cdot)$ are the normalised UPA steering complex column vectors of length equivalent to the number of antenna/reflecting elements of the transmitting/reflecting and receiving nodes of the n_{sc}-th subcarrier respectively, and $\mathbf{a}^{t\;H}_{n_{sc}}(\cdot, \cdot)$ is the Hermitian transpose of the vector $\mathbf{a}^t_{n_{sc}}(\cdot, \cdot)$; $\vartheta^{t,\cdot}_{l_i}$ and $\psi^{t,\cdot}_{l_i}$ are the azimuth and elevation angles of the transmitting/reflecting node of the l_i-th path respectively, where i is the channel in subject; $\vartheta^{r,\cdot}_{l_i}$ and $\psi^{r,\cdot}_{l_i}$ are the arrived azimuth and elevation angles of the receiving node of the l_i-th path respectively.

Generally, the UPA steering complex column vector is given by

$$\mathbf{a}(\vartheta, \psi) = \frac{1}{\sqrt{N_{UPA}}} e^{-j\frac{d_{e_x}(\vartheta,\psi)}{\lambda_c}\mathbf{n}_{UPA_x}} \otimes e^{-j\frac{d_{e_y}(\psi)}{\lambda_c}\mathbf{n}_{UPA_y}} \tag{4}$$

where $N_{UPA} = N_{UPA_x} \times N_{UPA_y}$ and λ_c are the number of UPA elements and wavelength of the carrier respectively; $\mathbf{n}_{UPA_x} \in \{0, 1, \ldots, N_{UPA_x} - 1\}$, $\mathbf{n}_{UPA_y} \in \{0, 1, \ldots, N_{UPA_y} - 1\}$, $d_{e_x}(\vartheta, \psi) = 2\pi d_e sin(\vartheta) cos(\psi)$ and $d_{e_y}(\psi) = 2\pi d_e sin(\psi)$, where the element of spacing is given as $d_e = \lambda_{sc}/2$. Furthermore, the complex gains (i.e., α_{l_g}, $\alpha^{Rj}_{l_{R_j}}$ and α_{l_H}) in (1), (2) and (3) are modelled as $\alpha_{l_i} = |\alpha_{l_i}| e^{j\phi_{l_i}}$, where ϕ_{l_i} is the phase of the l_i-th path of the i-th channel and $|\alpha_{l_i}| \in (0, 1]$ is the amplitude of the l_i-th path of the i-th channel.

B. Problem Formulation for the Generic System Model of K RISs

Assuming that all the RISs have equal number of reflecting elements (i.e., $P_1 = P_2 \ldots = P_K = P$), let the d-th time-domain delay tap of the cascaded channel between the UE_n and BS be $\mathbf{V}_{d,n} \in \mathbb{C}^{M \times P_1}$. That is,

$$\mathbf{V}_{d,n} = \mathbf{H}_{d,n}(n_{sc}) \left(\prod_{k=2,\ldots,K} \mathbf{R}_{d,j}(n_{sc}) \right) \times diag(\mathbf{g}_{d,n}(n_{sc})). \tag{5}$$

Therefore, by discrete fourier transformation (DFT), the frequency-domain cascaded channel of the n_{sc}-th subcarrier from UE_n to BS is given by

$$\mathbf{V}_n[n_{sc}] = \sum_{d=0}^{D} \mathbf{V}_{d,n} e^{-j2\pi\frac{n_{sc}}{N_{sc}}d}, \tag{6}$$

where

$$D = D_G D_H \left(\prod_{j=1,2,\ldots,K-1} D_{R_j} \right). \tag{7}$$

As in [30], the cascaded channel in (6) can be decomposed by using the virtual channel representation. That is,

$$\mathbf{V}_{n,q}[n_{sc}] = \mathbf{A}_{BS} \mathbf{\Delta}_{n,q}[n_{sc}] \mathbf{A}^H_{RIS_1}. \tag{8}$$

The $\mathbf{A}_{BS}$ and $\mathbf{A}_{RIS_1}$ in (8) represent the $M \times M$ and $P_1 \times P_1$ complex dictionary matrices of the BS's angle of arrival (AoA) and RIS_1's angle of departure (AoD) respectively. The symbol $\mathbf{\Delta}_n[n_{sc}]$ represents the $M \times P_1$ sparse matrix, and its nonzero entries are the BS's AoAs and RIS_1's AoDs.

The received signal at the BS for the n-th user at the q-th time slot where $q \in \{1, 2, \ldots, Q\}$, is given by

$$\mathbf{y}_{n,q}[n_{sc}] = \mathbf{V}_{n,q}[n_{sc}]\mathbf{\Theta}_{n,q}s_{n,q}[n_{sc}] + \mathbf{w}_{n,q}[n_{sc}]. \tag{9}$$

The $\mathbf{\Theta}_{n,q}$ is the combined reflecting coefficients of all the RISs at the q-th time slot for the n-th user given by

$$\mathbf{\Theta}_{n,q} = \prod_{k=1,2,\ldots,K} diag(\mathbf{\Theta}_{n,q,k}), \tag{10}$$

where $\mathbf{\Theta}_{n,q,k} = [\theta_{n,q,k,1}, \theta_{n,q,k,2}, \ldots, \theta_{n,q,k,P}]$ is an $P \times 1$ reflecting vector for the n-th user at the q-th time slot for the k-th RIS with $\theta_{n,q,k,p}$ representing the reflecting coefficient for the n-th user at the p-th element of RIS_k where $p \in \{1, 2, \ldots, P\}$.

The $\mathbf{w}_{n,q}[n_{sc}] \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_M)$ and $s_{n,q}[n_{sc}]$ in (9) are the noise vector and transmitted symbol at q-th time slot of the n-th UE respectively, where σ^2 is the noise power at the n_{sc}-th subcarrier and $\mathbf{I}_M$ is an $M \times M$ identity matrix.

From (8), (9) can therefore be represented as

$$\mathbf{y}_{n,q}[n_{sc}] = \mathbf{A}_{BS}\mathbf{\Delta}_{n,q}[n_{sc}]\mathbf{A}^H_{RIS_1}\mathbf{\Theta}_{n,q}s_{n,q}[n_{sc}] + \mathbf{w}_{n,q}[n_{sc}]. \tag{11}$$

Moreover, (11) can be reduced to

$$\tilde{\mathbf{y}}_{n,q}[n_{sc}] = \tilde{\mathbf{\Theta}}_{n,q}\mathbf{\Delta}_{n,q}[n_{sc}] + \tilde{\mathbf{w}}_{n,q}[n_{sc}], \tag{12}$$

where $\tilde{\mathbf{y}}_{n,q}[n_{sc}] = \mathbf{A}^H_{BS}\mathbf{y}_{n,q}[n_{sc}] \in \mathbb{C}^{M\times 1}$, $\tilde{\mathbf{w}}_{n,q}[n_{sc}] = \mathbf{A}^H_{BS}\mathbf{w}_{n,q}[n_{sc}] \in \mathbb{C}^{M\times 1}$ and $\tilde{\mathbf{\Theta}} = \mathbf{A}^H_{RIS_1}\mathbf{\Theta}_{n,q}s_{n,q}[n_{sc}] \in \mathbb{C}^{p\times 1}$ are the effective measurement, noise and sensing vectors for the n_{sc}-th subcarrier respectively.

III. The Two-Timescale, Sparsity Properties of the Channel and the Proposed DnCNN-GRU Architecture

In this section, the two time-scale and sparsity properties of the channel based on the system model discussed in Section II are pitched as adopted for the purpose of reducing the pilot overhead. Thereafter, the proposed deep learning-based CE architecture is presented.

A. The Two Time-Scale and Sparsity Properties of the Channel

In an RIS-assisted wireless channel, significantly high pilot overhead is required to estimate the channel due to huge number of channel coefficients to be estimated than in traditional wireless communication systems without RISs. The mobility of the transceivers further increases the pilot overhead [33]. As a result, the complexity and capacity of the system are negatively affected. It is therefore important to make efforts in reducing the pilot overheads in multi-RIS-assisted channels.

Since the proposed channel model in Section II inherits a two-timescale channel property [30]; that is, all RISs along with the BS are located at fixed positions and only the UEs are mobile, the two-timescale scheme is proposed here to significantly reduce the pilot overhead of the multi-RIS-assisted channel by taking advantage of the two-timescale channel property. The proposed channel model is assumed to have two coherence times T_{UE} and T_{RB} which are the small-timescale and large-timescale coherence times respectively. That is, since channels $\{\mathbf{g}_n\}_{n=1}^N$ are fast time-varying due to the mobility of the UEs, they are assumed to have a small-timescale coherence time T_{UE}, and all the channels from RIS_1 to the BS are assumed to be quasi-static and adopt the large-timescale coherence time T_{RB} due to the fixed positions of the RISs and the BS.

The two-timescale channel estimation scheme can be executed in two stages. In the first stage, the cascaded channel $\mathbf{\Delta}_n[n_{sc}]$ is estimated to estimate the channel amplitudes of the cascaded channel from RIS_1 to the BS since all the UEs share the RIS_1-BS cascaded channel. This is done for the large-timescale coherence time T_{RB}, assuming that the channel $\mathbf{g}_n$ is also quasi-static. That is, since the estimation of the channel $\mathbf{g}_n$ does not affect the estimation of the channels from RIS_1 to the BS for the first stage, the channel $\mathbf{g}_n$ can be initialised to have constant channel coefficients and it will not affect the estimation of the quasi-static channel RIS_1-BS. Thus, the channel $\mathbf{g}_n$ is assumed to be known and quasi-static in this first stage. Therefore, the pilots are transmitted from the UEs to the BS to estimate RIS_1-BS cascaded channel. This process significantly reduces the pilot overhead compared to the traditional channel estimation techniques because the quasi-static cascaded channel will be estimated only once during the large-timescale period.

In the second stage, the purpose is to estimate the fast time-varying channels $\{\mathbf{g}_n\}_{n=1}^N$. Using the estimated quasi-static channels from the first stage, the cascaded channel $\mathbf{\Delta}_n[n_{sc}]$ is once more estimated but with the unknown time-varying channel $\mathbf{g}_n$. This process is performed during the small-timescale channel stage for the small-timescale coherence time T_{UE}. The first stage solved the channel estimation ambiguity in the second stage since only channel $\mathbf{g}_n$ is unknown, reducing the CE complexity. Thus, fewer pilots are used for the estimation of the cascaded channel $\mathbf{\Delta}_n[n_{sc}]$ than in traditional channel estimation schemes where both the quasi-static and fast time-varying channels are estimated with the same number of iterations.

The estimation of the uplink cascaded channel between the UEs and the BS is possible mainly due to the channel sparsity. The CE in this paper is performed by taking advantage of the double-structured sparsity (adopted from [34]) of the proposed multi-RIS-assisted channel as demonstrated by the channel matrices in Fig. 2. The rows of $\mathbf{\Delta}_{i,q}$ and $\mathbf{\Delta}_{j,q}$ in Fig. 2 represent the paths for the cascaded channel RIS_1-BS while the columns represent the paths for the time varying-channel $\mathbf{g}_n$. Since all UEs completely share all the cascaded channel RIS_1-BS paths and some of the $\mathbf{G}$ paths [35], the sparse channels $\{\mathbf{\Delta}_{n,q}\}_{n=1}^N$ exhibit the double-structured sparsity. This means that the cascaded channels of $\{UE_n\}_{n=1}^N$ can be jointly estimated taking advantage of the double-structured sparsity property of the channels.

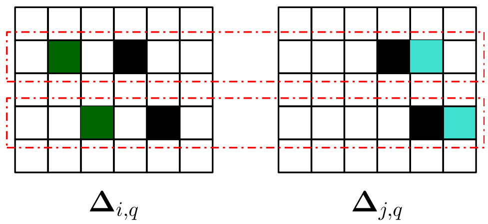


Fig. 2. The double-structured sparsity of the cascaded channels for the UE$_i$ and UE$_j$ for the q-th time slot.

B. The Proposed DnCNN-GRU Architecture

In this subsection, a deep-learning based algorithm is proposed to estimate the cascaded channel from the UEs to the BS through the multi-RISs. The proposed scheme in this section is based on the DnCNN scheme proposed in [16] where the double-structured DnCNN architecture was used for the support detection in an uplink single-RIS-assisted channel. First, since the UEs share quasi-static channel RIS-BS, the DnCNN was dedicated for estimating the RIS-BS channel. Furthermore, since the UEs do not share channel coefficients in the UEs-RIS channel, the second DnCNN was then dedicated to estimate the independent channels per UE taking advantage of the partial sparsity of the UEs-RIS_1. Basically, in [16], only the DnCNN was used to estimate the cascaded channels from the stationery UEs to the BS. However, if the UEs become mobile, the proposed scheme will not perform well.

Therefore, in this paper, with the assumption that the UEs are mobile and there are multiple RISs, the new multi-RIS-assisted CE scheme is proposed, utilising the DnCNN structure proposed in [16] accordingly. Unlike in [16], only a single-structured DnCNN architecture is proposed where the quasi-static cascaded channel from RIS_1 all the way to the BS is estimated using the DnCNN. To estimate the time-varying channel between the UEs and the RIS_1, the GRU scheme is proposed instead of another DnCNN as in [16] due to its robustness to fast time-varying channels.

Additionally, our proposed methodology allows the cascaded CE from RIS_1 to the BS without the stress of estimating the uplink channels between the RIS_1 and the BS individually. That is, the double-RIS-assisted CE of our proposed scheme can be simply extended to multi-RIS-assisted CE. This is possible by our assumption that the number of RIS elements for all RISs are equal, unlike the double-RIS schemes in [30] and [31] where the channels between the RIS_1 and the BS are estimated individually. For example, in [30] and [31], the double-RIS-assisted system was considered to estimate the channels between the UEs and RIS_1 (UEs-RIS_1), between the RIS_1 and second RIS (RIS_1-RIS_2), and between RIS_2 and the base station (RIS_2-BS). However, both works assumed that the UEs and the BS are in LoS with all two RISs. In this case, to estimate the UEs-RIS_1 channel, the RIS_2 is switched off and the pilots are reflected through the active RIS_1 only. This basically becomes the CE problem for a simpler single-RIS system. Similarly, to estimate RIS_2-BS channel, RIS_1 and RIS_2 are switched off and on respectively. Finally, it comes simpler to estimate the RIS_1-RIS_2 channel without the ambiguity of channels UEs-RIS_1 and RIS_2-BS. This methodology cannot be extended for scenarios with more than three RISs because after the estimation of UEs-RIS_1 and RIS_K-BS channels, there will still be more than two channels to be estimated between RISs (i.e., channels RIS_1-RIS_2, RIS_2-RIS_3, $\ldots$, RIS_{K-1}-RIS_K). The proposed architectures are explained below.

1) The DnCNN Architecture for Estimating the Quasi-Static Cascaded Channel: Since the received signal at the BS is a noisy version of the transmitted signal, the CNN is proposed to denoise the received signal for the recovery of the transmitted signal. There are several reasons for using the CNN for CE. First, the CNN is capable of reducing the dimensions of the input data for processing. For instance, since the multi-RIS assisted system has very high dimensions, the CNN can detect smaller, meaningful features of the cascaded $\mathbf{\Delta}_n$ for processing to reduce the complexity of CE [36]. This capability of the CNN consequently improves the statistical efficiency and reduces memory requirement of the system. Furthermore, the CNNs have advanced regularisation functions like rectified linear unit (ReLU), BN, and residual learning, which significantly increase the speed of the training process and improve the accuracy of denoising [37].

The proposed DnCNN architecture is shown in Fig. 3 and has L_C convolutional layers with each layer comprising of c_{l_c} different $D_{l_c,x} \times D_{l_c,y} \times D_{l_c,z}$ filters, where $l_c \in \{1, 2, \ldots, L_C\}$ is the l_c-th convolutional layer; the $D_{l_c,x}$, $D_{l_c,y}$ and $D_{l_c,z}$ are the dimensions of the filters in the x, y and z directions respectively [38]. The first layer (i.e., layer $l_c = 1$) is followed by the ReLU function while the succeeding convolutional layers except the last layer are each followed by both the BN and ReLU functions consecutively. The last layer (i.e., the L_C-th layer) constructs the output signal of the DnCNN by a single separate filter.

Since the users share the paths in the cascaded channel between RIS_1 and the BS, the DnCNN uses the $M \times 1$ correlation vector

$$\mathbf{c}_r = \sum_{n_{sc}=1}^{N_{sc}} \sum_{n=1}^{N} \mathbf{c}_n[n_{sc}]^H \tag{13}$$

as an input for path detection of all the UEs, where $\mathbf{c}_n[n_{sc}] = |\mathbf{s}_n \tilde{\mathbf{Y}}_n[n_{sc}]| \in \mathbb{C}^{1\times M}$, $\mathbf{s}_n = [s_{n,1}, s_{n,2}, \ldots, s_{n,Q}]^H$ and $\tilde{\mathbf{Y}}_n[n_{sc}] \in \mathbb{C}^{Q\times M}$ is the collection of the received effective signals for Q time slots of the n_{sc}-th subcarrier. Since the CNNs are constructed to specially process data that has a grid-like topology [36], for the correlation vector $\mathbf{c}_r$ to be used as an input to the DnCNN, it must therefore be formulated as a matrix $\mathbf{C}_r \in \mathbb{C}^{M_x \times M_y}$:

$$\mathbf{C}_r = vec2math(\mathbf{c}_r, [M_x, M_y]). \tag{14}$$

However, as shown in Fig. 3, the output of the DnCNN is the residual noise of the input. To find the vector associated with the cascaded channel, the residual noise can be subtracted from the input of the DnCNN as

$$\mathbf{g}_r = \mathbf{c}_r - \mathbf{w}_{RN} \in \mathbb{C}^{1\times M} \tag{15}$$

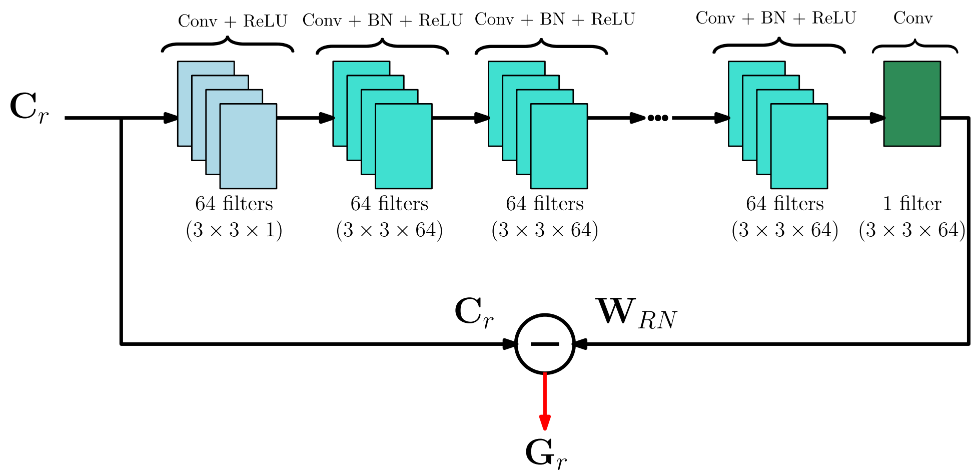


Fig. 3. The proposed DnCNN architecture.

where $\mathbf{w}_{RN} \in \mathbb{C}^{1\times M}$ is the residual noise vector and

$$\mathbf{g}_r = \sum_{n_{sc}=1}^{N_{sc}} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{p=1}^{P} |[\mathbf{\Delta}_n[n_{sc}]]_{:,n}|. \tag{16}$$

The vector $\mathbf{g}_r$ contains all the accumulated cascaded channel amplitudes coming from all the UEs. Similar to (5), the matrix version of $\mathbf{g}_r$ and $\mathbf{w}_{RN}$ are denoted by $\mathbf{G}_r$ and $\mathbf{W}_{RN} = \mathcal{W}(\mathbf{C}_r)$ respectively. Generally, the DnCNN is learning a mapping function $\mathcal{F}(\mathbf{C}_r) = \mathbf{G}_r$ for the estimation of the clean received signal from a noisy signal $\mathbf{C}_r$.

Therefore, the trainable parameter $\mathbf{\Theta}_n$ can be learned by a loss function

$$\ell_1(\mathbf{\Theta}_n) = \frac{1}{2I} \sum_{i=1}^{I} ||\mathcal{W}(\mathbf{C}_{r,i};\mathbf{\Theta}_n) - (\mathbf{C}_{r,i} - \mathbf{G}_{r,i})||_F^2 \tag{17}$$

where $\{\mathbf{C}_{r,i}, \mathbf{G}_{r,i}\}_{i=1}^{I}$ are the I training patch pairs.

In this DnCNN architecture, the BN and residual learning benefit from each other. As proven in [37], the residual learning without the BN performs drastically worser than the residual learning with BN. In the case of residual learning with BN, the BN improves the performance of the residual learning of the CNN by alleviating the problem of internal covariate shift. On the other hand, using BN in CNNs without residual learning results in relatively poor performance than the BN with residual learning. This is because the intensity of the input and the convolutional feature have high correlation with the neighbored ones and in each training mini-batch, there is a dependency of the distribution of the layer inputs. Incorporating the residual learning in this case produces a Gaussian-like distributions of the inputs of each layer which are less correlated.

2) The GRU Architecture for Estimating the Time-Varying Channels $\{\mathbf{g}_n\}_{n=1}^{N}$: Unlike in [16] where the DnCNN was used again to estimate the $\{\mathbf{g}_n\}_{n=1}^{N}$ channels, a GRU architecture is proposed as shown in Fig. 4a, where L_{GRU} is the total number of GRU layers.

The GRU is a variant of the Recurrent Neural Network (RNN) designed to mainly solve the vanishing gradient problem in RNNs. Unlike the DnCNN, the GRU architecture is able to handle the $\{UE_n\}_{n=1}^{N}$-RIS_1 time-varying channels as it is effective in capturing time series data dependencies [39]. Compared with the LSTM architecture which is also an RNN variant, the GRU has fewer gates and consequently, it has fewer parameters to learn and requires lesser memory [40]. Let $l_t \in \{1, 2, \ldots, L_T\}$ be the l_t-th cascaded path between RIS_1 and the BS where L_T is the total cascaded number of paths between RIS_1 and the BS. To estimate the $\mathbf{g}_n$ channel amplitudes, the proposed GRU algorithm takes a $P \times 1$ correlation vector

$$\mathbf{c}_{n,l_t,q} = \sum_{n_{sc}=1}^{N_{sc}} |\tilde{\mathbf{\Theta}}_{n,q}[\tilde{\mathbf{Y}}[n_{sc}]]_{:,\hat{\Omega}_r(l_t)}| \tag{18}$$

as an input. This input depends on the vector $\hat{\mathbf{g}}_r$ detected by the DnCNN where $\hat{\Omega}_r$ is a set containing the nonzero entries of $\hat{\mathbf{g}}_r$. Within $\hat{\Omega}_r$, $\hat{\Omega}_r(l_t)$ is the amplitude of the l_t-th path. The output of the GRU algorithm is the residual noise vector $\mathbf{w}_{n,l_t,q} \in \mathbb{C}^{P\times 1}$ where the vector corresponding to the channel amplitudes of the channel $\mathbf{g}_n$ is given by

$$\begin{aligned}&\mathbf{g}_{n,l_t,q}\\ &= \mathbf{c}_{n,l_t,q} - \mathbf{w}_{n,l_t,q} = \sum_{n_{sc}=1}^{N_{sc}} |\mathbf{\Delta}_{n,q}[n_{sc}]_{:,\hat{\Omega}_r(l_t)}| \in \mathbb{C}^{P\times 1}.\end{aligned} \tag{19}$$

The learnable parameters of the GRU are learned through the loss function

$$\begin{aligned}&\ell_2(\mathbf{\Theta}_n)\\ &= \frac{1}{2I} \sum_{i=1}^{I} ||\mathcal{W}(\mathbf{c}_{n,l_t,q,i};\mathbf{\Theta}_n) - (\mathbf{c}_{n,l_t,q,i} - \mathbf{g}_{n,l_t,q,i})||_F^2\end{aligned} \tag{20}$$

where $\{\mathbf{c}_{n,l_t,q,i}, \mathbf{g}_{n,l_t,q,i}\}_{i=1}^{I}$ are the I training patch pairs.

The internal structure of a single GRU is shown in Fig. 4b. Unlike the LSTM cell which has the forget, input and output gates, the GRU cell has only two gates which are the update and reset gates. The GRU's update gate is the combination of the LSTM's forget and input gates [41]. This combination significantly reduces the computational requirements of the GRU. In Fig. 4b, the update gate is represented by z_p and it is responsible for regulating the amount of information that

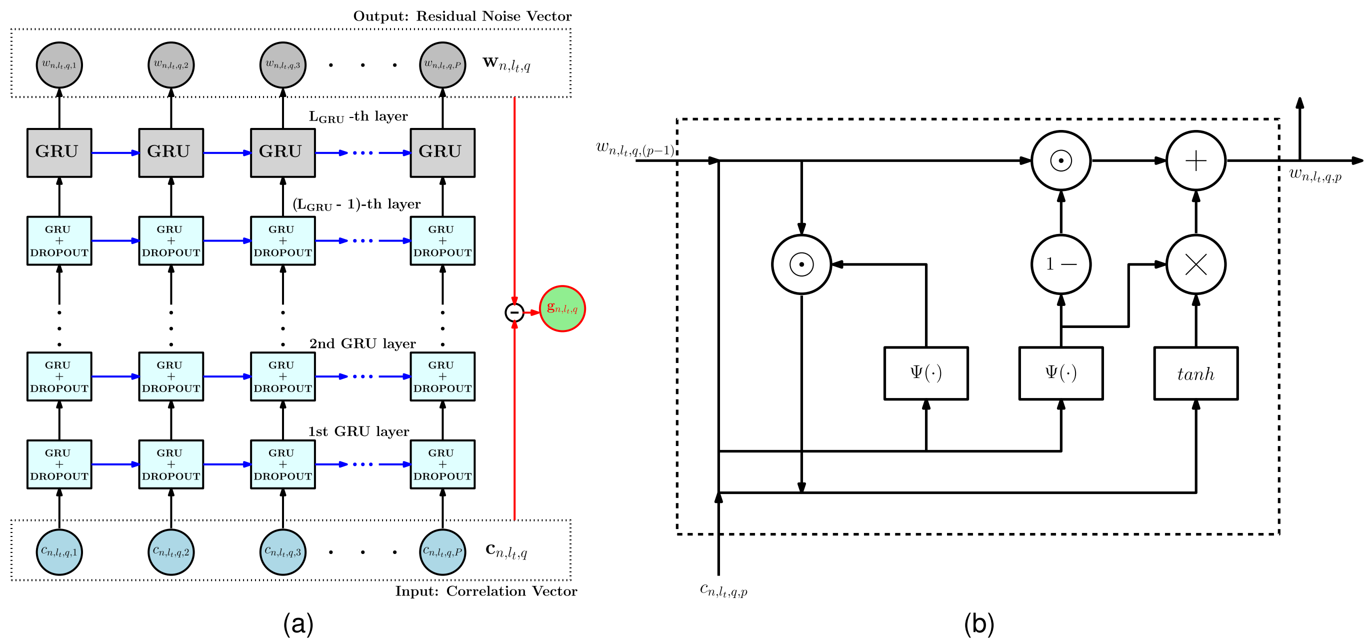


Fig. 4. The proposed GRU architecture: (a) The overall GRU architecture. (b) The internal structure of a single GRU.

can be stored between the current GRU cell and the previous GRU cells. On the other hand, the reset gate r_p regulates the amount of information of the previous GRU cells that should be ignored.

The output of the GRU cell is represented by

$$w_{n,l_t,q,p} = (1 - z_p) \odot w_{n,l_t,q,(p-1)} + (z_p \odot \hat{w}_{n,l_t,q,p}), \quad (21)$$

$$\hat{w}_{n,l_t,q,p} = tanh(\varrho_w c_{n,l_t,q,p} + U_w(r_p \odot w_{n,l_t,q,(p-1)}) + b_w). \quad (22)$$

Furthermore, the update and reset gates are computed as

$$z_p = \Psi(\varrho_z c_{n,l_t,q,p} + U_z w_{n,l_t,q,(p-1)} + b_z). \quad (23)$$

and

$$r_p = \Psi(\varrho_r c_{n,l_t,q,p} + U_r w_{n,l_t,q,(p-1)} + b_r) \quad (24)$$

respectively. In (22), (23) and (24), the symbols $\varrho_w, \varrho_z, \varrho_r, U_w, U_z, U_r, b_w, b_z$ and b_r are learnable parameters by the loss function in (20).

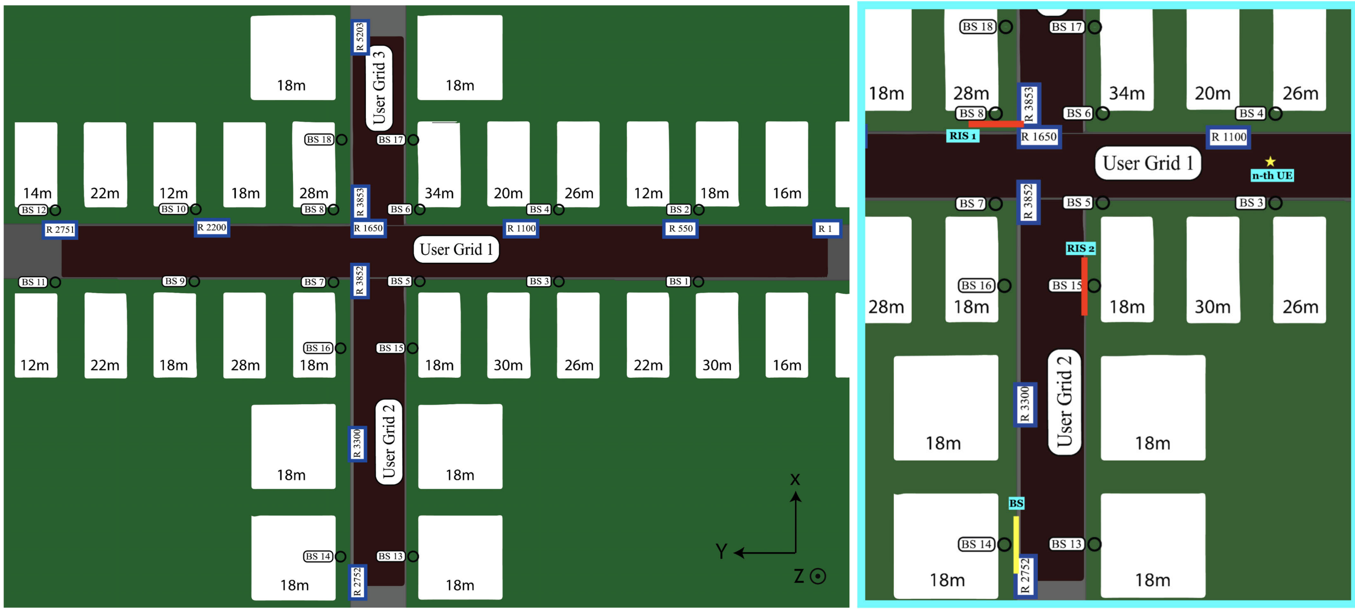
Algorithm 1 shows the complete CE process for the proposed scheme. In the initialisation stage, the $\tilde{\mathbf{Y}}_n[n_{sc}]$ data is generated by (12) for all q's, and specifying the number of paths per channel. Line 3 to 9 is the procedure for training the DnCNN algorithm. The final output is the estimation $\hat{\Omega}_r$. Line 10 to 14 shows how the GRU algorithm is trained to produce the estimation $\hat{\Omega}_{n,l_t,q}$, which is the set of nonzero entries of $\hat{\mathbf{g}}_{n,l_t,q}$. Finally, the estimated cascaded channel is reconstructed from line 15 to 27 with respect to $\hat{\Omega}_r$ and $\hat{\Omega}_{n,l_t,q}$ $\forall(n, n_{sc}, q)$.

Algorithm 1 The Proposed DnCNN-GRU

1: **Initialisation:** $\tilde{\mathbf{Y}}_n[n_{sc}]\ \forall(n_{sc}, n)$, $\boldsymbol{\Theta}_{n,q}\ \forall(n, q)$, L_G, L_H, $L_{R_j}\ \forall(R_j)$, $\hat{\Omega}_r$, $\hat{\Omega}_{n,l_t,q}\ \forall(n, n_{sc}, q, l_t)$
2: $\boldsymbol{\Delta}_{n,q}[n_{sc}] \leftarrow \mathbf{0}_{M\times P}\ \forall(n_{sc}, n, q)$
3: **procedure** ESTIMATE Ω_r
4: $\mathbf{C}_r \leftarrow vec2math(\mathbf{c}_r, [M_x, M_y])$.
5: $\hat{\mathbf{G}}_r \xleftarrow[\text{DnCNN}]{\text{Online}} \mathbf{C}_r$
6: $\hat{\mathbf{g}}_r \leftarrow vec(\hat{\mathbf{G}}_r)$
7: $\hat{\Omega}_r \leftarrow$ INDEXSORTDESCEND $(\hat{\mathbf{g}}_r, L_T)$
8: **Return** $\hat{\Omega}_r$
9: **end procedure**
10: **procedure** ESTIMATE $\Omega_{n,l_t,q}$
11: $\hat{\mathbf{g}}_{n,l_t,q} \xleftarrow[\text{GRU}]{\text{Online}} \mathbf{c}_{n,l_t,q}$
12: $\hat{\Omega}_{n,l_t,q} \leftarrow$ INDEXSORTDESCEND $(\hat{\mathbf{g}}_{n,l_t,q}, L_G)$
13: **Return** $\hat{\Omega}_{n,l_t,q}$
14: **end procedure**
15: **procedure** RECONSTRUCT THE CASCADED CHANNEL $\mathbf{V}_{n,q}[n_{sc}]$
16: **for** $q = 1 : Q$ **do**
17: **for** $n = 1 : N$ **do**
18: **for** $n_{sc} = 1 : N_{sc}$ **do**
19: **for** $l_t = 1 : L_T$ **do**
20: $\hat{\boldsymbol{\Delta}}_{n,q}[n_{sc}]_{(\hat{\Omega}_r, \hat{\Omega}_{n,l_t,q})}$
21: **end for**
22: $\hat{\mathbf{V}}_{n,q}[n_{sc}] = \mathbf{A}_{BS}\hat{\boldsymbol{\Delta}}_{n,q}[n_{sc}]\mathbf{A}^H_{RIS_1}$
23: **end for**
24: **end for**
25: **end for**
26: **return** $\hat{\mathbf{V}}_{n,q}[n_{sc}]\ \forall(n, n_{sc}, q)$
27: **end procedure**

IV. Simulation Results

In this section, the process of data generation for training the proposed DnCNN-GRU algorithm is explained. Furthermore, the simulation results of the proposed algorithm are presented. Finally, the comparison between the proposed DnCNN-GRU algorithm and other state-of-the-art channel estimators is performed. For every simulation result, the following are the



(a) (b)

Fig. 5. Simulation scenarios: (a) DeepMIMO "O1" scenario [24]. (b) The Double-RIS-assisted channel scenario.

parameters values settings, $N = 16$, $K = 2$, $M = 64$, $P = 256$, $N_{sc} = 2048$ and $B = 500$ MHz.

A. Simulation Scenario and Training Data Generation

In this paper, the DeepMIMO dataset [42] is utilised for training the proposed algorithms. The dataset is specially generated for training the machine learning based algorithms in mmWave massive-MIMO scenarios, thereby reducing the effort of generating large dataset for training the proposed machine learning algorithms. To emulate the proposed system model in Section II, the outdoor scenario in [42] called the "O1_60" scenario is adopted as shown in Fig. 5a. It is a top view of a two-intersected-street outdoor scenario with 18 BSs (i.e., BS1 to BS18), buildings of various heights and $1,184,923$ UEs distributed in rows (R1 to R5203) in an urban area operating at 60GHz. The white rectangles and squares in Fig. 5 represent the top view of buildings, and the numbers (i.e., 14m, 22m, etc.) inside them are their heights.

However, in this paper, the simulation results are generated for a double-RIS-assisted channel scenario as an example for the multi-RIS-assisted channels and therefore, only 16 UEs, 2 RISs and 1 BS are considered in Fig. 5a. The two RISs and the BS are indicated with red and yellow bars respectively in Fig. 5b. The n-th UE is indicated with a yellow star and it is assumed that all the 16 UEs' initial positions are in row 1000 (i.e., R1000) and they move parallel in the negative y-axis direction with different velocities. To communicate with the BS, all the UEs transmit information through RIS_1 and RIS_2 since it is clear from Fig. 5b that there is no LoS between the UEs and the BS. Similarly, the UEs can only communicate with RIS_2 through RIS_1.

The training data is therefore generated using the Wireless InSite ray-tracing simulator [42]. For training and validation of the proposed algorithms, 20 000 input-output dataset pairs were generated, where the training and validation datasets are 70% and 30% of the total generated dataset respectively. Similar procedure can be followed to generate training data for triple-RIS-assisted channel scenario and above.

B. DnCNN Convergence

In this subsection, the convergence of the proposed DnCNN channel estimator is investigated in terms of the learning rates and the number of convolutional layers that are required to accurately estimate the cascaded channel RIS_1-RIS_2-BS for the double-RIS-assisted channel in Fig. 5b. To identify the best learning rate for the proposed DnCNN, the loss function progression versus the number of training iterations is investigated, keeping the number of convolutional layers constant ($L_C = 3$). Once the best learning rate is identified, it is chosen and thereafter, the number of convolutional layers of the DnCNN is investigated which provide the optimal performance for the chosen learning rate.

Fig. 6 shows the convergence of the proposed DnCNN algorithm in terms of various learning rates and it is evident that the proposed DnCNN algorithm with 1×10^{-6} learning rate converges faster than the lower learning rates. Therefore, the 1×10^{-6} learning rate is chosen to train the DnCNN for double-RIS-assisted system. The higher learning rates than 1×10^{-6} are not considered in this case because they produce very high losses. For the 1×10^{-6} learning rate, the convergence of the proposed DnCNN algorithm is then investigated in terms of the number of convolutional layers, and the DnCNN with $L_C = 3$ is observed to converge faster and achieves lowest loss compared to DnCNNs with higher number of layers as shown in Fig. 7. Therefore, the DnCNN with $L_C = 3$ trained with a learning rate of 1×10^{-6} achieves the best CE for double-RIS-assisted channel. To prevent overfitting, the training of the DnCNN

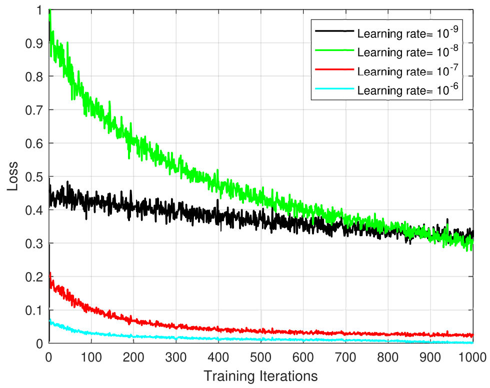


Fig. 6. The proposed DnCNN convergence in terms of training loss versus number of training iterations for various learning rates.

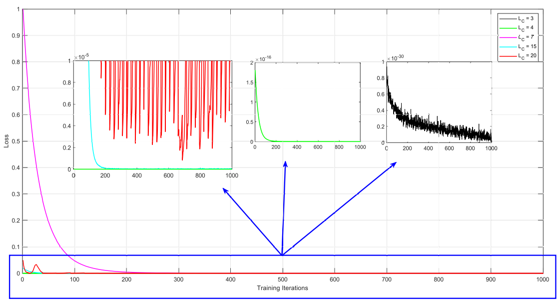


Fig. 7. The proposed DnCNN convergence (for 1×10^{-6} learning rate) in terms of training loss versus number of training iterations for various number of convolutional layers.

was terminated when the validation loss could not improve for 10 consecutive training iterations. For this, the training was terminated at the 100-th epoch for the training patch size of 256.

The optimisation of the proposed DnCNN scheme for triple-RIS-assisted channel and above follows the same procedure, where the learning rates and the number of layers are investigated for the optimal DnCNN.

C. GRU Convergence

Similar to the investigation of the proposed DnCNN convergence, the convergence of the proposed GRU is investigated. For that purpose, the 1×10^{-3} learning rates is chosen to train the GRUs of various layers. Fig. 8 shows that the GRU scheme for double-RIS-assisted channel with $L_{GRU} = 6$ converges faster and achieves the lowest loss. The GRU schemes with $L_{GRU} > 6$ produce unstable losses. This procedure can be applied also for triple-RIS-assisted channel and above to determine the optimal number of GRU layers. The convergence rate of the proposed GRU scheme depends largely on the number of UEs than the number of RISs deployed in the environment since the ambiguity of the quasi-static channels (i.e., $\mathbf{H}$ and $\{\mathbf{R}_k\}_{k=1}^{K}$) was eliminated by the DnCNN estimation.

D. The NMSE Comparison and Complexity Analysis

In this section, the benchmark is provided in terms of the NMSE metric for various Signal to Noise Ratios (SNRs) to show the superiority of the proposed DnCNN-GRU scheme.

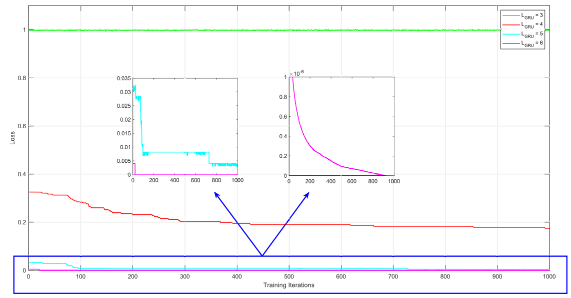


Fig. 8. The proposed GRU convergence (for 1×10^{-3} learning rate) in terms of training loss versus number of training iterations for various number of GRU layers.

The following benchmark CE schemes are compared with the proposed DnCNN-GRU scheme:

- **CNN-SC** [31]: The CNN-SC algorithm takes the LS channel estimate as preamble and formulates the CE as a denoising problem. It consists of 6 convolutional layers. The first convolutional layer has 128 filters of size 3×3. Each convolutional layer from the second to the fifth consists of the self-attention layer. To produce an output of the CNN-SC, the last convolutional layer takes the sum of the output of the first convolutional layer and the output of the fifth convolutional layer. All the convolutional layers have 128 filters of size 3×3 except the last convolutional layer which has 256 filters of size 3×3.
- **Oracle LS**: This provides the lowest bound of CE. It estimates the channel by assuming that the supports are perfectly known [35]. The Oracle LS in this case is used as a benchmark to see how far is the estimation of the proposed algorithm to the lowest possible estimate.
- **LMMSE channel estimator**: This LMMSE estimator is as derived in [31]. Although the LMMSE estimator was compared with the CNN-SC in [31], the LMMSE estimator in this paper is simulated for larger number of RIS elements (i.e., $P = 256$) and mobile UEs as opposed to 32 RIS elements and stationary UEs.

The NMSE is realised by

$$\text{NMSE} = \frac{1}{N}\sum_{n=1}^{N}\frac{\sum_{n_{sc}=1}^{N_{sc}}||\hat{\mathbf{V}}_n[n_{sc}] - \mathbf{V}_n[n_{sc}]||_F^2}{\sum_{n_{sc}=1}^{N_{sc}}||\mathbf{V}_n[n_{sc}]||_F^2} \tag{25}$$

which is averaged over many realisations per algorithm for fair comparison. The NMSE performance is investigated for different SNR levels from -10 dB to 15 dB as shown in Fig. 9.

The proposed DnCNN-GRU performs better than both the CNN-SC and LMMSE in terms of NMSE as shown in Fig. 9, and its performance is about 1 dB lower than the Oracle LS. This proves the superiority of the proposed algorithm.

Unlike the CNN-SC, the proposed DnCNN-GRU learns the residual mapping than the original unreferenced mapping [37]. The residual learning advantages the DnCNN-CNN largely over the CNN-SC since it is easier to learn the residual mapping, resulting with more accurate CE. Furthermore, the GRUs can handle the time-varying channel than the CNN-SC scheme. Additionally, in the training process, the BN of the DnCNN-GRU improves the CE performance by changing the

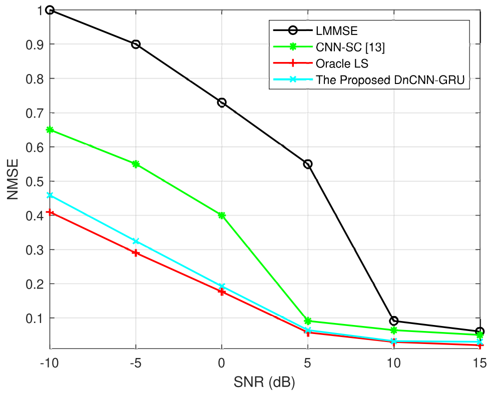


Fig. 9. NMSE with respect to SNR.

internal distribution of the training data, resulting in a faster and stable training.

The NMSE performance of the conventional LMMSE is worse than both the proposed DnCNN-GRU and CNN-SC mainly because the LMMSE finds it hard to learn the non-linear mapping of the double-RIS-assisted channel. It can only perform better if the spatial features of a channel are linear. This proves the inferiority of conventional methods in estimating the high dimensional RIS-assisted channels.

The overall online complexity of the proposed DnCNN-GRU scheme can be analysed from **Algorithm 1**. For the estimation of Ω_r from line 3 to 8, the complexity cost is of order $\mathcal{O}(NN_{sc}MQ)$ for computing the correlation matrix $\mathbf{C}_r$. The DnCNN processes the correlation matrix $\mathbf{C}_r$ to produce the estimation $\hat{\mathbf{G}}_r$. For the DnCNN with L_C number of convolutional layers, the complexity cost order is given by

$$\mathcal{O}\left(\sum_{l_c=1}^{L_C} D_{l_c,x} D_{l_c,y} D_{l_c,z} c_{l_c} MQ\right).$$

For the estimation $\hat{\Omega}_{n,l_t,q}$, the computational complexity cost order of the GRU scheme from line 10 to 13 of **Algorithm 1** is given by $\mathcal{O}\left(2PQ^2L_{GRU}\right)$, and the complexity of line 16 to 26 is given by $\mathcal{O}(NN_{sc}(L_TM + L_GP))$ for the q-th time slot.

Finally, the overall computational complexity of **Algorithm 1** is given by $\mathcal{O}(2PQ^2L_{GRU} + NN_{sc}(L_TM + L_GP + MQ) + \sum_{l_c=1}^{L_C} D_{l_c,x} D_{l_c,y} D_{l_c,z} c_{l_c} MQ)$.

V. Future Work and Conclusion

In this paper, the DnCNN-GRU is proposed to estimate the multi-RIS-assisted channel in a dynamic environment. Taking advantage of the two-timescale and double-structured sparsity properties of the multi-RIS-assisted channel, the uplink cascaded channel from the first RIS to the BS is estimated using the DnCNN in the first stage. The BN and residual learning of the DnCNN significantly increased both the speed of the DnCNN training process and the CE accuracy. In the second stage, the GRU scheme with its ability to capture time series data dependencies is used to estimate the time-varying uplink channel between the UEs and the first RIS. Finally, a benchmark of the proposed DnCNN-GRU scheme with the CNN-SC, LMMSE and Oracle LS is provided, and the proposed scheme achieved a performance of about 1 dB lower than the lower bound scheme (i.e., lower than the oracle LS). The results showed the superiority of the proposed scheme.

This work can be extended by investigating how the number of UEs, number of RIS elements and the machine-learning parameters like dropout rates of the GRU and the size the batch (how the BN of the DnCNN performs for various batch sizes) can affect the accuracy of the proposed scheme. Furthermore, it will be interesting to investigate the CE performance for the proposed scheme in a SIM-aided system.

References

[1] B. G. Kashyap et al., "A reconfigurable intelligent surface for 5G wireless communication applications," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting (APS/URSI)*, Dec. 2021, pp. 111–112, doi: 10.1109/APS/URSI47566.2021.9704068.

[2] S. Kisseleff, W. A. Martins, H. Al-Hraishawi, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for smart cities: Research challenges and opportunities," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1781–1797, 2020, doi: 10.1109/OJCOMS.2020.3036839.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013, doi: 10.1109/TCOMM.2013.020413.110848.

[4] X. Mo, L. Gui, X. Sang, and X. Diao, "Wideband sparse cascaded channel estimation for RIS-assisted wireless systems," in *Proc. 14th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Nov. 2022, pp. 495–500, doi: 10.1109/WCSP55476.2022.10039189.

[5] J. Wu, S. Kim, and B. Shim, "Parametric sparse channel estimation for RIS-assisted terahertz systems," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5503–5518, Jun. 2023, doi: 10.1109/TCOMM.2023.3285759.

[6] W. Yang, R. Liu, W. Cai, M. Li, and Q. Liu, "Channel estimation and pilot allocation for practical RIS-aided wideband OFDMA systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Rome, Italy, May 2023, pp. 4347–4352, doi: 10.1109/icc45041.2023.10279798.

[7] X. Su, R. He, P. Zhang, and B. Ai, "Generalized approximating message passing based channel estimation for RIS-aided THz communications with beam split," in *Proc. IEEE 99th Veh. Technol. Conf. (VTC-Spring)*, Singapore, Singapore, Jun. 2024, pp. 1–5, doi: 10.1109/vtc2024-spring62846.2024.10683198.

[8] Z. Huang, C. Liu, Y. Song, H. Wang, H. Zhou, and S. Shen, "Manifold optimization empowered two-timescale channel estimation for RIS-assisted systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 15034–15048, Oct. 2024, doi: 10.1109/TVT.2024.3405955.

[9] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, Nov. 2021, doi: 10.1109/TCOMM.2021.3072729.

[10] J. Sun, L. Zhao, C. Liu, and Y. Gao, "Sensing-based two-timescale channel estimation for RIS-assisted hybrid millimeter wave systems," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Hangzhou, China, Oct. 2022, pp. 1–6, doi: 10.1109/ISWCS56560.2022.9940404.

[11] K. Zhi et al., "Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI," *IEEE Trans. Inf. Theory*, vol. 69, no. 5, pp. 3001–3033, May 2023, doi: 10.1109/TIT.2022.3227538.

[12] L. Yu et al., "Channel estimation for STAR-RIS-Aided communications based on deep iterative networks," in *Proc. 3rd Int. Conf. Intell. Commun. Comput. (ICC)*, Nanchang, China, Nov. 2023, pp. 119–123, doi: 10.1109/icc59986.2023.10421001.

[13] W. Shen, Z. Qin, and A. Nallanathan, "Deep learning enabled channel estimation for RIS-aided wireless systems," in *Proc. GLOBECOM - IEEE Global Commun. Conf.*, Rio de Janeiro, Brazil, Dec. 2022, pp. 4226–4231, doi: 10.1109/GLOBECOM48099.2022.10000900.

[14] H. Feng, Y. Xu, and Y. Zhao, "Deep learning-based joint channel estimation and CSI feedback for RIS-assisted communications," *IEEE Commun. Lett.*, vol. 28, no. 8, pp. 1860–1864, Aug. 2024, doi: 10.1109/LCOMM.2024.3413729.

[15] K. Kim and C. S. Hong, "Pruned autoencoder based mmWave channel estimation in RIS-assisted wireless networks," in *Proc. 23rd Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Sep. 2022, pp. 1–4, doi: 10.23919/APNOMS56106.2022.9919942.

[16] A. Abdallah, A. Celik, M. M. Mansour, and A. M. Eltawil, "RIS-aided mmWave MIMO channel estimation using deep learning and compressive sensing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3503–3521, May 2023, doi: 10.1109/TWC.2022.3219140.

[17] H. Feng and Y. Zhao, "MmWave RIS-assisted SIMO channel estimation based on global attention residual network," *IEEE Wireless Commun. Lett.*, vol. 12, no. 7, pp. 1179–1183, Jul. 2023, doi: 10.1109/LWC.2023.3265648.

[18] Y. Jin, J. Zhang, X. Zhang, H. Xiao, B. Ai, and D. W. K. Ng, "Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11083–11088, Oct. 2021, doi: 10.1109/TVT.2021.3109937.

[19] C. Nguyen, T. M. Hoang, and A. A. Cheema, "Channel estimation using CNN-LSTM in RIS-NOMA assisted 6G network," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 1, pp. 43–60, 2023, doi: 10.1109/TMLCN.2023.3278232.

[20] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, J.-I. Baik, D.-S. Kim, and H. Song, "Deep learning-based improved cascaded channel estimation and signal detection for reconfigurable intelligent surfaces-assisted MU-MISO systems," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 3, pp. 1515–1527, Jan. 2023, doi: 10.1109/TGCN.2023.3237132.

[21] M. A. Haider, Y. D. Zhang, and E. Aboutanios, "Channel estimation and prediction in wireless communications assisted by semi-passive RIS," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 8601–8605, doi: 10.1109/icassp48485.2024.10446555.

[22] C. Xu et al., "Channel estimation for reconfigurable intelligent surface assisted high-mobility wireless systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 718–734, Jan. 2023, doi: 10.1109/TVT.2022.3203818.

[23] C. Xu et al., "Reconfigurable intelligent surface assisted multi-carrier wireless systems for doubly selective high-mobility Ricean channels," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4023–4041, Apr. 2022, doi: 10.1109/TVT.2022.3147859.

[24] W. Xu, J. An, Y. Xu, C. Huang, L. Gan, and C. Yuen, "Time-varying channel prediction for RIS-assisted MU-MISO networks via deep learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 4, pp. 1802–1815, Dec. 2022, doi: 10.1109/TCCN.2022.3188153.

[25] J. An et al., "Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2380–2396, Aug. 2023, doi: 10.1109/JSAC.2023.3288261.

[26] J. An et al., "Stacked intelligent metasurface-aided MIMO transceiver design," *IEEE Wireless Commun.*, vol. 31, no. 4, pp. 123–131, Aug. 2024, doi: 10.1109/MWC.013.2300259.

[27] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[28] C. Liu et al., "A programmable diffractive deep neural network based on a digital-coding metasurface array," *Nature Electron.*, vol. 5, no. 2, pp. 113–122, Feb. 2022.

[29] Y. Wang, W. Zhang, Y. Chen, C.-X. Wang, and J. Sun, "Novel multiple RIS-assisted communications for 6G networks," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1413–1417, Jun. 2022, doi: 10.1109/LCOMM.2022.3163587.

[30] S. Yang, W. Lyu, Y. Xiu, Z. Zhang, and C. Yuen, "Active 3D double-RIS-aided multi-user communications: Two-timescale-based separate channel estimation via Bayesian learning," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3605–3620, Jun. 2023, doi: 10.1109/TCOMM.2023.3265115.

[31] M. Liu, X. Li, B. Ning, C. Huang, S. Sun, and C. Yuen, "Deep learning-based channel estimation for double-RIS aided massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 70–74, Jan. 2023, doi: 10.1109/LWC.2022.3217294.

[32] S. Gherekhloo, K. Ardah, A. L. F. De Almeida, M. Maleki, and M. Haardt, "Nested PARAFAC tensor-based channel estimation method for double RIS-aided MIMO communication systems," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, Helsinki, Finland, Sep. 2023, pp. 1674–1678, doi: 10.23919/eusipco58844.2023.10290107.

[33] B. P. S. Sahoo, S. Narang, P. Jain, D. Kumar, and S. K. Vankayala, "Mobility-aware hybrid channel estimation scheme for RIS-aided millimeter-wave systems," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Bangalore, India, Jul. 2022, pp. 1–5, doi: 10.1109/CONECCT55679.2022.9865721.

[34] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021, doi: 10.1109/LCOMM.2021.3052787.

[35] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath Jr., "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018, doi: 10.1109/TWC.2018.2804943.

[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. Accessed: Nov. 15, 2023. [Online]. Available: https://www.deeplearningbook.org

[37] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017, doi: 10.1109/TIP.2017.2662206.

[38] A. Abdallah, A. Celik, M. M. Mansour, and A. M. Eltawil, "Deep learning-based frequency-selective channel estimation for hybrid mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3804–3821, Jun. 2022, doi: 10.1109/TWC.2021.3124202.

[39] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [review article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018, doi: 10.1109/MCI.2018.2840738.

[40] J.-H. Wang, H. He, K. Tamura, S. Kojima, J. Cha, and C.-J. Ahn, "Vehicle-to-Vehicle communication channel estimator based on gate recurrent unit," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Taipei, Taiwan, Jan. 2024, pp. 1–4, doi: 10.1109/iceic61013.2024.10457159.

[41] M. H. Essai Ali, M. L. Rabeh, S. Hekal, and A. N. Abbas, "Deep learning gated recurrent neural network-based channel state estimator for OFDM wireless communication systems," *IEEE Access*, vol. 10, pp. 69312–69322, 2022, doi: 10.1109/ACCESS.2022.3186323.

[42] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2019, pp. 1–8.

Thabang C. Rapudu received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from the University of Johannesburg, Johannesburg, South Africa, in 2021 and 2022, respectively. He is currently pursuing the Ph.D. degree in electrical and information engineering with the University of the Witwatersrand at Johannesburg, Johannesburg. His current research interests include signal processing, 5G and beyond 5G technologies, and the application of machine learning techniques for wireless communication systems.

Olutayo O. Oyerinde (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from the University of KwaZulu-Natal, Durban, South Africa, in 2011. Since 2013, he has been with the School of Electrical and Information Engineering, University of the Witwatersrand at Johannesburg, South Africa, where he is currently an Associate Professor. He is a National Research Foundation (NRF) C2-Rated Scientist. His current research interests include wireless communications with specific interests in 5G and beyond 5G technologies, and signal processing techniques for wireless communication systems. He is a Corporate Member of NSE. He is also an Editorial Member of the Editorial Board of *Wireless Communications and Control*. He is an Associate Editor of IEEE ACCESS, IEEE COMMUNICATIONS LETTERS, the *Journal of the Franklin Institute*, and the *Franklin Open*. He is a registered Professional Engineer (Pr.Eng.) with the Engineering Council of South Africa (ECSA) and a registered Engineer (R.Eng.) with COREN.