Cascaded and Separate Channel Estimation based on CNN for RIS-MIMO Systems

Wala'a Hussein
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Department of Chemical Engineering and Petroleum Refining, Basrah University for Oil and Gas, Iraq | Department of Computer Technology Engineering, Faculty of Engineering, Iraq University College, Iraq
walaahussein613@gmail.com (corresponding author)

Nor K. Noordin
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Wireless and Photonics Networks Research Center of Excellence (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Malaysia
nknordin@upm.edu.my (corresponding author)

Kamil Audah
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Department of Electronics Technologies, Southern Technical University, Iraq
thegenusnabster@gmail.com

Mod Fadlee B. A. Rasid
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Wireless and Photonics Networks Research Center of Excellence (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Malaysia
fadlee@upm.edu.my

Alyani Binti Ismail
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Malaysia | Wireless and Photonics Networks Research Center of Excellence (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Malaysia
alyani@upm.edu.my

Aymen Flah
MEU Research Unit, Middle East University, Jordan | Applied Science Research Center, Applied Science Private University, Jordan
flahaymening@yahoo.fr

Received: 15 April 2024 | Revised: 25 April 2024 | Accepted: 27 April 2024
Licensed under a CC-BY 4.0 license \ Copyright (c) by the authors \ DOI: https://doi.org/10.48084/etasr.7499

ABSTRACT

With the dramatic increase in mobile users and wireless devices accessing the network, the performance of 5G wireless communication systems is severely challenged. Reconfigurable Intelligent Surface (RIS) has received much attention as one of the promising technologies for 6G due to its ease of deployment, low power consumption, and low price. This study aims to improve accuracy, reliability, and the capacity to estimate channel characteristics between transmitter and receiver. However, this is practically challenging for the following reasons. Due to the lack of active components for baseband signal processing, low-cost passive RIS elements can only reflect incident signals but without the capability to transmit/receive pilot signals for channel estimation as active transceivers in conventional wireless communication systems. This
Emerging developments in the field of machine learning provide appealing substitutes for the RIS phase transition [17]. This study presents a new deep-learning architecture employing two interconnected neural networks. Each network is specifically trained to address the distinct channel characteristics encountered in two scenarios. Compared to traditional reinforcement learning approaches, the proposed double neural network exhibits exceptional performance, even in scenarios with a large number of RIS components. A 5G link level-based RIS-MIMO MATLAB 2023a simulator was utilized to model the time-frequency response of the channel as a 2D picture and Anaconda Spyder for training on the dataset.

II. SYSTEM AND CHANNEL MODELS

Figure 1 depicts the passive/hybrid RIS architecture in conjunction with MIMO communication systems based on a limited number of RF chains.

![Diagram of RIS architecture](image)

**Fig. 1.** The proposed RIS design uses a surface-distributed array of $K$ active/passive channel sensors. These sensors operate in two modes: (i) a channel sensing mode connected to the baseband for channel estimation and (ii) a reflection mode applying phase shifts to incident signals. The remaining RIS elements are passive reflectors without baseband connections.

In scenarios where direct communication between the source (BS) and the destination (UE) nodes is impeded by severe obstruction or is blocked, the channel efficiency from the $n^{th}$ antenna at the BS to the $k^{th}$ element of the RIS experiences Rician fading, as outlined below:

$$H_{R,K} = \frac{k_{k,n}}{\sqrt{k_{k,n}+1}} h_{k,n}^{(LOS)} + \frac{k_{k,n}}{k_{k,n}+1} h_{k,n}^{(NLOS)}$$  \hspace{1cm}(1)$$

where $k_{k,n}$ represents the Rician fact respective link, and $h_{k,n}^{(LOS)}$ and $h_{k,n}^{(NLOS)}$ are the components of the Rician fading channel for Line Of Sight (LOS) and Non Line Of Sight (NLOS). The expression for the LOS component is as follows:

$$h_{k,n}^{(LOS)} = \sqrt{\beta} e^{-j(k-1)\text{min}0} \frac{e^{-\frac{|d_{k,n}|^2}{2\sigma^2}}}{d_{k,n}}$$  \hspace{1cm}(2)$$

In scenarios where direct communication between the source (BS) and the destination (UE) nodes is impeded by severe obstruction or is blocked, the channel efficiency from the $n^{th}$ antenna at the BS to the $k^{th}$ element of the RIS experiences Rician fading, as outlined below:

where $k_{k,n}$ represents the Rician fact respective link, and $h_{k,n}^{(LOS)}$ and $h_{k,n}^{(NLOS)}$ are the components of the Rician fading channel for Line Of Sight (LOS) and Non Line Of Sight (NLOS). The expression for the LOS component is as follows:
In this equation, \( \theta \) denotes the angle of arrival at the RIS, and \( f_0 \) represents the path loss at the reference distance of 1 m. In the NLOS scenario:

\[
h_{k,n}^{\text{(NLOS)}} = \frac{1}{\sqrt{d_{k,n}}} \hat{h}_{k,n}
\]

where \( \hat{h}_{k,n} \) is the complex Gaussian small-scale fading with a unit variance, and zero mean is denoted by \( H_{k,R} \). Similarly, the channel connecting the \( R \)th RIS elements and the \( m \)th antenna at the UE is denoted as \( H_{k,R} \). The signal received at the destination is expressed as:

\[
Y = H_{T,R} \Theta H_{R,U} x + n
\]  

(3)

In this context, \( Y \) belongs to the \( \mathbb{C}^{M \times 1} \), \( H_{T,R} \in \mathbb{C}^{N \times K} \) is the channel matrix between UE and RIS, and \( H_{R,U} \in \mathbb{C}^{M \times N} \), is the channel matrix representing the communication link between the RIS and UE. Additionally, \( n \) belongs to the \( \mathbb{C}^{M \times 1} \) space and signifies the additive noise between the BS and RIS, and the phase shift matrix at the RIS is represented as \( \Theta \), which is given by:

\[
\Theta = \text{diag}(a_1e^{j\theta_1}, \ldots, a_Ke^{j\theta_K})
\]

(4)

In this context, \( a_k \) signifies the amplitude, and \( \theta_k \) represents the phase shift at the \( k \)th element. Although \( \theta_k \) is typically presumed to remain constant in numerous existing methodologies, it is also acknowledged that it can be subject to variation.

A. Cascaded Channel Estimation Scenario

Cascaded channel estimation aims to capture the combined influence of the BS-RIS and RIS-UE channels in a single step. When employing fully passive RIS without sensing devices, it becomes impractical to estimate the BS-RIS and RIS-UE channels separately. Instead, only the combined BS-RIS-UE channel can be estimated, typically at one endpoint, such as the UE, in the downlink communication system. For passive RIS, direct estimation is limited to the overall BS-RIS-UE channel. Separate channel estimates require additional capabilities at the RIS, typically at the UE, in the downlink communication system. During time slot \( t \), the BS transmits pilot signals, and the properties of the signal received by the UE can be described by:

\[
Y_{U,E}^{(t)} = \sum_{k=1}^{K} \sqrt{P_{RIS}} \frac{1}{\sqrt{d_{k,n}}} H_{T,R,k} \theta^{(t)} H_{R,U,k} x_k^{(t)} + n_{U,E}^{(t)}
\]

(5)

where \( x_k^{(t)} \in \mathbb{C}^{M_R \times 1} \) represents the transmitted signal of the pilot by the BS, \( \theta^{(t)} = \text{diag}(\theta^{(t)}_1, \ldots, \theta^{(t)}_K) \) is represented as a diagonal reflection matrix during time slot \( t \) at RIS, and the AWGN at the user can be represented as \( n_{U,E}^{(t)} \in \mathbb{C}^{M_R \times 1} \). Using the product properties of Khatri-Rao the following equation can be derived:

\[
\text{vec}(H_{T,R,k} \theta^{(t)} H_{R,U,k}) = H_{R,U,k}^{(t)} H_{T,R,k} \theta^{(t)}, \quad k = 1..K
\]

(6)

In this expression, \( H_{R,U,k}^{(t)} \in \mathbb{C}^{M_R M_U \times N} \) signifies the cascaded channel of user \( k \), \( \text{vec}(\cdot) \) denotes vectorization, and \( \text{O} \) denotes the Khatri-Rao product.

B. Separate Channel Estimation Scenario

Unlike the combined channel estimation approach, the separate channel estimation approach treats the transmitter-to-RIS (BS-RIS) and RIS-to-receiver (RIS-UE) channels as independent and estimates them individually. This method is particularly applicable in scenarios with semi-passive or hybrid RIS equipped with dedicated sensing devices (Ns). When such sensing devices are available, they can be used to estimate the individual channels between the BS or UE and the RIS using dedicated pilot signals sent by either the BS or UE. Channel estimation is conducted at the RIS. TDD systems can effectively leverage channel reciprocity to obtain the Channel State Information (CSI) from the RIS to the BS or UE. This is because the same channel is deployed for transmission in both directions. Applying this method in Frequency-Division Duplexing (FDD) systems is complex. FDD engages separate frequencies for transmission and reception, eliminating channel reciprocity. To reduce hardware costs in semi-passive RIS, a limited number of low-resolution Analog-to-Digital Converters (ADCs) might be employed. However, this necessitates active sensors on the RIS, capable of transmitting and receiving pilots for channel estimation in FDD systems, leading to increased power consumption. Specifically, the channels from the BS and UE to the Ns sensing devices are represented as follows:

\[
\mathbf{G}(H_{s,x}) \in \mathbb{C}^{N_s \times M_B} \quad \text{and} \quad \mathbf{H}(H_{R,U}) \in \mathbb{C}^{N_s \times N_U}
\]

Additionally, \( X_0 \in \mathbb{C}^{M_B \times T} \) and \( X_{ID} \in \mathbb{C}^{N_0 \times T} \) represent the pilot sequences transmitted by the BS and UE, in which the channel sensing period is represented by \( T \) pilot symbols. Hence, the following is a possible expression for the signal received at the Ns sensing devices:

\[
Y_S = Q(\sqrt{P_k} \mathbf{G} X_B + \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H} k X_U + n_S)
\]

(7)

where \( Q(\cdot) \) is the quantization function that depends on the precision of the ADCs. The transmit power at the BS and the UE can be represented as \( P_k \), respectively. The AWGN at the sensing devices is represented as \( n_S \in \mathbb{C}^{N_s \times T} \). This is particularly advantageous in millimeter-wave [18] or THz [19] frequency bands.

III. JOINT PHASE SHIFT AND BEAMFORMING

A. Cascaded Channel

In a fully passive RIS scenario, estimating the combined effect of the BS and the RIS on the signal requires considering both the beamforming applied at the BS and the phase shifts introduced by the RIS elements. This combined influence is reflected in the received signal at the UE, which can be expressed mathematically as:

\[
Y_{U,E} = F \phi_1 x + n
\]

(8)

where \( \phi = \text{diag}(\phi_x, \ldots, \phi_x) \in \mathbb{C}^{J_X \times J_X} \), and \( J_X = [1, \ldots, 1] \). The matrix \( J_X \) is a \( K \times N \) matrix where in the \( n \)th column, all elements are set to one and all other elements are set to zero. The beamforming at the BS can be derived as:

\[
x = W s
\]

(9)

where \( W = [w_1, \ldots, w_Q] \), where, \( w_q \) represents the \( q \)-th beamforming vector, and \( s = [s_1, \ldots, s_Q]^T \), where \( s_q \) signifies
the $q^b$th transmit data, with the assumption that $E[s_q^2] = 1$ for all $q$. Upon substituting (8) into (9), we obtain:

$$y_{UE} = F \phi_1 F W s + \eta$$

(10)

The joint optimization of beamforming and phase shifting aims to maximize the capacity of the channel between the source (S) and destination (D), as:

$$\max_{w, \phi} C = \log \det \left[ I + \frac{AWW^H A^H}{\sigma^2} \right]$$

(11)

where $A = F \phi_1 F \in C^{M \times N}$. The Singular Value Decomposition (SVD) of matrix $A$ can be represented as $A = U \Sigma V^H$, where $U$ is a matrix in $C^{M \times M}$ representing the left singular vectors, $V$ is a matrix in $C^{N \times N}$ representing the right singular vectors, and $\Sigma$ is a matrix in $C^{M \times N}$ representing the singular values. When $\phi_1$ is fixed, the optimal beamforming vectors are:

$$w_q = \sqrt{p_q} v_q, \quad q = 1, 2, \ldots, Q$$

(12)

Consider $v_q$ as the right singular vector of $q^b$th in $V$, and $p_q$ denoting the power allocated for the transmit data of $q^b$th to ensure the satisfaction of the transmit power constraint $\sum_{q=1}^{Q} p_q \leq P_\text{t}$, obtained through water-filling. Substituting (12) can give:

$$\max_{w, \phi} C = \log \det \left[ I + \frac{p - H}{\sigma^2} \right]$$

(13)

where $P = \text{diag}(P_1, \ldots, P_Q, 0, \ldots, 0)$. $H = \frac{1}{2}$.

To achieve the highest possible system capacity, the cascaded network undergoes training. This is achieved through the minimization of the negative values of the network’s outputs, essentially aiming for the most positive values:

$$\min \left\{ \sum_{i=(C)} (C_{p(i), q(i)})^2 \right\}$$

(14)

### B. Separate Channel

This section focuses on designing the phase control matrix at the RIS and the beamforming vectors for both the BS and UE when separate estimation of the BS to RIS (BS-RIS) and RIS to the UE (RIS-UE) channels, denoted by $H_{T,K}$ and $H_{K,R}$ respectively, is performed using a semi-passive RIS. The equation representing the signal received at the RIS’s sensing device ($y_{RIS}$) is given by:

$$y_{RIS} = y_{BS} + y_{UE}$$

(15)

$$y_{RIS} = \sum_{k=1}^{K} P_{BS} H_{T,K} \Theta_k^o x_k + \eta_2$$

(16)

$$y_{RIS} = \sum_{k=1}^{K} P_{UE} H_{K,R} \Theta_k^o x_k + \eta_2$$

(17)

Assume a diagonal matrix $\phi$ of size $MN \times MN$ with all diagonal elements equal to $\phi_k$, represented as:

$$\phi = \text{diag}(\phi_1, \ldots, \phi_N) \in C^{MN \times MN}$$

Let $x_{BS}$ and $x_{UE}$ denote the pilot sequences transmitted by the BS and UE, respectively, $T$ represents the number of pilot symbols transmitted during the channel sensing period. The beamforming matrix at the sensing device, denoted by $N_s$, can be obtained as:

$$x_{RIS} = x_{BS} + x_{UE}$$

(18)

$$x = WS$$

(19)

The objective of jointly optimizing beamforming and phase shifting is to maximize the capacity of the channel between the S and D, where W represents the $q^b$th beamforming vector and $s$ denotes the $q^b$th transmit data from both S and D.

$$\max_{w, \phi} C = \sum_{k=1}^{K} \log_2 (1 + \text{SNR}) (h_{T,K} \Theta_k^o h_{K,R}^T \phi_k)^2$$

(20)

The uplink channels from the transmitter and receiver to the RIS at the $k^b$th subcarrier are denoted as $H_{T,K}$ and $H_{K,R}$, respectively. Leveraging reciprocity, the downlink channels from the RIS to the transmitter and receiver are given by the transposes $H_{T,K}^H, H_{K,R}^H$. With this notation, the received signal at the receiver can be expressed as:

$$y_{UE} = H_{T,K}^o H_{T,K} x_k + \eta_k$$

(21)

### IV. DEEP LEARNING MODEL DESIGN FOR OPTIMAL PHASE SHIFT AND HIGH-CAPACITY DATA RATE

#### A. Deep Learning Structure

The updated deep learning model employs two distinct neural networks, Net1 and Net2, as observed in Figure 2. Net1 is the main operating network. The system’s inputs, which are $N \times M \times K$, stand for each item in a collection of $2^b$ dimensional category vectors that are cascaded $K$ times. There is one RIS element for every set of $R$ bits, and each set indicates a different phase shift.

![Deep Learning Network Structure](image)

Fig. 2. The deep learning network structure.

The primary network Net1 is trained indirectly by combining Net1 and Net2 to create a powerful learning pipeline of cascaded networks. This pipeline trains without the need for labeled data by minimizing the negative of its final output. After the training phase, Net1 generates phase shifts to maximize the network capacity for a given channel (represented as $F$). With these optimal phase shifts, the ideal beamforming configuration at the source is determined by (12). Net2 efficiently processes the channel information ($F$) and the preceding matrix $\phi_k$ to directly estimate the resulting channel capacity $C_{\phi}$, $\phi$.

#### B. The Training Phase of Deep Neural Networks

Algorithm 1 describes the training phase of Net1 and Net2 for cascaded and separate channel estimation.
two-antenna users were linked to two BSs equipped with antennas via a RIS with phase shift matrix \( \Phi \).

Obtain the equivalent channel matrix from \( F \) and \( \Phi \)

\[ A = F \operatorname{diag}(\Phi_1, \ldots, \Phi_i) \]

Obtain optimal phase control matrix \( \Phi_v \)

\[ \text{Input: 40,000 frames channel matrix} \]

\[ \text{2. Train Network 1} \]

\[ \text{Input: 40,000 frames channel matrix} H_{\text{H}}(X,K) \text{, true phase shift } \Phi_{\text{true}}, \text{ and random phase shift } \Phi_v \]

\[ i = 0 \text{ (initial epoch)} \]

\[ \text{Initialize optimizer = Adam} \]

\[ \text{While } i \leq 600 \text{ (max epochs)} \]

\[ \text{Obtain the equivalent channel matrix from } H_{\text{H}}(X,K) \]

\[ \text{Obtain optimal phase control matrix } \Phi_v \]

\[ \text{End While} \]

\[ \text{Return } \Phi \]

Output: \( \Phi_v \) (estimated optimal phase shifts)

3. Cascade Net 1 with Net 2 after training Net 1

4. Train Network 2

\[ \text{Input: 40,000 frames channel matrix} H_{\text{H}}(X,K) \text{ optimal phase shift matrix } \Phi_v \]

\[ i = 0 \text{ (initial epoch)} \]

\[ \text{Initialize optimizer = Adam} \]

\[ \text{Initialize loss function = MSE} \]

\[ \text{While } i \leq 600 \text{ (max epochs)} \]

\[ \text{Obtain the equivalent channel matrix from } F \text{ and } \Phi \]

\[ A = F \operatorname{diag}(\Phi_1, \ldots, \Phi_i) \]

\[ \text{Utilize the water-filling algorithm to determine the transmit power } P \]

\[ \text{Obtain the capacity as} \]

\[ \text{max}_{\psi_{\text{opt}}} C = \log \text{det}\left( I + \frac{P}{\sigma^2} A \right) \]

where \( A = F \psi_{\text{opt}} \in \mathbb{C}^{K \times N} \)

\[ \text{Calculate the loss function as} \]

\[ \text{min}_{i < \text{max epochs}} \left( \sum \log(1 + \psi_{\text{opt}}(\phi_{i})^2) - C_{\text{opt}}(\theta_i) \right)^2 \]

\[ \text{End While} \]

\[ \text{Return } C \]

Output: \( C_{\text{opt}} \) (corresponding capacity)

End Procedure

V. SIMULATION RESULTS

An RIS-supported network was tested, in which \( M = 2 \), two-antenna users were linked to two BSs equipped with \( N = 2 \) antennas via a RIS with \( K_{\text{reconfig}} = 8 \) and \( K_{\text{reflective}} = 80 \) components. All simulations presented below assume that the channels undergo Rician fading with a Rician factor \( K_{\text{Rician}} \) of 10 dB. Additionally, the phase shifts at each RIS element are quantized using three bits. Figure 3 demonstrates the capacities of the RIS-MIMO network, where both the source and destination are equipped with multiple antennas. The results portrayed are for RIS configurations with varying numbers of elements, specifically 8, 32, and 64. Figure 3 indicates that the data rates achieved by both the cascade and separate channels are high, which is an amazing performance considering the signal-to-noise ratio.

Figures 4 and 5 illustrate the normalized path loss performance across various observation angle values with \( K = 8 \), exhibiting a similar trend. The simulation parameters are consistent with Figure 3. The spatial correlation in \( H \) is greater at \( \alpha = \lambda/2 \) compared to when \( \alpha = \lambda/5 \) gives better performance. However, when the distance \( \alpha \) is equal to the wavelength \( \lambda \), the path loss is notably high, albeit less than in the case of \( \alpha = \lambda/2 \) wavelength, and it represents the worst performance compared to a wavelength of \( 4\lambda \). The proposed technique is advantageous in both cases, leading to higher-quality estimated channels. However, the route loss is quite large when the distance \( \alpha \) is equal to the wavelength \( \lambda \), but it is lower than when \( \alpha = \lambda/2 \), and it is the worst performance when compared to a wavelength of \( 4\lambda \). Positioning the reconfigurable intelligent surfaces either near the source or far from the destination results in a very high level of overall channel gain. The channel gain is maximum when the reflecting surface is close to the source or target, or when the distance between the RIS source and the RIS destination is zero. Figure 7 manifests the difference between LOS and NLOS. The data rate was also calculated. For the first scenario, the data rate is extremely high when employing LOS, presuming there is no obstruction between the sender and the receiver.
Cascaded and Separate Channel Estimation based on CNN for RIS-MIMO Systems

The second scenario employs the reflecting surface method while utilizing NLOS technology, assuming that there is an obstruction preventing a direct path between the transmitter and the receiver. In this case, the use of reflecting surfaces has a noticeable impact. The results disclose that the proposed reflected surfaces yield a higher data rate than the standard surface. Figure 8 compares the performance of various PSBA-based separate channel estimator architectures, such as CBDNet [20], GAN-CBDN [21], CV-DnCNN [22], and MDRN [21], against conventional channel estimating methods, including ADMM [23] and PARAFAC [24]. The results of the simulations were averaged over 400 iterations for the proposed method. Notably, both PSBA-H_{K,R} and PSBA-H_{T,K} exhibit superior NMSE performance compared to GAN-CBD and CBDNet by 5.63 and 4.51 dB, respectively. Additionally, when compared to CV-DnCNN, which also employs CNN, along with traditional ADMM and PARAFAC methods, PSBA demonstrates lower complexity while achieving comparable NMSE performance.

Figure 9 compares various models, namely MRDN, CBDNet, and GAN-CBD. It can be observed that PSBA-H_{T,K,R} shows superior NMSE performance and faster convergence compared to other models. This advantage stems from its network judgment capability, which outperforms PSBA-H_{T,K} and PSBA-H_{K,R}. Additionally, PSBA-H_{T,K,R} offers reduced computational complexity in training and offline operation, leading to enhanced robustness of the channel estimator across different scenarios. The average running time of PSBA-H_{T,K,R} is 0.0073 s, while PSBA-H_{T,K} and PSBA-H_{K,R} exhibit running times of 0.0096 and 0.0092 s, respectively. Furthermore, the computational complexity in both training and offline operations for PSBA-H_{T,K,R} is lower than that of PSBA-H_{T,K} and PSBA-H_{K,R}.
VI. CONCLUSION AND FUTURE WORK

This study presented a DL network that simultaneously optimizes phase shifts and beamforming in the RIS-MIMO system, taking into account cascaded and separate channels. A two-stage channel estimation approach was proposed for the RIS-MIMO communication system, with the first stage estimating the cascaded channel between the BS-RIS-UE and the second stage estimating the separate channels between the BS-RIS and the RIS-UE. Then, to make the most of the user's channel gain, a phase shift design and a beamforming approach were demonstrated for the RIS. The incorporation of RIS into bands with higher frequencies, such as the THz band, also poses a formidable obstacle. More accurate channel estimate algorithms are needed to guarantee dependable communication while deploying RIS in complex and dynamic contexts, such as cities with many buildings, moving cars, or satellite communications [25].

REFERENCES


