Two Time-Scale Learning for Beamforming and Phase Shift Design in RIS-aided Networks

1st Joohyun Cho
Dept. of ECE
Univ. of Utah
joohyun.cho@utah.edu

2th Xiang Huang
Dept. of ECE
Univ. of Utah
eric.xiang.huang@utah.edu

3rd Rong-Rong Chen
Dept. of ECE
Univ. of Utah
rchen@ece.utah.edu

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838777

***Abstract*—In this work, we develop a two time-scale deep learning approach for beamforming and phase shift (BF-PS) design in time-varying RIS-aided networks. In contrast to most existing works that assume perfect CSI for BF-PS design, we take into account the cost of channel estimation and utilize Long Short-Term Memory (LSTM) networks to design BF-PS from limited samples of estimated channel CSI. An LSTM channel extrapolator is designed first to generate high resolution estimates of the cascaded BS-RIS-user channel from sampled signals acquired at a slow time scale. Subsequently, the outputs of the channel extrapolator are fed into an LSTM-based two stage neural network for the joint design of BF-PS at a fast time scale of per coherence time. To address the critical issue that training overhead increases linearly with the number of RIS elements, we consider various pilot structures and sampling patterns in time and space to evaluate the efficiency and sum-rate performance of the proposed two time-scale design. Our results show that the proposed two time-scale design can achieve good spectral efficiency when taking into account the pilot overhead required for training. The proposed design also outperforms a direct BF-PS design that does not employ a channel extrapolator. These demonstrate the feasibility of applying RIS in time-varying channels with reasonable pilot overhead.**

I. Introduction

In recent years, reconfigurable intelligent surface (RIS)-aided networks have attracted significant attention as an emerging technology to increase the network capacity. Smart radio environments can utilize RIS to manipulate the propagation of incident electromagnetic waves in a programmable manner to actively alter the channel realization. This turns the wireless channel into a controllable system block that can be optimized to improve overall system performance.

A major challenge in RIS-aided multiuser communication lies in the joint design of base station (BS) transmit beamforming matrix $\mathbf{W}$ and RIS reflection phase shift $\boldsymbol{\theta}$. Due to the intricate relationship between $\mathbf{W}$ and $\boldsymbol{\theta}$ in determining the sum-rate of a multi-user network, the computation complexity for determining the optimal beamforming and phase shift (BF-PS) is often high [1], [2], especially when the number of RIS elements N is large. To address the complexity issue of conventional algorithms, machine learning approaches have been applied to jointly design BF-PS. For instance, reinforcement learning algorithms were adopted in [3], [4] to design optimal BF-PS. These, however, can only be applied to find the optimal BF-PS for one channel realization and one needs to re-run the learning algorithms whenever channel changes. In [5], [6], deep neural networks were proposed to compute the optimal BF-PS such that the same trained network can be used when channel changes. We note that the above existing works (both conventional and machine learning approaches) typically assume that perfect channel state information (CSI) is available for the optimal BF-PS design. This assumption is not practical for RIS-aided communications. The very high dimension of channel matrices, resulting from a large N (in the order of tens to hundreds), together with multiple transmit antennas at the BS and multiple users sharing the spectrum, can induce significant training overhead for channel estimation. In the literature, a lot of works [7]–[10] consider channel estimation for RIS-aided communications. Some works have also applied machine learning [11]–[13] to improve the RIS-aided channel estimation. However, the impact of imperfect channel estimation on the joint BF-PS design, especially on the machine-learning aided BF-PS design, has not been studied.

In [14], deep learning-based channel extrapolation is implemented over both antenna and time domains for the acquisition of time-varying cascaded channels for a RIS-aided system. In [15], an LSTM based neural network design was proposed to utilize the past location information of the mobile users to predict the phase shift of the RIS. Two neural networks are trained separately, one for the phase shift design from location information. The other one is for designing the optimal beamforming matrix $\mathbf{W}$, assuming that the optimal phase shift matrix has been determined from the first network.

In this work, we develop a two time-scale deep learning approach for BF-PS design in time-varying RIS-aided networks. In contrast to most existing works that assume perfect CSI for BF-PS design, we take the cost of channel estimation into account and utilize neural networks to design BF-PS from limited samples of estimated channel CSI. This approach reduces pilot overhead needed for channel estimation and makes the use of RIS in time-varying channels possible. We adopt a joint optimization for the beamforming matrix $\mathbf{W}$ and RIS phase shift $\boldsymbol{\theta}$. This is in contrast to [15] where the optimization is done separately using two independently trained networks. Furthermore, our design does not require the knowledge of

This work is supported by NSF Grants 1817154 and 1824558.

past location information. The main contributions of this pap
are summarized as follows:

- We propose LSTM-based neural networks for the BF-I design over time-varying channels. An LSTM chanr extrapolator is designed first to generate high resoluti estimates of the cascaded BS-RIS-user channel fro sampled signals acquired at a slow time scale. Subs quently, the outputs of the channel extrapolator are f into an LSTM-based two stage neural network for t joint design of BF-PS at a fast time scale of per coheren time.
- To address the critical issue that training overhead i creases linearly with the number of RIS elements, v consider various pilot structures and sampling patter in time and space to evaluate the efficiency and sum-rate performance of the proposed two time-scale design. We find that with our current network architecture, a higher temporal resolution in channel sampling has a more positive impact on the sum-rate performance than a higher spatial resolution does.
- Our results show that the proposed two time-scale design can achieve a sum-rate performance that is close to that of the design assuming perfect CSI. It also outperforms a direct BF-PS design that does not employ a channel extrapolator. These demonstrate the feasibility of applying RIS in time-varying channels with reasonable pilot overhead.

II. System Model

A. Channel Model

We consider a system similar to that of [2], shown in Fig. 1, consisting of one base station equipped with M antennas, one RIS with N unit-cells, and K single-antenna users. Let $\mathbf{G} \in \mathbb{C}^{N\times M}$ denote the channel from BS to RIS. Since the BS and the RIS are at fixed locations, we assume that $\mathbf{G}$ is fixed. Let $\mathbf{h}_{r,k}(t) \in \mathbb{C}^{N\times 1}$ denote the reflected channel from RIS to user k at time t. Channels $\mathbf{G}, \mathbf{h}_{r,k}(t)$ include effects of large scale path loss and small scale fading (with both line-of-sight (LOS) and non-line-of-sight (NLOS) parts). Elements of channel matrices can be independent under the rich scattering assumption of the NLOS part, or correlated, when considering antenna correlations or sparse scattering environment. For a multi-user setting, we assume that $M \geq K$.

The channels $\mathbf{G}$ and $\mathbf{h}_{r,k}(t)$ are modeled by

$$\mathbf{G} = L_1\Big(\sqrt{\frac{\epsilon}{\epsilon+1}}\mathbf{a}_N(\vartheta)\mathbf{a}_M(\psi)^* + \sqrt{\frac{\epsilon}{\epsilon+1}}\bar{\mathbf{G}}\Big) \tag{1}$$

$$\mathbf{h}_{r,k}(t) = L_{r,k}(t)\Big(\sqrt{\frac{\epsilon}{\epsilon+1}}\mathbf{a}_N(\zeta_k(t)) + \sqrt{\frac{\epsilon}{\epsilon+1}}\bar{\mathbf{h}}_{r,k}(t)\Big) \tag{2}$$

where L_1 and $L_{r,k}(t)$ are the path losses, ϵ is the Rician factor. $\mathbf{a}_D(\phi)$ is the steering vector defined as $[1, e^{j\pi sin(\phi)}, ..., e^{j\pi sin(\phi)(D-1)}]^T \in \mathbb{C}^{D\times 1}$ with angular parameter ϕ and antenna number D. ϑ, ψ and $\zeta_k(t)$ are the corresponding angular parameters. $\bar{\mathbf{G}}$ and $\bar{\mathbf{h}}_{r,k}(t)$ are the

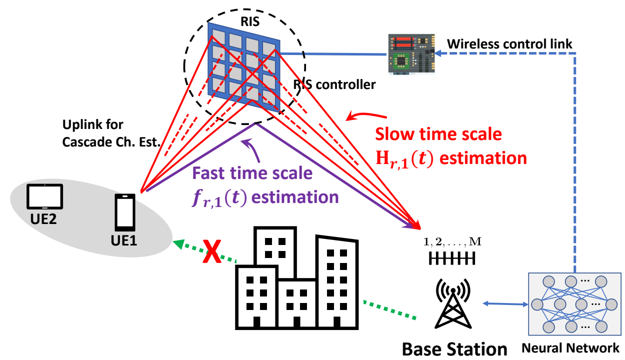


Fig. 1: An RIS-aided multiuser network.

We define the phase shift matrix of the RIS as $\mathbf{\Theta}(t) = \text{diag}(\theta_1(t), \theta_2(t), \cdots, \theta_N(t))$, where $\theta_n(t) = e^{j\phi_n(t)}$ is the phase shift of the n-th RIS element at time t. Let $\mathbf{s}(t) = (s_1(t), \cdots, s_K(t))'$ be the transmitted symbol vector, where $s_k(t)$ denotes the transmitted data symbol to user k at time t. Define the beamforming matrix at the BS as $\mathbf{W}(t) = [\mathbf{w}_1(t), \cdots, \mathbf{w}_K(t)]$, where $\mathbf{w}_k(t) \in \mathbb{C}^{M\times 1}$ denotes the beamforming vector for user k at time t. Then the transmitted signal vector at the BS is $\mathbf{x}(t) = \mathbf{W}(t)\mathbf{s}(t) = \sum_{k=1}^{K}\mathbf{w}_k(t)s_k(t)$. The signal received at user k is given by (4) where $u_k(t)$ is a complex Gaussian random variable with zero mean and variance σ_0^2.

Then, the received signal at user k at time t is given by

$$\begin{aligned} y_k(t) &= \mathbf{h}_{r,k}(t)'\mathbf{\Theta}(t)\mathbf{G}\mathbf{x}(t) + u_k(t), &(3)\\ &= \mathbf{f}_{r,k}(t)'\mathbf{x}(t) + u_k(t), &(4)\end{aligned}$$

where $\mathbf{f}_{r,k}(t)' = \mathbf{h}'_{r,k}(t)\mathbf{\Theta}(t)\mathbf{G} \in \mathbb{C}^{1\times M}$ is defined as the *effective* channel vector of the BS-RIS-user-k channel. The received signal-to-interference-ratio (SINR) at each user k is then given by

$$\gamma_k(t) = \frac{\|\mathbf{f}_{r,k}(t)'\mathbf{w}_k(t)\|^2}{\sum_{i=1,i\neq k}^{K}\|\mathbf{f}_{r,k}(t)'\mathbf{w}_i(t)\|^2 + \sigma_0^2} \tag{5}$$

The objective is to design the optimal BF-PS ($\mathbf{W}(t)$, $\mathbf{\Theta}(t)$) such that the sum-rate in (6) is maximized. The optimization is subject to a total power constraint (7).

$$\max_{\mathbf{W}(t),\mathbf{\Theta}(t)} \sum_{k=1}^{K}\log(1+\gamma_k(t)), \tag{6}$$

$$\text{subject to } \sum_{k=1}^{K}\|\mathbf{w}_k(t)\|^2 \leq P_T. \tag{7}$$

For the optimization of $\mathbf{\Theta}(t)$, it is convenient to introduce $\boldsymbol{\theta}(t) = (\theta_1(t), \cdots, \theta_K(t))'$ to denote the RIS phase shift vector at time t. Then, we can rewrite (4) as (8) such that

$$y_k(t) = \boldsymbol{\theta}'\mathbf{H}_{r,k}\mathbf{x}(t) + u_k(t), \tag{8}$$

where $\mathbf{H}_{r,k}(t) = \text{diag}(\mathbf{h}_{r,k}(t))\mathbf{G} \in \mathbb{C}^{N\times M}$ denotes the *cascaded* channel matrix of the BS-RIS-user-k channel. Here, the notation $\text{diag}(\mathbf{h}_{r,k}(t))$ denotes the diagonal matrix consisting of diagonal elements from $\mathbf{h}_{r,k}(t)$.

We see from (8) that in order to design $\mathbf{\Theta}(t)$, we need to estimate the cascaded channel matrix $\mathbf{H}_{r,k}(t)$ for every user k. This requires an estimation of $K \times M \times N$ channel elements. On the other hand, once $\mathbf{\Theta}(t)$ is given, it is sufficient to estimate the effective channel vector $\mathbf{f}_{r,k}(t)$ for every user k in order determine the optimal beamforming vector. This requires an estimation of only $M \times K$ channel elements. Since the training overhead required to estimate the effective channel is much lower than that is needed for the cascaded channel, in the remainder of the paper, we assume that $\mathbf{f}_{r,k}(t)$ is estimated at the fast time scale (for every coherence time T_c), but the estimation is done less frequently (at a slower time-scale) for $\mathbf{H}_{r,k}(t)$ due to large pilot overhead.

III. Proposed Two Time-Scale Design

A. Transmission Protocols

Let T_{ca} denote the time duration required to estimating the cascaded channels and let T_{c} denote the channel coherence time. Depending on the relationships between T_{ca} and T_c, we consider three data transmission formats as shown in Fig. 2.

- Case 1: $T_{\text{ca}} \ll T_c$. This case applies to a scenario with a small number of RIS elements N. In this case, we can take T_{ca} symbol times within each coherence interval to estimate the cascaded channels. This will not reduce the transmission efficiently significantly.
- Case 2: $T_{\text{ca}} \approx T_c$. This applies to a moderate value of N. Since channel remains roughly constant over the time duration of T_{ca}, we estimate the full cascaded channel using least square estimation. In this case, we assume that the cascaded channel is estimated for every Q_1 coherence times. With each T_c, we will estimate the effective channel using a time duration of T_{ef}, followed by data transmission.
- Case 3: $T_{\text{ca}} \gg T_c$. This case is for a larger value of N. In this case, if one estimates the full cascaded channel, the channel would have changed during T_{ca}. Thus, we choose to turn off some RIS elements and only estimate the cascaded channels for those elements that are on. We denote the time needed to estimate the subset of active RIS elements at $T_{\text{ca}}^{\text{on}}$. After every Q_2 coherence time, we

erent

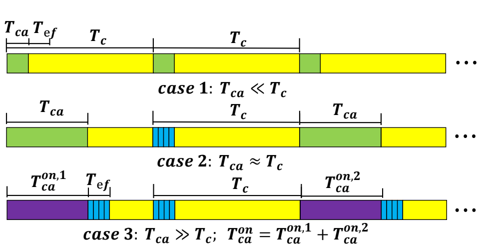


Fig. 2: Three transmission data formats

B. Estimation of Cascaded Channel

To estimate the cascaded channel, we adopt the uplink channel estimation algorithm in [11]. Due to channel reciprocity, the downlink channel can be estimated as the transpose of the uplink channel estimates. Following the channel estimation protocol of [11], each channel estimation frame consists of C sub-frames. In each sub-frame, the RIS is set as one column of the matrix $\mathbf{P} = [\Phi_1, ..., \Phi_C] \in \mathbb{C}^{N \times C}$ and K orthogonal message vectors of length L are transmitted from users, where $L \geq K$. The message vector of user k is denoted as $\mathbf{v}_k = [v_{k,1}, ...v_{k,L}]^H$ and $\mathbf{v}_k^H \mathbf{v}_k = \mathcal{P}$, where $\mathcal{P}$ is the transmitted power of each user over L symbol time. Therefore, at BS, the received vector at l-th time slot via the RIS Φ_c can be expressed as, $\mathbf{r}_{c,l}(t) = \sum_{k=1}^{K} \mathbf{H}'_{r,k}(t)\Phi_c v_{k,l} + \mathbf{n}_{c,l}(t)$, where $\mathbf{n_{c,l}}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$. Based on the assumption that the cascaded channel is constant during channel estimation, the stacked received matrix $\mathbf{R}_c \in \mathbb{C}^{M \times L}$ of L received vectors at c-th sub-frame can be expressed as, $\mathbf{R}_c = \sum_{k=1}^{K} \mathbf{H}'_{r,k} \Phi_c \mathbf{v}_k^H + \mathbf{N}_c$, where $\mathbf{N}_c = [\mathbf{n}_{c,1}, ..., \mathbf{n}_{c,L}] \in \mathbb{C}^{M \times L}$. According the orthogonality of message vectors, the received vector $\mathbf{s}_{c,k}$ subject to the k-th user at the c-th sub-frame can be extracted by multiplying message vector $\mathbf{v}_k$ as,

$$\mathbf{s}_{c,k} = \frac{1}{\mathcal{P}} \sum_{k=1}^{K} \mathbf{H}'_{r,k} \Phi_c \mathbf{v}_k^H \mathbf{v}_k + \frac{1}{\mathcal{P}} \mathbf{N}_c \mathbf{v}_k = \mathbf{H}'_{r,k} \Phi_c + \mathbf{n}'_c,$$

By stacking all received vectors subject to the k-th user over C sub-frames as the matrix $\mathbf{S}_k = [\mathbf{s}_{1,k}, ..., \mathbf{s}_{C,k}] \in \mathbb{C}^{M \times C}$ as, $\mathbf{S}_k = \mathbf{H}'_{r,k} \mathbf{P} + \mathbf{N}'$ where $\mathbf{N}' = [\mathbf{n}'_1, ..., \mathbf{n}'_C] \in \mathbb{C}^{M \times C}$. Then, the Least Square (LS) estimation of uplink cascaded channel $\mathbf{H}'_{r,k}$ can be given by multiplying the pseudoinverse of $\mathbf{P}$ as, $\hat{\mathbf{H}}'_{r,k} = \mathbf{S}_k \mathbf{P}^H (\mathbf{P}\mathbf{P}^H)^{-1}$ The design of matrix $\mathbf{P}$ is also based on the discrete Fourier transform (DFT) as proposed in [11]. However, due to lack of direct channels between users and the BS in our setting, our matrix $\mathbf{P}$ removes the all-one row of the DFT matrix.

C. Proposal two time-scale neural network design

In this section, we propose a two time-scale neural network design, which consists of three modules as shown in Fig. 3. A Space-Time Channel Extrapolation Neural Network(STN) generates extrapolated cascade channel data in time and RIS element space domain every t time slot based on estimated cascade channel. Next, a two-stage neural network, consisting of a Phase Shift Neural Network (PSN) and a Beamforming Neural Network (BFN) is trained jointly to optimize the BF-PS. The PSN-BFN two-stage architecture is motivated from a similar architecture first proposed in [6]. The difference here is that we apply LSTM to both PSN and BFN in order to utilize history channel information, whereas architecture of [6] is designed under the assumption of perfect CSI. The role of the proposed PSN is to obtain the optimized phase shift matrix $\mathbf{\Theta}(t)$ with consideration of both spatial and temporal features of time varying channel. Once the phase shift matrix $\mathbf{\Theta}(t)$ is given, estimated *effective* channel matrix $\mathbf{f}_r(t) = [\mathbf{f}_{r,1}(t), \cdots, \mathbf{f}_{r,K}(t)] \in \mathbb{C}^{K \times M}$ is sent to a

Beamforming Neural Network (BFN) to acquire the optimized beamforming matrix $\mathbf{W}(t)$. After acquiring the beamforming matrix $\mathbf{W}(t)$, the sum rate can be calculated with the *effective* channel matrix according to (6). Therefore, we can design the loss function as the negative of sum rate over L time slots:

$$Loss = -\frac{1}{L}\sum_{t=1}^{L}\sum_{k=1}^{K}\log(1+\gamma_k(t)) \tag{9}$$

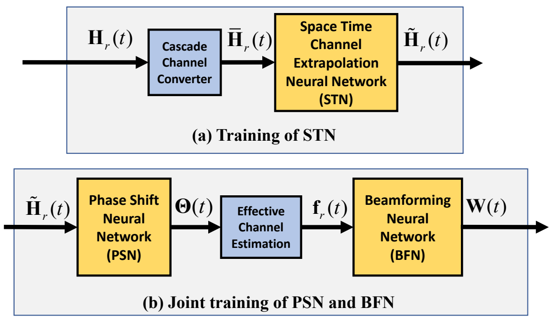


Fig. 3: A system diagram for Two Time-Scale Architecture

We convert the estimated cascade channel matrix $\mathbf{H}_r(t) = [\mathbf{H}_{r,1}(t), \cdots, \mathbf{H}_{r,K}(t)] \in \mathbb{C}^{K\times N\times M}$ of BS-RIS-all users channel into the channel matrix $\bar{\mathbf{H}}_r(t) = [\bar{\mathbf{H}}_{r,1}(t), \cdots, \bar{\mathbf{H}}_{r,N}(t)] \in \mathbb{C}^{N\times K\times M}$ in which $\bar{\mathbf{H}}_{r,n}(t)$ is the converted cascade channel matrix of transmission from base station to all users through n-th RIS element at t time slot. Depending on the transmission protocol, sequence of converted cascade channels has different form. In the case 1, T_{ca} is short enough to estimate cascade channel matrix every T_c and the sequence of converted cascade channel matrix has the form described in Fig. 4(a). However, in case 2, cascade channel $\mathbf{H}_{r,k}(t)$ cannot be estimated every T_c because transmission burden and the sequence of converted cascade channel matrix has zero padded form in temporal axis as Fig. 4(b) in which the grey matrices denote zero matrices. For case 3, T_{ca} is longer than T_c and, as described in Section III.C, cascade channel estimation is performed within T_c by turning off some RIS elements. If κ-th RIS element is turned off in T_{ca}, then the converted cascade channel matrix $\bar{\mathbf{H}}_{r,\kappa}(t)$ is the zero matrix because $\bar{\mathbf{H}}_{r,\kappa}(t)$ is the channel matrix of BS-κth RIS element-all users. Transmission protocol Case 3 has two degree of freedom in time and RIS element space and they are denoted by R_T and R_S, respectively. In Fig. 4(e), half of RIS elements are turned off in each T_C and the cascade channel data is transmitted alternately in time. Thus, the spatial rate R_S is $\frac{1}{2}$, temporal rate R_T is $\frac{1}{2}$, and the total rate $R = R_T \cdot R_S = \frac{1}{4}$.

Space-Time channel extrapolation Neural Network (STN) structure is shown in Fig. 5. It is trained under supervised learning with the MSE loss function. In case 1 ($R_T = R_S = 1$), the converted cascade channel data can be obtained every

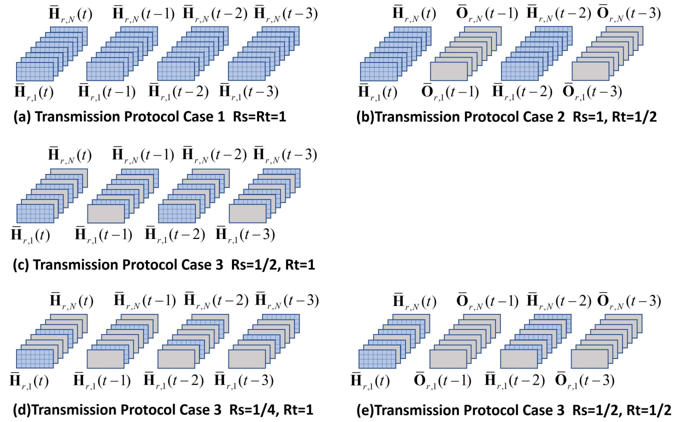


Fig. 4: Estimated Cascade Channel

T_c and STN just passes over the channel matrix sequence to the PSN. However, if R_T or R_S has fractional value, STN performs extrapolation in time domain and RIS element space domain. STN adopts LSTM to take advantage of temporal features from the input sequence but it loses some spatial information via serializing converted cascade channel matrices. We observe that STN has better performance in time than spatial domain in Fig. 6 as well.

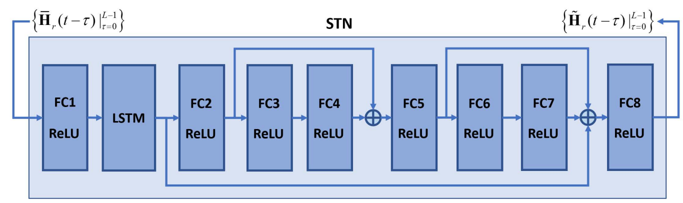


Fig. 5: The architecture of STN

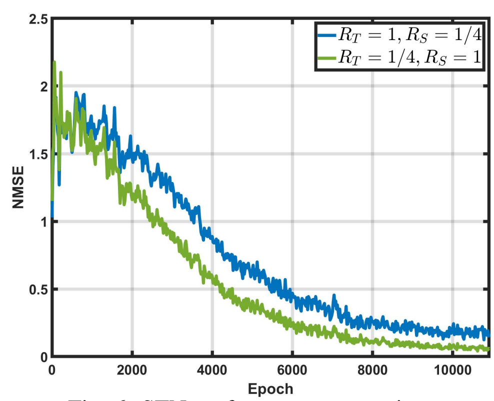


Fig. 6: STN performance comparison

Phase Shift Neural Network (PSN) is composed of Convolutional Neural Network (CNN) module, Long short-term memory (LSTM) cell, and fully Connected Layer, as illustrated in Fig. 7. PSN has sequence of extrapolated converted cascade channel matrices from STN as an input. Considering converted cascade channel matrix's spatial characteristic, CNN module is suitable choice to pre-process input to the LSTM cell. CNN module also reduces the dimension of input to LSTM cell especially in the case of large number of RIS elements environment. Extracted features from CNN module is serialized and concatenated to be sent to LSTM cell for temporal feature exploration. To make use of temporal features of time varying channel, L sequence of extracted features from CNN module are fed to LSTM cell. Hidden cell output from LSTM is sent to a fully connected layer to generate phase shift matrix $\mathbf{\Theta}(t)$.

Given phase shift matrix $\mathbf{\Theta}(t)$, estimated effective channel matrix $\mathbf{f}_r(t)$ is obtained by uplink estimation in T_{ef} as described in Fig. 2. Here, $\mathbf{f}_r(t)$ is flattened and stacked with L numbers in time sequence to be sent to LSTM cell of Beamforming Neural Network (BFN). Fully connected network exploits features from hidden cell output of LSTM cell and makes beamforming matrix $\mathbf{W}(t)$. At the end of BFN, power normalization is conducted to fulfill the total transmit power constraint (7). When PSN and BFN are trained jointly, pre-trained STN is used.

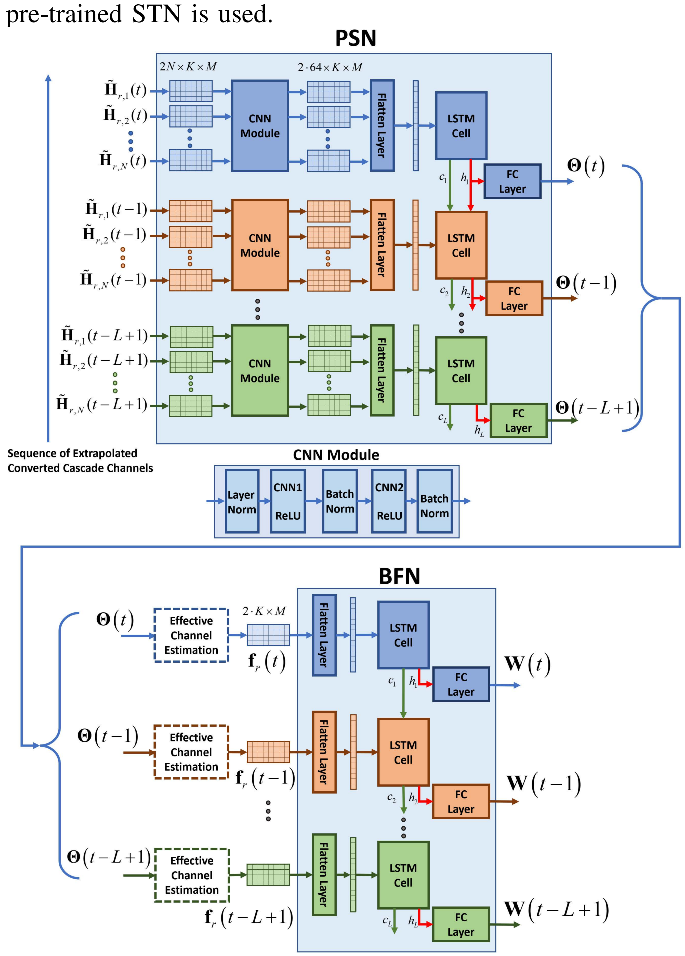


Fig. 7: The architecture of PSN and BFN

IV. Simulation Results

In this section, we evaluate the performance of the proposed two time-scale learning approach. We first describe the channel environment setup and then present the simulation results.

Wireless channel environment setting is as follows:

- Carrier Frequency: $f_c = 3.5$ GHz
- Transmission Bandwidth: $BW = 180$ KHz
- Mobility: $v = 3.85$ m/s
- Maximum Doppler: $f_d = f_c \cdot \frac{v}{c} = 45$ Hz
- Coherence Time: $T_c = \frac{1}{8 f_d} = 2.8$ ms
- Normalized Maximum Doppler: $f_d \cdot T_s = 2.5 \times 10^{-4}$.
- # of symbols per coherence time: $N_C = \frac{T_c}{T_s} = T_c \cdot BW = 500$ symbols

We consider an RIS-aided network, where one BS, equipped with $M = 8$ antennas, communicates $K = 4$ single antenna users via an RIS with N elements. It requires $(N+1) \times K$ symbol times to estimate all elements in the cascaded BS-RIS-user channels. Assume the users are randomly distributed in a circle with a radius of 10 m and the direct links between BS and users are blocked, as illustrated in Fig. 8.

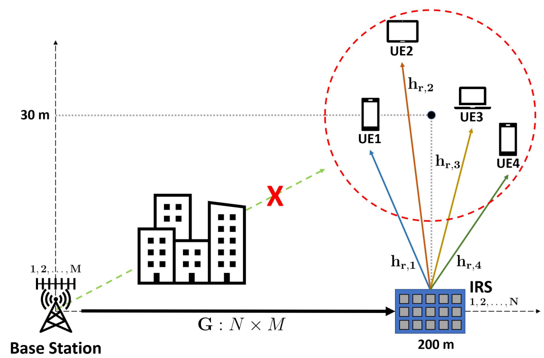


Fig. 8: Simulation Setup

To generate training data, users' locations are randomly selected per 100 T_c. Since the distance travelled as a user moves in 100 T_c is 1.08 meters, the path loss change is less than 0.5 dB. Thus, the path loss is approximately constant within this time (280 ms), while small scale fading changes. We use the filter method to simulate the continuous, time-varying channel with small scale fading. Complex Gaussian with zero mean and unit variance is filtered by the auto-correlation function $R_0[n] = 2a^2 \pi J_0(n \cdot 2\pi f_n)$ in which $f_n = f_d \cdot T_s$ is the normalized maximum Doppler frequency and J_0 is the zeroth-order Bessel function of the first kind: $J_0(x) := \frac{1}{\pi} \int_0^{\pi} e^{jxcos\theta} d\theta$.

In Fig. 9, we first compare the sum rate achieved by three architectures. The first one, CNN+FC with full CSI represents a performance upper bound where full CSI at every T_c is fed into the CNN and Fully Connected (FC) layer architecture [6] for optimal BF-PS design. The performance lower bound is obtained using CNN+FC with $R_T = 1/2$. For this architecture, we feed the two time-scale channel inputs to the same architecture [6]. Here, to apply the temporal rate ($R_T = 1/2$) input, the network is modified to be trained with two time scale transmission protocol. When zero padded cascade channel is fed to CNN, output phase shift matrix holds previous value and CNN blocks are detached from computation graph. Only FC layers for beamforming matrix are updated by backpropagation algorithm. Our proposed architecture is denoted by STN+PSN+BFN, which shows superior performance to the CNN+FC with $R_T = 1/2$, while showing a performance gap compared to the idealized CNN+FC with full CSI.

Next, we will make performance comparisons based on spectral efficiency, which takes pilot overhead into account. With our proposed frame structure, pilot overhead rate for uplink channel estimation can be calculated by $O_p = \frac{(N+1)\cdot K\cdot R_S+(\frac{1}{R_T})\cdot K}{(N+1)\cdot K\cdot R_S+(\frac{1}{R_T})\cdot N_C} = \frac{(N+1)\cdot K\cdot R+K}{(N+1)\cdot K\cdot R+N_C}$ in which $R = R_S \cdot R_T$. Thus, the system spectral efficiency that takes into account pilot overhead is $(1-O_p)\cdot\sum_{k=1}^{K}\log(1+\gamma_k(t))$.

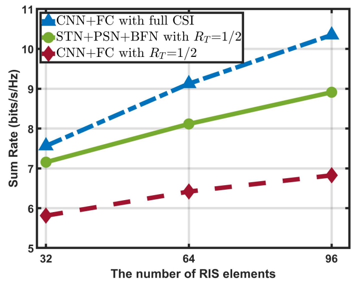


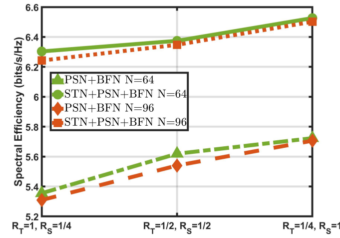
Fig. 9: Sum rate comparisons over various architectures.

Fig. 10a compares the spectral efficiency over three combinations of temporal rate R_T and spatial rate R_S while fixing the total rate R to be $\frac{1}{4}$. Two architectures are compared, one with STN, and one without. Fig. 10a shows that having the STN improves the spectral efficiency by about 1 bits/s/Hz. We also note that increasing N from 64 to 96 only improves the spectral efficiency slightly. This is due to the higher pilot overhead required for the case of $N = 96$. As seen from Fig. 10a, assigning a lower temporal rate R_T than R_S provides higher spectral efficiency.

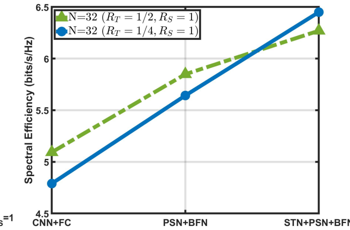
Fig. 10b makes performance comparison over three architectures. LSTM cell in PSN and BFN can exploit previous cascade channel and we can observe taking into account previous channel gives advantage in the time varying channel environment by comparing the architecture (CNN+FC) with the proposed PSN+BFN spectral efficiency. With STN, spectral efficiency of $R = \frac{1}{2}$ and $R = \frac{1}{4}$ are reversed. Reducing pilot overhead inevitably generates more zero padded channel data both in time and RIS element space and LSTM cells in PSN and BFN can deal with can deal with sparse zero padded input. However, more zero padded input is fed, a module dedicated to channel extrapolation such as SFN can improve spectral efficiency as shown in Fig. 10b.

V. Conclusion

This paper proposed a two-time scale phase shift and beamforming neural network architecture to reduce channel estimation overhead in time-varying channel for RIS-aided networks. In the proposed phase-shift and beamforming architecture, LSTM and CNN are combined to exploit temporal features from time-varying channel and spatial features from the converted cascade channel. Phase shift and beamforming matrix are optimized jointly. To reduce channel estimation overhead, two time-scale transmission protocol is presented and a channel extrapolation Neural Network architecture is designed to manipulate zero padded channel data from the two-time scale protocol. Our results show that the proposed approach is effective to enable BF-PS design for time-varying channels with good spectral efficiency performance.



(a) Over various protocol designs



(b) Over various architectures

Fig. 10: Spectral Efficiency comparison

References

[1] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[2] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064–3076, 2020.

[3] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for miso communication systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 745–749, 2020.

[4] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser miso systems exploiting deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, 2020.

[5] J. Gao, C. Zhong, X. Chen, H. Lin, and Z. Zhang, "Unsupervised learning for passive beamforming," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1052–1056, 2020.

[6] H. Song, M. Zhang, J. Gao, and C. Zhong, "Unsupervised learning based joint active and passive beamforming design for recongurable intelligent surfaces aided wireless networks," *IEEE Communications Letters*, 2020.

[7] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, 2020.

[8] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive mimo," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210–214, 2019.

[9] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mimo systems," *arXiv preprint arXiv:1912.03619*, 2019.

[10] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for ris-empowered multi-user miso wireless communications," *IEEE Transactions on Communications*, 2021.

[11] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *arXiv preprint arXiv:2009.01423*, 2020.

[12] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-wave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1447–1451, 2020.

[13] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Deep learning for large intelligent surfaces in millimeter wave and massive mimo systems," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[14] M. Xu, S. Zhang, J. Ma, and O. A. Dobre, "Deep learning-based time-varying channel estimation for ris assisted communication," *arXiv preprint arXiv:2108.05661*, 2021.

[15] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Transactions on Wireless Communications*, 2021.