ACTIVE IRS-ASSISTED MIMO CHANNEL ESTIMATION AND PREDICTION

Mirza A. Haider [1], *Saidur R. Pavel* [1], *Yimin D. Zhang* [1], *and Elias Aboutanios* [2]

[1] Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA
[2] School of Electrical Engineering and Telecommunications,
University of New South Wales, Sydney, Australia 2052

ABSTRACT

This paper considers a wireless network assisted by an intelligent reflecting surface (IRS) to enhance data transmission between the base station and mobile users. Our objective is to estimate and predict the user-IRS channels by exploiting a small number of sparsely distributed active elements with a low pilot overhead. The Hermitian and Toeplitz properties of the data covariance matrices are used to perform covariance matrix interpolation for enhanced estimation of the time-varying user-IRS multipath channels, and a machine learning-based channel predictor is developed to predict the channels based on prior channel estimates so as to shorten the required training pilot signals and enhance the transmission data rate. Simulation results verify the effectiveness of the proposed method for accurate channel estimation and prediction.

Index Terms— Intelligent reflecting surface, sparse array, structured matrix completion, channel estimation, machine learning.

1. INTRODUCTION

The development of ubiquitous machine-to-machine and human-to-machine communications in next-generation wireless networks requires faster data transmission, higher channel capacity, more reliable communication, and lower latency [1–4]. Recently, communication networks have expanded to the millimeter-wave frequency band to enable high data-rate applications such as high-quality video transmission, vehicle-to-infrastructure communications, intra- and inter-vehicle messaging, device-to-device communications, Internet of Things frameworks, tactile Internet [5–8].

Millimeter-wave channels have limited scattering and thus are not viable to support mobile users who are not in the line-of-sight (LOS) of the base station (BS) transmitter [9]. Therefore, it is crucial to build extra links between the BS and the mobile users in order to have them reliably connected. Intelligent reflecting surfaces (IRS), which can alter the wireless channels between the BS and the mobile users, are a promising emerging technology for 5G and beyond [10–13]. An IRS is typically a rectangular metasurface made of a high number of passive reflecting elements that may each be digitally modified to induce a specific amplitude change and/or phase shift of the incident signal, enabling broadcasters and receivers to work together to change the wireless links. Designing an IRS with a small number of active components can result in high-precision performance for pilot overhead reduction, wireless channel reconfiguration, and channel estimation [14–16]. Note that active elements in this context imply that the signals received by these elements can be digitized for processing.

In [14], the effectiveness of active IRS-assisted channel estimation, equipped with randomly distributed sparse active IRS elements and using compressive sensing and machine learning (ML) techniques, is demonstrated. References [15, 16] show that exploitation of structural deployment of sparse active IRS elements further enhances signal parameter estimation using subspace-based direction-of-arrival (DOA) estimation methods, such as ESPRIT and MUSIC. It is revealed that the number of training data required for channel estimation is independent of the number of IRS elements so the training overhead is manageable even for a large IRS. Exploiting active IRS elements offers several attractive advantages, particularly for target recognition, user localization, and secure connectivity [17–19].

Several methods have been developed for multiple-input multiple-output (MIMO) and multiple-input single-output (MISO) systems to predict channels with realistic priors channel state information (CSI) [20–22]. In [21], a comparative analysis between a Kalman filter-based channel predictor and an ML-based channel predictor using the realistic channels from the spatial channel model. According to the analysis, the overall complexity of the Kalman filter-based channel prediction is lower than that of ML-based techniques. However, because ML-based techniques can train the neural network (NN) offline, once the neural network is trained, the ML-based predictor requires much lower complexity than one based on the Kalman filter.

The objective of this paper is, while providing accurate channel estimation between the IRS and mobile users, to enhance the data transmission by predicting the channels based on the prior estimated channel knowledge. A MISO channel model is shown in Fig. 1. We first estimate the fast-fading multipath channel between the mobile user and the IRS with the aid of an L-shaped sparse array of active IRS elements. We then train a multilayer perceptron (MLP) network to forecast the subsequent channel based on the previous estimates of the IRS-user channels. Simulation results verify that the proposed technique increases the transmission data rate between the BS and mobile users via the IRS by reducing the number of required training pilots and achieving faster channel estimation.

Notations: We use lower-case and upper-case bold characters to denote vectors and matrices, respectively. In particular, $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{\dagger}$ respectively represent the transpose, Hermitian, and pseudo-inverse operations of a matrix or a vector. $\|.\|_*$ and $\|.\|_{\mathrm{F}}$ respectively represent the nuclear norm and the Frobenius norm of a matrix. $\mathrm{Tr}(\cdot)$ represents the trace operator, and $\mathrm{diag}(\cdot)$ forms a diagonal matrix from a vector. $\otimes$ computes the Kronecker product. $\mathbf{A} \succcurlyeq 0$ implies that matrix $\mathbf{A}$ is positive semidefinite. In addition, $\mathbb{E}[\cdot]$ stands for the statistical expectation operator and $\mathbb{C}^{M\times N}$ denotes the $M \times N$ complex space.

2. SYSTEM MODEL

Consider a MISO downlink communication system with an IRS, equipped with a small number of active elements, deployed to aid communications between a single-antenna user and the BS equipped

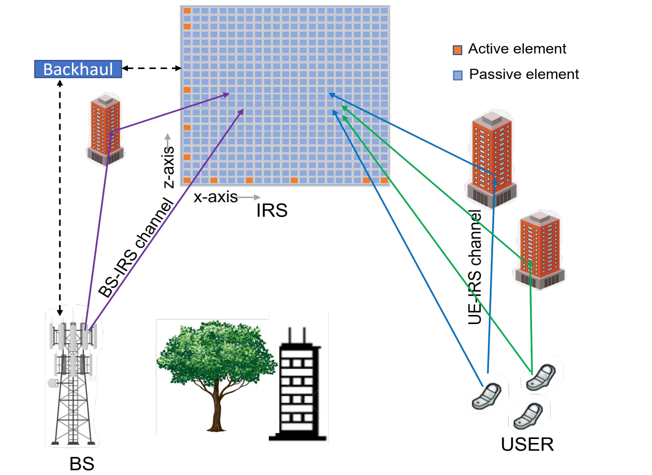

Fig. 1. IRS-assisted channel model.

with N_b antennas. The IRS, which is a uniform planar array consisting of $M = M_x \times M_z$ elements, is deployed in a scenario where the direct links between the user and the BS are significantly obstructed. Among the M reflecting elements, the IRS contains $\bar{M}$ active elements, which have both sensing and reflecting capabilities. As illustrated in Fig. 1, the active elements are arranged in a sparse L-shape, which consists of a horizontally (x-axis) placed subarray and a vertically (z-axis) placed subarray. The numbers of the active IRS elements in these two subarrays are respectively $\bar{M}_x$ and $\bar{M}_z$, and the total number of the active elements is $\bar{M} = \bar{M}_x + \bar{M}_z - 1$ because the element in the corner is shared by both subarrays. The two-dimensional (2-D) steering vector $\mathbf{a}_{\mathrm{IRS}}(\varphi, \vartheta)$ of the entire IRS can be expressed as [23]

$$\mathbf{a}_{\mathrm{IRS}}(\varphi, \vartheta) = \mathbf{a}_z(\varphi) \otimes \mathbf{a}_x(\vartheta), \tag{1}$$

where $\varphi \in [-\pi/2, \pi/2]$ and $\vartheta \in [-\pi/2, \pi/2]$ are the azimuth and elevation angles, and

$$\mathbf{a}_x(\varphi) = [1, e^{-j\frac{2\pi}{\lambda} d \sin(\varphi)}, \ldots, e^{-j\frac{2\pi}{\lambda} d (M_x - 1)\sin(\varphi)}]^{\mathrm{T}}, \tag{2}$$

$$\mathbf{a}_z(\vartheta) = [1, e^{-j\frac{2\pi}{\lambda} d \sin(\vartheta)}, \ldots, e^{-j\frac{2\pi}{\lambda} d (M_z - 1)\sin(\vartheta)}]^{\mathrm{T}}. \tag{3}$$

In the above expressions, λ is the wavelength and d is the inter-element spacing between IRS elements in both x and z directions.

The M-dimensional channel vector between the IRS and the mobile user is denoted as the superposition of L paths, expressed as

$$\mathbf{h} = \sum_{l=1}^{L} \beta_l \mathbf{a}_{\mathrm{IRS}}^{\mathrm{H}}(\vartheta_l, \varphi_l), \tag{4}$$

where β_l is the path gain for $l = 1, 2 \cdots, L$.

The structure of the time frames used for channel detection and prediction is shown in Fig. 2, where B denotes the channel order. The channel within the short time period of each time frame $T_n = T - n + 1$ for $n = B, B-1, \cdots, 1$ can be considered slowly time-varying. Each time frame T_n contains an IRS transmission mode and an IRS sensing mode. In the sensing mode, the IRS collects $\dot{T}$ samples for each time frame to estimate the user-IRS channel $\mathbf{h}_{T_n} \in \mathbb{C}^{M\times 1}$.

3. CHANNEL ESTIMATION AND PREDICTION

We consider channel estimation and prediction in two phases. In phase I, we estimate the uplink channels $\mathbf{h}_{T_n}$ between the user and the IRS using the limited number of active IRS elements. To enhance the channel estimation capability and performance, we use the nuclear norm-based interpolation technique to estimate the DOAs of the user-IRS multipath channel and the associated channel gains at the IRS to reconstruct the full channel. Based on estimated channels $\mathbf{h}_{T-B+1}, \mathbf{h}_{T-B}, \cdots, \mathbf{h}_T$ for consecutive time frames $T - B + 1, T - B, \cdots, T$, we then use an MLP network in phase II to predict channel $\mathbf{h}_{T+1}$ without requiring training samples in the $(T+1)$th time frame, so that the full frame at time $T+1$ can be used for data transmission. These two phases are described in detail in the following two subsections.

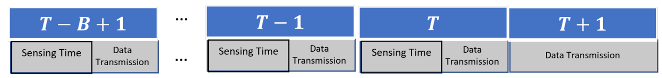

Fig. 2. The transmission frame structure.

3.1. Phase I: User-IRS Channel Estimation

In phase I, the active IRS elements are used to estimate the multipath channel between the user and IRS. The uplink signal is characterized by L uncorrelated far-field narrowband multipath components impinging to the IRS from DOAs $\{\varphi_l, \vartheta_l\}$ for $l = 1, \cdots, L$.

3.1.1. 2-D DOA and Path Gain Estimation

During the sensing mode, the IRS uses its active elements to receive signals from the user. At any time t in time frame T_n, the received signal at the IRS corresponding to the x-axis and the z-axis subarrays are respectively expressed as [15,23]

$$\mathbf{x}(t) = \sum_{l=1}^{L} \beta_l \mathbf{a}_X(\varphi_l) s_u(t) + \mathbf{n}_X(t), \tag{5}$$

$$\mathbf{z}(t) = \sum_{l=1}^{L} \beta_l \mathbf{a}_Z(\vartheta_l) s_u(t) + \mathbf{n}_Z(t), \tag{6}$$

where $s_u(t) = \sqrt{P} s(t)$ denotes the source signal transmitted by the mobile user, P is the transmitted power, $s(t)$ is the transmitted waveform with $\mathbb{E}(|s(t)|^2) = 1$, and $\mathbf{n}_X(t)$ and $\mathbf{n}_Z(t)$ are the additive white Gaussian noise (AWGN) vectors.

We use the optimized non-redundant array (ONRA) structure for both sparse linear subarrays of active elements in the x- and the z-axes [24, 25]. The positions of the active elements along the two subarrays are represented by $\mathbb{X} = \{p_0, p_1, \cdots, p_{\bar{M}_x - 1}\}\lambda/2$ and $\mathbb{Z} = \{q_0, q_1, \cdots, q_{\bar{M}_z - 1}\}\lambda/2$, respectively, where p_i and q_i are integers for all i, and $p_0 = q_0 = 0$ is assumed. We also denote $W_x = p_{\bar{M}_x - 1} + 1$ and $W_z = p_{\bar{M}_z - 1} + 1$ as the lengths of active and passive elements included within the respective apertures of the two subarrays. In addition, $\mathbf{a}_X(\varphi_l) \in \mathbb{C}^{W_x \times 1}$ and $\mathbf{a}_Z(\vartheta_l) \in \mathbb{C}^{W_z \times 1}$ respectively denote the steering vectors of the two subarrays along the x and z axes.

Assuming that the noise is uncorrelated to the signals, we use the MUSIC algorithm to estimate the DOAs of the multipath signals. We first consider the elevation angles and stack the received signals at the vertical subarray in the following matrix form at each time frame, i.e.,

$$\mathbf{Z} = [\mathbf{z}(1), \mathbf{z}(2), \cdots, \mathbf{z}(\dot{T})] \in \mathbb{C}^{W_z \times \dot{T}}. \tag{7}$$

We use $\mathbf{Z}$ to estimate the covariance matrix of $\mathbf{z}(t)$, given as

$$\mathbf{R}_{Z_{\mathrm{IRS}}} = \mathbf{Z}\mathbf{Z}^{\mathrm{H}} = \mathbf{A}_Z \mathbf{R}_s \mathbf{A}_Z^{\mathrm{H}} + \sigma_n^2 \mathbf{I}_{W_z}, \tag{8}$$

where $\mathbf{R}_s = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_L^2)$, $\sigma_l^2 = |\beta_l|^2$ represents the power of the l-th path signal, and σ_n^2 denotes the noise power. Because of the sparse spacing between the elements, the covariance matrix $\mathbf{R}_{Z_{\mathrm{IRS}}}$ is sparse with missing elements in the half-wavelength space. We consider matrix interpolation of $\mathbf{R}_{Z_{\mathrm{IRS}}}$ to obtain the covariance matrix of the uniform linear array along the z-axis from the following nuclear norm minimization problem [26,27]:

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \|\mathcal{T}(\boldsymbol{w})\mathbf{V} - \mathbf{R}_{Z_{\mathrm{IRS}}}\|_{\mathrm{F}}^2 + \zeta\|\mathcal{T}(\boldsymbol{w})\|_* \\ \text{subject to} \quad & \mathcal{T}(\boldsymbol{w}) \succcurlyeq 0, \end{aligned} \tag{9}$$

where $\mathcal{T}(\boldsymbol{w})$ denotes the Hermitian-Toeplitz matrix with $\boldsymbol{w}$ as its first column, $\|\mathcal{T}(\boldsymbol{w})\|_* = \mathrm{Tr}(\sqrt{\mathcal{T}^{\mathrm{H}}(\boldsymbol{w})\mathcal{T}(\boldsymbol{w})})$ is the nuclear norm of $\mathcal{T}(\boldsymbol{w})$, and ζ is a tunable regularization parameter. In addition, $\mathbf{V} = \mathbf{v}_p\mathbf{v}_p^{\mathrm{T}}$ is the binary mask matrix indicating whether each element in the sparse covariance matrix is measured, where

$$\langle \mathbf{v}_p \rangle_g = \begin{cases} 1, & gd \in \mathbb{Z}, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

and $g \in \{p_0, p_1, \cdots, p_{N_x-1}\}$ is the index of the active element locations and $\langle \cdot \rangle_g$ denotes the element corresponding to the position at gd. The interpolated covariance matrix is denoted as $\hat{\mathbf{R}}_{z_{\mathrm{IRS}}} = \mathcal{T}(\boldsymbol{w}) \in \mathbb{C}^{W_z \times W_z}$. We apply the MUSIC algorithm to $\hat{\mathbf{R}}_{Z_{\mathrm{IRS}}}$ to estimate the elevation DOAs of the user-IRS multipath signals. The same way we can also determine the interpolated $\hat{\mathbf{R}}_{x_{\mathrm{IRS}}}$.

3.1.2. Pair-Matching for 2-D DOA Estimation

In a multipath propagation environment, it is important to correctly pair between the estimated elevation angles and the corresponding azimuth angles. The array manifold matrix of the z-axis subarray can be constructed according to the estimated elevation angles as

$$\hat{\mathbf{A}}_z = [\mathbf{a}_z(\hat{\vartheta}_1), \mathbf{a}_z(\hat{\vartheta}_2), \cdots, \mathbf{a}_z(\hat{\vartheta}_L)] \in \mathbb{C}^{M_z \times L}. \tag{11}$$

The interpolated cross-covariance matrix between $\mathbf{z}(t)$ and $\mathbf{x}(t)$ can be achieved as [28]

$$\hat{\mathbf{R}}_{zx} = \hat{\mathbf{V}}_{zs}(\hat{\mathbf{\Lambda}}_{zs} - \sigma_{\mathrm{n}}^2\mathbf{I})^{\frac{1}{2}}(\hat{\mathbf{\Lambda}}_{xs} - \sigma_{\mathrm{n}}^2\mathbf{I})^{\frac{1}{2}}\hat{\mathbf{V}}_{xs}^{\mathrm{H}}, \tag{12}$$

where $\hat{\mathbf{V}}_{xs}$ and $\hat{\mathbf{V}}_{zs}$ denote the estimated signal subspaces for the two linear arrays, whereas $\hat{\mathbf{\Lambda}}_{xs}$ and $\hat{\mathbf{\Lambda}}_{zs}$ are the diagonal matrices containing the eigenvalues corresponding to the signal subspaces. Noticing that $\mathbf{R}_s$ can be estimated as

$$\hat{\mathbf{R}}_s = \hat{\mathbf{A}}_z^{\dagger}\hat{\mathbf{V}}_{zs}(\hat{\mathbf{\Lambda}}_s - \sigma_{\mathrm{n}}^2\mathbf{I})\hat{\mathbf{V}}_{zs}^{\mathrm{H}}(\hat{\mathbf{A}}_z^{\dagger})^{\mathrm{H}}, \tag{13}$$

we can estimate the steering matrix $\hat{\mathbf{A}}_x$ as [23]

$$\hat{\mathbf{A}}_x = \left(\mathbf{R}_s^{-1}\hat{\mathbf{A}}_z^{\dagger}\mathbf{R}_{zx}\right)^{\mathrm{H}}. \tag{14}$$

From the estimated $\hat{\mathbf{A}}_x \in \mathbb{C}^{M_x \times L}$, we can determine the azimuth angle sequence to reconstruct the steering matrix of the IRS for the user-IRS channel as $\hat{\mathbf{A}}_{\mathrm{IRS}} = [\hat{\mathbf{a}}_{\mathrm{IRS}}(\varphi_1, \vartheta_1), \cdots, \hat{\mathbf{a}}_{\mathrm{IRS}}(\varphi_L, \vartheta_L)] \in \mathbb{C}^{M \times L}$.

3.1.3. Path Gain Estimation

The path gains are identical for the x- and z-axis subarrays. Therefore, to estimate the path gain of the user-IRS channel, computation in one of these two subarrays will suffice. The received signal at the z-direction subarray can be rearranged as

$$\mathbf{y}_z(t) = \mathbf{A}_z\mathbf{g}s_u(t) + \mathbf{n}_z(t), \tag{15}$$

where $\mathbf{g} = [\beta_1, \beta_2, \cdots, \beta_L]^{\mathrm{T}}$ represents the path gains and can be estimated from

$$\hat{\mathbf{g}} = \frac{1}{\sigma_s^2}(\mathbf{A}_z^{\mathrm{H}}\mathbf{A}_z)^{-1}\mathbf{A}_z^{\mathrm{H}}\bar{\mathbf{y}}_z, \tag{16}$$

and $\bar{\mathbf{y}}_z = \mathbb{E}\{\mathbf{y}_z(t)s_u^*(t)\}$. From the above DOA and path gain estimates, we can reconstruct the user-IRS multipath channel $\mathbf{h}_r$.

3.2. Phase II: Machine Learning-Based Channel Prediction

In this section, we develop an ML-based algorithm for predicting channel $\mathbf{h}_{T+1}$. As shown in Fig. 3, the MLP structure has three levels: an input layer, an output layer, and K fully connected hidden layers.

To prepare the dataset for network training, we first concatenated the consecutive channels as $\boldsymbol{v} = [\boldsymbol{h}_{T-B+1}^{\mathrm{T}}, \cdots, \boldsymbol{h}_T^{\mathrm{T}}]^{\mathrm{T}}$. Denoting $\boldsymbol{W}_k$, $\boldsymbol{b}_k$, and $\mathcal{A}_k$ as the weights, biases, and activation function of the kth network for $k = 1, 2, \cdots K$, the predicted output can be obtained as

$$\hat{\boldsymbol{h}}_{T+1} = \mathcal{A}_k(\boldsymbol{W}_k\mathcal{A}_{k-1}(\cdots, \mathcal{A}_1(\boldsymbol{W}_1\boldsymbol{v} + \boldsymbol{b}_1)\cdots) + \boldsymbol{b}_K). \tag{17}$$

This procedure leads to the direct prediction of the upcoming channels from the estimated channels. Note that the same procedure can be used to predict the channels at $t > T + 1$. However, the accuracy of the estimated channels would degrade.

In the MLP training phase, the inputs to the MLP network are the channel vectors $\hat{\mathbf{h}}_{T-B+1}, \hat{\mathbf{h}}_{T-B}, \cdots, \hat{\mathbf{h}}_T$ estimated in phase I, and the output is the reference channel vector $\mathbf{h}_{T+1}$ at the $(T+1)$-th time slot. To exploit a real-valued MLP architecture, we reshape the inputs to a $2BM$-dimensional input layer containing the real and imaginary parts of the complex-valued input vectors, i.e., $\{\mathrm{Re}(\hat{\mathbf{h}}_{T-B+1}), \mathrm{Im}(\hat{\mathbf{h}}_{T-B+1}), \cdots, \mathrm{Re}(\hat{\mathbf{h}}_T), \mathrm{Im}(\hat{\mathbf{h}}_T)\}$. Each hidden layer uses $2BM$ nodes. The output layer is designed to have $2M$ dimensions, which correspond to the real and imaginary parts of channel vector $\{\mathrm{Re}(\mathbf{h}_{T+1}), \mathrm{Im}(\mathbf{h}_{T+1})\}$ at the $(T+1)$-th time slot. The last reshaped layer combines the real and imaginary parts to reconstruct the complex-valued predicted channel vector $\tilde{\mathbf{h}}_{T+1}$.

We use the adaptive moment estimation (ADAM) optimizer and the following mean square error (MSE) is used as the loss function to train the network:

$$C_{\mathrm{loss}} = \frac{1}{N_{\mathrm{train}}}\sum_{n=1}^{N_{\mathrm{train}}}\|\hat{\mathbf{h}}_{\mathrm{T+1}} - \mathbf{h}_{\mathrm{T+1}}\|^2, \tag{18}$$

where $\hat{\mathbf{h}}_{T+1}$ and $\mathbf{h}_{T+1}$ are the predicted and reference channels, respectively, and N_{train} is the number of observations for a channel sequence.

4. SIMULATION RESULTS

The IRS has a total of $M = 23 \times 23 = 529$ elements, among which $\bar{M} = 11$ are active. Each of the two linear subarrays in the L-shape sparse active array comprises 6 sensors, and the positions of the active elements along the x- and the z-axes are $\mathbb{X} = \mathbb{Z} = \{0, 3, 7, 12, 20, 22\}\lambda/2$. The corresponding non-negative lags of the two linear subarrays are $\mathbb{D}_{\mathrm{self}}^X = \mathbb{D}_{\mathrm{self}}^Z = \{0, 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 15, 17, 19, 20, 22\}\lambda/2$.

The BS, RIS, and the initial position of the user are located at (0 m, 0 m, 10 m), (30 m, 4 m, 10 m), and (30 m, 34 m, 1.5 m), respectively. The number of paths between the user and the IRS is $L = 2$. The large-scale path loss for user-IRS distance r meters is given as $\mathrm{PL}(r)[\mathrm{dB}] = 10\log_{10}(4\pi f_c/c)^2 + 10\alpha\log_{10}(r/r_0)$, where

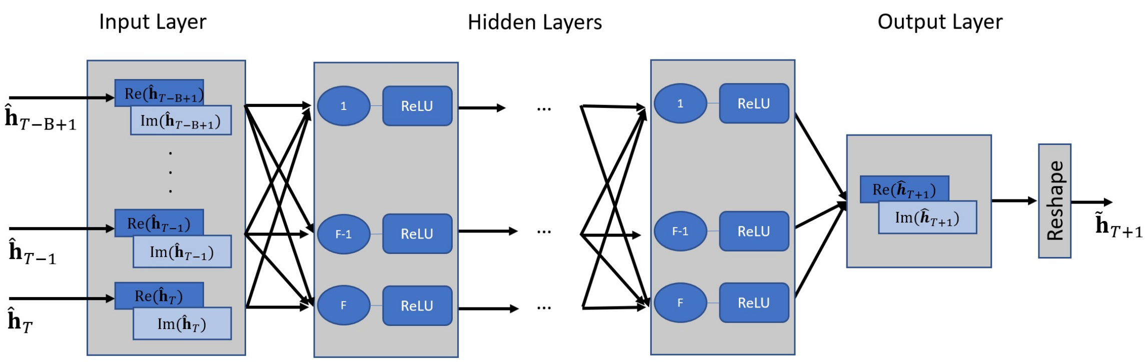

Fig. 3. MLP structure for $\mathbf{h}_{T+1}$ estimation.

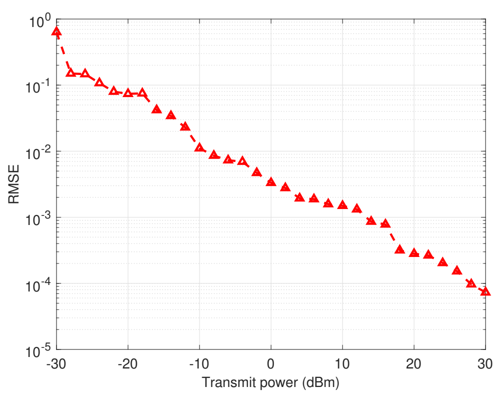

Fig. 4. RMSE of channel estimates versus user transmission power.

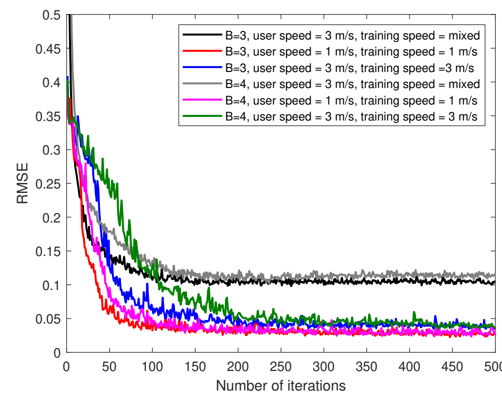

Fig. 5. RMSE of the predicted channels.

f_c, α, and r_0 are the carrier frequency, the path loss exponent, and the reference distance ($r_0 = 1$ m), respectively [29]. In this paper, $f_c = 28$ GHz is assumed.

The path loss exponents of the two paths between the IRS and the mobile user are set to 2.2 and 2.1, respectively. The training and testing data sets are generated based on the user's constant walking speed, which is assumed to be either 1 m/s or 3 m/s in two different scenarios. The length of each time frame is $T_n = 140$ ms, where the sensing time is 40 ms and that for data transmission is 100 ms.

We use the root-mean-square error (RMSE) to evaluate the channel estimation and prediction performance. The RMSE of the estimated channel for the nth time frame is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{Q}\sum_{q=1}^{Q} \|\hat{\mathbf{h}}_{q,n} - \mathbf{h}_{q,n}\|^2}, \tag{19}$$

where Q is the number of independent trials.

Fig. 4 shows the RMSE performance of the estimated user-IRS channel with respect to the transmit power of the mobile user, where $\dot{T} = 1,000$ samples are used. The channel noise is fixed at $\sigma_n^2 = -80$ dBm. The channel RMSE decreases as the transmit power increases. For channel prediction, $N_{\text{train}} = 11^3$ training data sets are generated based on different estimated user-IRS channels based on the user moving speeds, and the ADAM training algorithm is used with a batch size of 64. All channels fed into the training network are estimated in phase I. The user-transmitted power was 25 dBm, and the ReLU activation function is used in the hidden layers.

Fig. 5 shows the RMSE performance of the test datasets with respect to the number of iterations. The test results are based on three separately trained networks employing different user moving speeds (1 m/s and 3 m/s). The first network is trained using data generated with the user moving speed of 1 m/s and the second one is trained based on the user with a moving speed of 3 m/s. The third network is trained using mixed data sets of these two speeds. The RMSE of the predicted channels decreases with the number of iterations (epoch). When the user speed is fixed (1 m/s or 3 m/s), using three or four consecutive input channels ($B = 3$ or $B = 4$) provides similar performance. With higher user speeds, more iterations are needed to saturate predicted channel RMSE for input channels ($B = 3$ or $B = 4$). When trained with mixed-speed data, RMSE decreases faster, but the floor is higher with more iterations.

5. CONCLUSION

In this paper, we presented a novel approach to improve user quality of services through the design of an IRS setup that utilizes both estimated and predicted channels based on communication requirements. By training the neural network based on the CSI of the previous time frames, the channels are predicted to reduce the pilot overhead required for channel estimation. Additionally, structured matrix completion and pair-matching methods are applied to enable multipath channel detection with an L-shaped structured IRS, resulting in an increased array aperture and reduced computational overhead. Simulation results confirmed the effectiveness and performance of the proposed approach.

6. REFERENCES

[1] M. W. Akhtar, S. A. Hassan, R. Ghaffar, H. Jung, S. Garg, and M. S. Hossain, "The shift to 6G communications: Vision and requirements," *Human-Centric Comput. Info. Sci.*, vol. 10, no. 1, pp. 1–27, 2020.

[2] M. Xu, W. C. Ng, W. Y. B. Lim, J. Kang, Z. Xiong, D. Niyato, Q. Yang, X. S. Shen, and C. Miao, "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Commun. & Tut.*, vol. 25, no. 1, pp. 656–700, 2023.

[3] E. C. Strinati and S. Barbarossa, "6G networks: Beyond shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, p. 107930, 2021.

[4] M. Maier and A. Ebrahimzadeh, "Towards immersive tactile internet experiences: Low-latency FiWi enhanced mobile networks with edge intelligence," *J. Opt. Commun. Netw.*, vol. 11, no. 4, pp. B10–B25, 2019.

[5] A. M. Elbir, K. V. Mishra, M. Shankar, and S. Chatzinotas, "The rise of intelligent reflecting surfaces in integrated sensing and communications paradigms," *arXiv preprint arXiv:2204.07265*, 2022.

[6] H. Elayan, O. Amin, R. M. Shubair, and M.-S. Alouini, "Terahertz communication: The opportunities of wireless technology beyond 5G," in *Proc. IEEE Int. Conf. Adv. Commun. Technol. Netw. (CommNet)*, 2018, pp. 1–5.

[7] Z. Pi and F. Khan, "A millimeter-wave massive MIMO system for next generation mobile broadband," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2012, pp. 693–698.

[8] S. K. Sharma, I. Woungang, A. Anpalagan, and S. Chatzinotas, "Toward tactile internet in beyond 5G era: Recent advances, current issues, and future directions," *IEEE Access*, vol. 8, pp. 56 948–56 991, 2020.

[9] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, 2017.

[10] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2019.

[11] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.

[12] W. Long, R. Chen, M. Moretti, W. Zhang, and J. Li, "A promising technology for 6G wireless networks: Intelligent reflecting surface," *J. Commun. Info. Netw.*, vol. 6, no. 1, pp. 1–16, 2021.

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

[14] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, 2021.

[15] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1132–1142, 2021.

[16] M. Asif Haider, M. W. T. Chowdhury, and Y. D. Zhang, "Sparse channel estimation for IRS-aided systems exploiting 2-D sparse arrays," in *Proc. IEEE Sensor Array and Multich. Signal Process. Workshop (SAM)*, 2022, pp. 31–35.

[17] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, 2019.

[18] M. Hua, L. Yang, Q. Wu, C. Pan, C. Li, and A. L. Swindlehurst, "UAV-assisted intelligent reflecting surface symbiotic radio system," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5769–5785, 2021.

[19] X. Song, D. Zhao, H. Hua, T. X. Han, X. Yang, and J. Xu, "Joint transmit and reflective beamforming for IRS-assisted integrated sensing and communication," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2022, pp. 189–194.

[20] W. Jiang and H. D. Schotten, "Deep learning for fading channel prediction," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 320–332, 2020.

[21] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, 2020.

[22] C. Li, J. Zhang, S. Song, and K. B. Letaief, "Selective uplink training for massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2016, pp. 1–6.

[23] F. Wu, F. Cao, X. Ni, C. Chen, Y. Zhang, and J. Xu, "L-shaped sparse array structure for 2-D DOA estimation," *IEEE Access*, vol. 8, pp. 140 030–140 037, 2020.

[24] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1377–1390, 2015.

[25] A. Ahmed and Y. D. Zhang, "Generalized non-redundant sparse array designs," *IEEE Trans. Signal Process.*, vol. 69, pp. 4580–4594, 2021.

[26] C. Zhou, Y. Gu, Z. Shi, and Y. D. Zhang, "Off-grid direction-of-arrival estimation using coprime array interpolation," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1710–1714, 2018.

[27] M. W. T. Chowdhury and Y. D. Zhang, "Direction-of-arrival estimation exploiting distributed sparse arrays," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2021, pp. 231–235.

[28] M. Asif Haider, Y. D. Zhang, and E. Aboutanios, "ISAC system assisted by RIS with sparse active elements," *EURASIP J. Adv. Signal Process.*, vol. 2023, no. 20, pp. 1–22, 2023.

[29] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, 2015.