Multi-Task Learning-Based Channel Estimation for RIS Assisted Multi-User Communication Systems

Wenwu Xie, Jian Xiao, Peng Zhu, and Chao Yu

***Abstract*— In this letter, we propose a multi-task learning (MTL)-based joint channel estimation scheme for reconfigurable intelligent surface (RIS) assisted millimeter-wave communication system, where the direct channel (DC) and cascaded channel (CC) are estimated at the same coherence time by learning the feature of shared pilots. Since the dimension of CC is much larger than the DC, we design a learnable joint loss function based on homoscedastic task uncertainty to balance the training of two subtasks. Meanwhile, the residual shrinkage blocks are introduced into the multi-task network architecture to release the noise effect. Simulation results show that the estimation accuracy of MTL with less pilot overhead outperforms conventional channel estimation scheme, and significantly reduces training overhead compared with the single-task network.**

***Index Terms*— Reconfigurable intelligent surface, channel estimation, multi-task learning.**

I. Introduction

RECONFIGURABLE intelligent surfaces (RISs) can impose the required phase shift on the incident electromagnetic wave through a planar surface with a number of metamaterial units, which enhance the communication coverage and capacity with low deployment costs [1], especially for millimeter-wave (mmWave) communication with significant path loss. The effective design of passive beamforming depends on whether the base station (BS) or user can acquire accurate channel state information (CSI). Since the passive RISs are not equipped with radio frequency (RF) chains, the channel can only be sensed at the terminals. However, the dimension of cascaded channel (CC) is much large for the RIS with hundreds of elements, which increase the difficulty of channel estimation compared with conventional communication system [2].

In [3], classic least square (LS) algorithm was applied to CC estimation based on the element-by-element ON/OFF protocol. In this protocol, only one RIS element reflect the pilot signal at the single time slot, which reduced the accuracy of channel estimation. In [4], a efficient transmission protocol was designed for RIS-aided orthogonal frequency division multiplexing communication systems, where the DFT-based reflection pattern was firstly used to the joint channel estimation, and the optimality analysis of DFT protocol was deduced in [5]. Reference [6] proposed a PARAllel FACtor (PARAFAC) decomposition-based CC estimation scheme, where two iterative estimation algorithms were designed to estimate the BS-RIS channel and RIS-users channel. Considering the sparsity in RIS-aided mmWave communication, compressed sensing (CS) was widely used to the channel estimation for reducing pilot overhead [7], [8]. In [7], a sparse representation of CC was deduced by utilizing properties of KatriRao and Kronecker products. By exploiting the common sparsity of angular CC for multiple users, [8] proposed the double-structured orthogonal matching pursuit (DS-OMP)-based channel estimation scheme.

Deep learning (DL) technology has shown great potential to further improve the performance for RIS-aided communication system [9], [10]. In [11], LS was used to estimate initial channel matrix with noise firstly, and then deep residual network was adopted to obtain the channel matrix after denoising. Similarly, [12] combined the CS algorithm and deep complex-valued denoising network to reconstruct a CC from the limited measured data, but few RF chains need to be equipped to the RIS. In [13], channel estimation was regarded as super-resolution reconstruction in computer vision and only considered the channel from transmitter to the RIS. In the semi-passive channel estimation with active RIS elements [12], [13], cable or power supply must be configured, which limited the application of RIS. The design idea of [14] was similar with conventional two-stage channel estimation schemes, where two independent neural networks (NNs) were used to realize DC estimation and CC estimation by turning off/on all RIS elements alternately. Since two single-task learning (STL)-based networks were independent trained and deployed, the training overhead was large for communication device, including memory and computation complexity.

Considering the redundant training overhead of STL-based channel estimation, we develop a joint channel estimation scheme based on multi-task learning (MTL) [15], where DC and CC are jointly estimated by utilizing the shared pilot signal. In the proposed MTL network, the weights and parameters at the bottom layers are shared to learn the common knowledge and representation, while two independent and paralleled network layers are designed to complete both DC and CC estimation tasks at the top layers. To suppress the

Manuscript received October 28, 2021; revised November 26, 2021; accepted December 20, 2021. Date of publication December 23, 2021; date of current version March 10, 2022. This work was supported by the Open Fund of Advanced Cryptography and System Security Key Laboratory of Sichuan Province under Grant SKLACSS-202115, the Natural Science Foundation of Hunan Province under Grant 2021JJ40228, Grant 2020JJ4341, the Key Projects of Hunan Provincial Department of Education Department under Grant 21A0408, and the Outstanding Youth Project of Hunan Provincial Education Department under Grant 20B267, Grant 20B269. The associate editor coordinating the review of this letter and approving it for publication was C. Huang. *(Corresponding author: Chao Yu.)*

Wenwu Xie is with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China, and also with the Advanced Cryptography and System Security Key Laboratory, Chengdu 610103, China (e-mail: gavinxie@hnist.edu.cn).

Jian Xiao, Peng Zhu, and Chao Yu are with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China (e-mail: victor@vip.hnist.edu.cn; jianchongren@sina.com; ycyuchao_001@hotmail.com).

Digital Object Identifier 10.1109/LCOMM.2021.3138082

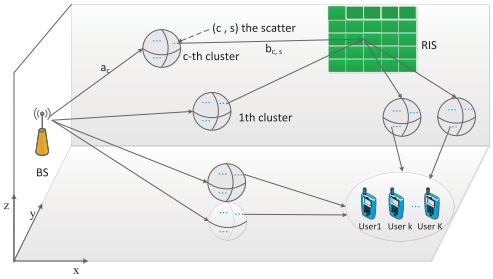

Fig. 1. Generic RIS assisted communication environment with random clusters.

noise effect of communication system, deep residual shrinkage network (DRSN) architectures are introduced into the shared bottom layers [16], which improve the estimation accuracy at low signal-to-noise ratio (SNR).

II. System Model and Problem Formulation

We consider a RIS-aided downlink mmWave communication system, where the BS and the RIS respectively are deployed with the M uniform planar array (UPA) antennas and the N UPA elements to communication with K single-antenna users. Following the 3GPP standardization [17] and 5G mmWave physical channel modeling [18], the cluster statistical MIMO channel model is used for RIS-aided mm-Wave communications [19]. Fig.1 shows the three-dimensional (3D) geometry model, where BS and RIS lie on the yz plane and xz plane, respectively. We assume all scatters between BS(RIS) and users (RIS) are grouped under C clusters, each having $S_c(c = 1, 2, \ldots, C)$ sub-rays. a_c and $b_{c,s}$ denote the distance between the BS and the cluster c and the distance between the $(c,\ s)$th scatterer and the RIS, respectively. The BS-RIS channel $\boldsymbol{G}$ can be modeled as

$$\begin{aligned}\boldsymbol{G} &= \underbrace{\sqrt{G_e\left(\theta_{\mathrm{LOS}}^{G_r}\right) L_{\mathrm{LOS}}^{G_r}}\boldsymbol{a}\left(\phi_{\mathrm{LOS}}^{G_r}, \theta_{\mathrm{LOS}}^{G_r}\right)\boldsymbol{a}^T\left(\phi_{\mathrm{LOS}}^{G_t}, \theta_{\mathrm{LOS}}^{G_t}\right)}_{\boldsymbol{G}_{\mathrm{LOS}}} \\ &+ \underbrace{\bar{\gamma}\sum_{c=1}^{\bar{C}}\sum_{s=1}^{\bar{S}_c}\bar{\beta}_{c,s}\sqrt{G_e\left(\theta_{c,s}^{G_r}\right) L_{c,s}^{G_r}}\boldsymbol{a}\left(\phi_{c,s}^{G_r}, \theta_{c,s}^{G_r}\right)\boldsymbol{a}^T\left(\phi_{c,s}^{G_t}, \theta_{c,s}^{G_t}\right)}_{\boldsymbol{G}_{\mathrm{NLOS}}}\end{aligned} \tag{1}$$

where $\boldsymbol{G}$ be decomposed into two line of sight (LOS) component $\boldsymbol{G}_{\mathrm{LOS}}$ and non-LOS (NLOS) component $\boldsymbol{G}_{\mathrm{NLOS}}$. Due to the space limitation, the parameters of $\boldsymbol{G}_{\mathrm{NLOS}}$ will be specified and $\boldsymbol{G}_{\mathrm{LOS}}$ is similar. $\bar{C}$ and $\bar{S}$ denote the number of clusters and sub-ray per cluster for the transmission link from the BS to RIS, respectively. $\bar{\gamma} = \sqrt{\frac{1}{\sum_{c=1}^{\bar{C}}\bar{S}_c}}$ represents a normalization factor. $\bar{\beta}_{c,s} \sim \mathcal{CN}(0,1)$ denotes the complex path gain. $L_{c,s}^{G_r}$ represents the fading associated with the (c,s)th path. $\phi_{c,s}^{G_r}$ $(\theta_{c,s}^{G_r})$ and $\phi_{c,s}^{G_t}$ $(\theta_{c,s}^{G_t})$ represent the azimuth (elevation) angle at the RIS, and the azimuth (elevation) angle at the BS for the (c,s)th path, respectively. The azimuth departure angles $(\phi_{c,s}^{G_t}, s = 1, \ldots, \bar{S}_c)$ follows conditionally Laplacian distribution $\phi_{c,s}^{G_t} \sim \mathcal{L}\left(\phi_c^{G_t}, \sigma_\phi\right)$, where $\phi_c^{G_t}$ follows a uniform distribution $\phi_c^{G_t} \sim \mathcal{U}[-\pi/2, \pi/2]$ and σ_ϕ denotes a constant angular spread [18]. Similarly, the elevation departure angles $\theta_{c,s}^{G_t} \sim \mathcal{L}\left(\theta_c^{G_t}, \sigma_\theta\right)$, where $\theta_{c,s}^{G_t} \sim \mathcal{U}[-\pi/4, \pi/4]$ [18] and σ_θ denotes angular spread. $\boldsymbol{a}\left(\phi, \theta\right)$ represents the UPA array response.

$$\begin{aligned}\boldsymbol{a}\left(\phi,\theta\right) = \Big[&1 \ldots e^{j2\pi d(x sin\theta + y\sin\phi\cos\theta)/\lambda} \ldots \\ &e^{j2\pi d((\sqrt{N}-1)sin\theta + (\sqrt{N}-1)\sin\phi\cos\theta)/\lambda}\Big]\end{aligned} \tag{2}$$

where $0 \le x \le \sqrt{N}-1$ and $0 \le y \le \sqrt{N}-1$. λ represents the carrier wavelength and d denotes the antenna spacing.

$G_e\left(\theta_{c,s}^{G_r}\right)$ represents the radiation of RIS elements for the (c,s)th scatterer.

$$G_e\left(\theta_{c,s}^{G_r}\right) = 2(2q+1)\cos^{2q}\left(\theta_{c,s}^{G_r}\right) \tag{3}$$

where $\cos^q$ pattern is used for reflectarrays and q determines the element gain [20].

We adopt the generic 5G path loss model for $L_{c,s}^{G_r}$ [21].

$$\begin{aligned}L_{c,s}^{G_r} = -20\log_{10}\left(\frac{4\pi}{\lambda}\right) - 10n\left(1 + b\left(\frac{f-f_0}{f_0}\right)\right) \\ \times \log_{10}\left(d_{c,s}\right) - X_{\sigma_x}\end{aligned} \tag{4}$$

where $d_{c,s} = a_c + b_{c,s}$ represents the ray length of the $(c,\ s)$th path. n denotes the path loss exponent. b and f_0 denote a model parameter and reference frequency, respectively. $X_{\sigma_x} \sim \mathcal{CN}(0, {\sigma_x}^2)$ represents a shadow factor.

Similarly, the RIS-user k channel $\boldsymbol{h_k}$ can be represented as

$$\begin{aligned}\boldsymbol{h_k} = &\underbrace{\sqrt{G_e\left(\theta_{\mathrm{LOS}}^{r,k}\right) L_{\mathrm{LOS}}^{r,k}}\boldsymbol{a}\left(\phi_{\mathrm{LOS}}^{r,k}, \theta_{\mathrm{LOS}}^{r,k}\right)}_{\boldsymbol{h}_{\mathrm{LOS}}} \\ &+ \underbrace{\tilde{\gamma}\sum_{c=1}^{\hat{C}}\sum_{s=1}^{\hat{S}_c}\hat{\beta}_{c,s}\sqrt{G_e\left(\theta_{c,s}^{r,k}\right) L_{c,s}^{r,k}}\boldsymbol{a}\left(\phi_{c,s}^{r,k}, \theta_{c,s}^{r,k}\right)}_{\boldsymbol{h}_{\mathrm{NLOS}}}\end{aligned} \tag{5}$$

where $\hat{C}$ and $\hat{S}$ represent the number of clusters and sub-ray per cluster for the transmission link from the RIS to user, respectively. $\hat{\gamma} = \sqrt{\frac{1}{\sum_{c=1}^{\hat{C}}\hat{S}_c}}$. $\hat{\beta}_{c,s} \sim \mathcal{CN}(0,1)$. $G_e\left(\theta_{c,s}^{r,k}\right)$ represents the RIS element gain. $L_{c,s}^{G_{r,k}}$ denotes the path loss. $\phi_{c,s}^{r,k}$ $(\theta_{c,s}^{r,k})$ is the azimuth (elevation) angle at the RIS.

The BS-user k channel $\boldsymbol{d_k}$ can be given by

$$\begin{aligned}\boldsymbol{d_k} = &\underbrace{\sqrt{L_{\mathrm{LOS}}^{d_t,k}}\boldsymbol{a}\left(\phi_{\mathrm{LOS}}^{d_t,k}, \theta_{\mathrm{LOS}}^{d_t,k}\right)}_{\boldsymbol{d}_{\mathrm{LOS}}} \\ &+ \underbrace{\tilde{\gamma}\sum_{c=1}^{\tilde{C}}\sum_{s=1}^{\tilde{S}_c}\tilde{\beta}_{c,s}\sqrt{L_{c,s}^{d_t,k}}\boldsymbol{a}\left(\phi_{c,s}^{d_t,k}, \theta_{c,s}^{d_t,k}\right)}_{\boldsymbol{d}_{\mathrm{NLOS}}}\end{aligned} \tag{6}$$

where $\tilde{C}$ and $\tilde{S}$ denote the number of clusters and sub-ray per cluster for the transmission link from the BS to user, respectively. $\tilde{\gamma} = \sqrt{\frac{1}{\sum_{c=1}^{\tilde{C}}\tilde{S}_c}}$. $\tilde{\beta}_{c,s} \sim \mathcal{CN}(0,1)$. $L_{c,s}^{d_t,k}$ represents the path loss. $\phi_{c,s}^{d_t,k}$ $(\theta_{c,s}^{d_t,k})$ denotes the azimuth (elevation) angle at the BS.

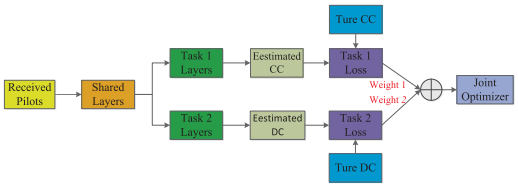

Fig. 2. The low-level shared multi-task learning frame.

Let $\boldsymbol{\theta} = [e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}]^T \in \mathbb{C}^N$ denotes the reflecting vector at the RIS, where $\theta_i (i = 1, 2, \ldots, N)$ represents the phase shift at i-th RIS element. Following [3], [5], we consider the channel reciprocity under time-division duplexing mode, where the downlink channel can be obtained by estimating uplink channel. The received pilot signal $\boldsymbol{y}_{k,t} (t = 1, 2, \ldots, T)$ in the t-th time slot at the BS for the k-th user can be expressed as

$$\begin{aligned} \boldsymbol{y}_{k,t} &= (\boldsymbol{G}^T \mathrm{diag}(\boldsymbol{\theta}_t) \boldsymbol{h}_k^T + \boldsymbol{d}_k^T) s_{k,t} + \boldsymbol{n}_{k,t} \\ &= (\boldsymbol{\theta}_t \boldsymbol{G}^T \mathrm{diag}(\boldsymbol{h}_k^T) + \boldsymbol{d}_k^T) s_{k,t} + \boldsymbol{n}_{k,t} \end{aligned} \tag{7}$$

where $s_{k,t}$ is the transmitted pilot signals at the t-th time slot. $\boldsymbol{n}_{k,t} \in \mathbb{C}^M$ is white Gaussian noise. $\boldsymbol{H}_k = \boldsymbol{G}^T \mathrm{diag}(\boldsymbol{h}_k^T) \in \mathbb{C}^{M \times N}$ is defined as the CC.

III. Proposed Method

A. Dataset Construction

When we keep RIS elements active, the received signal of the BS include the direct signal and reflecting signal according to (7). From the data-driven perspective, we can design a end-to-end MTL model to realize the estimation of DC and CC at once, which can avoid additional training overhead for direct channels [4], [5]. Consequently, the total overhead of joint channel estimation is equal to the pilot overhead of CC estimation. Following the reflecting protocol of CC estimation in [3], we generated the received pilot signal $\boldsymbol{Y}_k \in \mathbb{C}^{M \times N}$ as the input data $\boldsymbol{I}$, where $\boldsymbol{I} = [\mathrm{Re}(\boldsymbol{Y}_k), \mathrm{Im}(\boldsymbol{Y}_k)] \in \mathbb{R}^{M \times N \times 2}$ represents the real and imaginary part of $\boldsymbol{Y}_k$. Similarly, the label data $\boldsymbol{O}_d = [\mathrm{Re}(\boldsymbol{d}_k^T), \mathrm{Im}(\boldsymbol{d}_k^T)] \in \mathbb{R}^{M \times 2}$ and $\boldsymbol{O}_H = [\mathrm{Re}(\boldsymbol{H}_k^T), \mathrm{Im}(\boldsymbol{H}_k^T)] \in \mathbb{R}^{M \times N \times 2}$ represent the real and imaginary part of the DC and CC, respectively. Meanwhile, the received signal and the channel data are normalized by the maximum absolute value of their elements.

B. Multi-Task Learning

Fig.2 shows a low-level shared MTL frame [15], where Task 1 layers and Task 2 layers are used to realize the CC and DC estimation, respectively. Since the dimension of DC is usually lower than CC, the difficulty of two channel estimation tasks is also different for NN. Conventional methods use a fixed uniform weight to combine a joint loss function, which is unfair for CC estimation task. We combine different loss functions of subtasks based on homoscedastic task uncertainty [22], which can automatic learn the optimal weight by NN. The joint loss function of MTL can be represented as

$$L_{\mathrm{jo\,int}}(w, \sigma_1, \sigma_2) \approx \frac{1}{\sigma_1^2} L_1(w) + \frac{1}{\sigma_2^2} L_2(w) + \log \sigma_1^2 + \log \sigma_2^2 \tag{8}$$

where w is the weight parameter of the network. $\sigma_i^2 (i = 1, 2)$ is a observation noise scalar of the i-th task and is a learnable parameter by the NN. The loss function of the i-th task $L_i(w)$ can be represented as

$$L_i(w) = |a_i - {f_i}^w(x)| \tag{9}$$

where a_i is the true channel for i-th subtask, and ${f_i}^w(x)$ is the output of the i-th subtask when x is the input data. In (9), large σ_i^2 will decrease the contribution of $L_i(w)$, while smaller σ_i^2 will increase its contribution. The whole objective function is penalized by $\log \sigma_i^2$ and will be a constant when σ_i^2 is too large.

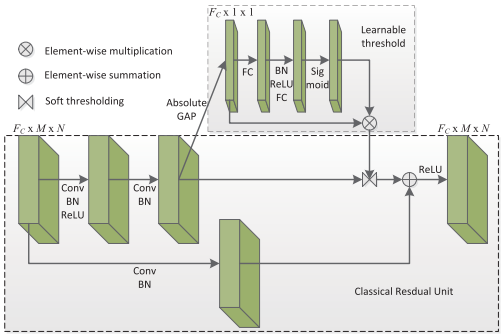

Fig. 3. The architecture of DRSN-CS block.

TABLE I
THE MULTI-TASK NETWORK WHEN $M = 64$ AND $N = 256$

Shared layers	Task 1 layers	Task 2 layers
CBR(64, 3)	CBR(64, 3)	CBR(64, 3, [1,4])
DRSN-CS(64, 3)	Conv(2, 3),Tanh	CBR(64, 3, [1,4])
DRSN-CS(64, 3)		CBR(64, 3, [1,4])
		Conv(2, 3, [1,4]),Tanh

C. Multi-Task Network Architecture

In the channel estimation of communication system, the estimation accuracy is easily affected by noise. Soft thresholding (ST) is the key step in signal denoising algorithm and can be expressed as

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau \le x \le \tau \\ x + \tau & x < -\tau \end{cases} \tag{10}$$

where x and y represent the input and output of ST, respectively. τ is the threshold.

It is a challenging problem to select an suitable threshold τ in traditional algorithms. Fig.3 shows the DRSN with channel-wise thresholds (DRSN-CS) block, where F_c denotes the number of channel for feature map. Motivated by attention

mechanism in DL, ST is inserted as a specialized network layers into the classic residual unit [23], which can automatically learn the τ in the DRSN-CS block.

In the shared layers of MTL model, a baseline convolutional block, composed of convolutional layer, batch normalization (BN) layer and LeakyReLU activation layer (CBR), is used to extract the low-level feature. Two DRSN-CS blocks are stacked to learn the deep features and further ease the influence of noise. The CC estimation is completed by the Task 1 layers, where a CRR convolutional layer is used to extract the CC feature map and a convolutional layer is designed to reduce the channel dimension of the output, respectively. Since the dimension of DC is lower than the dimension of the received signal, convolutional layers with $(F_{\mathrm{d1}}, F_{\mathrm{d2}})$ strides are used to reduce the size of feature map in the Task 2 layers, where the number of convolutional layers is $L_{\mathrm{d}} = \log_{F_{\mathrm{d1}}}^{N}$. In the output of network, the Tanh activation is used to constraint the range of output data. Table I summarized the whole network architecture of MTL model. The first number in the bracket is the number of convolutional kernels F_{c}, and the second number represents the size of convolutional kernel $(F_{\mathrm{s}}^1, F_{\mathrm{s}}^2)$, where the shape of convolutional kernel is square in our MTL model, i.e. $F_{\mathrm{s}} = F_{\mathrm{s}}^1 = F_{\mathrm{s}}^2 = 3$. The third number in some brackets represents $(F_{\mathrm{d1}}, F_{\mathrm{d2}}) = (1, 4)$.

IV. Numerical Simulation

In the simulation, we set $M = 8 \times 8$, $N = 16 \times 16$ and $K = 6$. The mmWave communication frequency band is set as 28 GHz and 73 GHz. $C(\bar{C}, \hat{C}, \tilde{C})$ is modeled by Poisson distribution $C \sim \max\{P(\lambda_p), 1\}$, where $\lambda_p = 1.8$ in 28 GHz and $\lambda_p = 1.9$ in 73 GHz [18]. $S_c(\bar{S}_c, \hat{S}_c, \tilde{S}_c) \sim \mathcal{U}[1, 30]$ follows uniform distribution between 1 and 30 [24]. We consider the path loss model of Urban Microcellular environment, where $n = 3.19$, $b = 0$ and $\sigma = 8.2$ dB for NLOS component, while $n = 3.19$ and $\sigma = 3.1$ dB for LOS component [21]. We generated 5×10^3 paired samples for each user. Consequently, the number of total sample is 3×10^4 and are randomly divided into training, test and validation sets by the ratios of 60%, 20%, and 20%. We used the normalized mean squared error (NMSE) as the performance evaluations metric, i.e., $\mathrm{NMSE}_{H_k} = \mathbb{E}[||\hat{\boldsymbol{H}}_k - \boldsymbol{H}_k||_F^2 / ||\boldsymbol{H}_k||_F^2]$.

The NMSE performance of the MTL-based channel estimation scheme is compared with the LS [3]–[5], LMMSE, orthogonal matching pursuit (OMP) [8] and STL [14]. Table. II summarized the model size, parameters, the number of addition and multiplication (Mult-Adds) and pilot overhead of STL and MTL. In Fig.4 and Fig.5, the "$-$ON/OFF" and"$-$DFT" denote that the reflecting vector Θ follows ON/OFF protocol and DFT protocol, respectively. STL-OFF denotes the RIS is turned off when we estimate DC [14].

Fig.4 and Fig.5 show the NMSE performance of different algorithms for DC and CC estimation, respectively, where the estimation accuracy of MTL is superior to other schemes in the most of SNR ranges. The channel estimation performance depends on the RIS reflection protocol and estimator in RIS-aided communication system. Although the DFT protocol has been proved to be the optimal reflection protocol, the

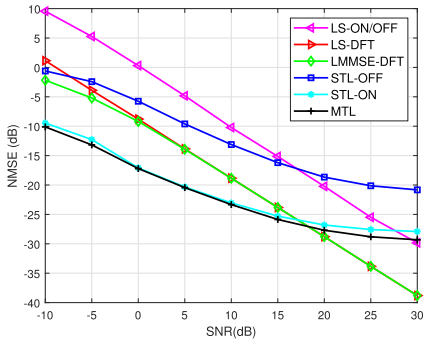

Fig. 4. The NMSE performance of DC estimation.

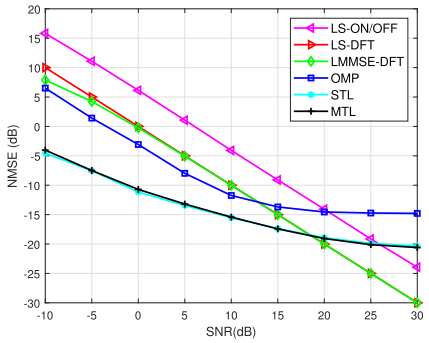

Fig. 5. The NMSE performance of CC estimation.

conventional LS and LMMSE estimator are sub-optimal. With the increase of SNR, the performance gap between MTL and conventional schemes is closing. The channel estimation can be regarded as a high-dimensional regression task, while data-driven neural network can be regarded as a universal approximator. When the noise is small enough, e.g., SNR $\geq$ 20 dB, the approximate error of neural network is larger than the inverse of noise in LS algorithm.

Although the training data I_{DC} of STL-OFF avoid the interference of CC in the DC estimation, the data size $I_{\mathrm{DC}} \in \mathbb{R}^{M \times 2}$ of STL-OFF is less than $I \in \mathbb{R}^{M \times N \times 2}$ of STL-ON and MTL. The performance improvement of DL depends on more sufficient data. However, larger dimension of input data bring more computation overhead, so the Mult-Adds of STL-ON is much larger than STL-OFF in Table II. The NMSE of MTL is similar with STL, while STL needs two independent networks to estimate DC and CC, respectively. Consequently, the total parameters and computation complexity of STL are larger than MTL. Since the sparsity is variable and relatively large in the clustered statistical MIMO channel model, the estimation performance of OMP is limited.

In Fig.6, the NMSE performance has been compared for different network architectures in shared layers. The CNN and ResNet have the similar estimation accuracy, while DRSN architecture can achieve higher estimation accuracy under lower SNR. The DRSN introduced the branch of learnable ST in classic residual unit, which can degrade the effect of noise on channel estimation. According to the cluster MIMO channel model and 5G path loss model, the change of communication frequency band will affects the distribution of cluster C and

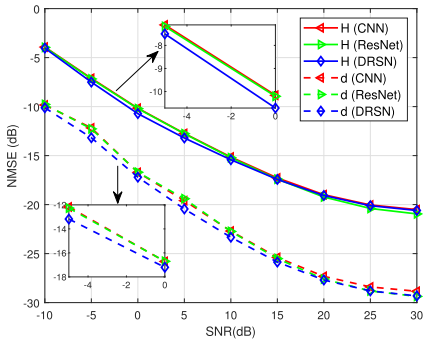

Fig. 6. The NMSE performance of different network architectures.

TABLE II
THE TRAINING OVERHEAD FOR DIFFERENT NETWORKS

	MTL	**STL-ON**	**STL-OFF**
Model size(MB)	1.21	0.78+1.06	0.78×2
Parameters(K)	316.55	204.24+278.36	204.24×2
Mult-Adds(G)	3.28	3.06+2.63	3.06+0.012
Pilot overhead	N	N	$N+1$

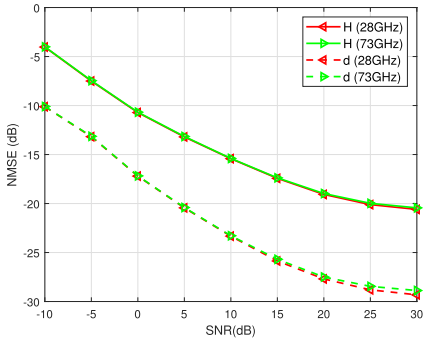

Fig. 7. The NMSE performance under different frequency bands.

the path loss $L_{c,s}$. Fig.7 shows the robustness of MTL model for different frequency bands, where the trained MTL model under 28 GHz can work well on 73 GHz.

V. CONCLUSION

To improve the estimation accuracy of conventional algorithms and reduce the training overhead of STL, we proposed a MTL-based channel estimation scheme for RIS-aided mmWave multi-user communication system, which can realize DC and CC estimation at the same time slots by the end-to-end learning. The MTL can achieve similar NMSE performance as STL with less training overhead. Meanwhile, the robustness of MTL model has been verified in different frequency bands. In the future work, we will extend the MTL model to the receiver with multiple antennas communication scenario.

REFERENCES

[1] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[2] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part I: Fundamentals, solutions, and future opportunities," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1398–1402, May 2021.

[3] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 4659–4663.

[4] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[5] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2020, pp. 5000–5004.

[6] 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 Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.

[7] 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.

[8] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.

[9] B. Yang, X. Cao, C. Huang, C. Yuen, L. Qian, and M. D. Renzo, "Intelligent spectrum learning for wireless networks with reconfigurable intelligent surfaces," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3920–3925, Apr. 2021.

[10] X. Cao *et al.*, "Reconfigurable intelligent surface-assisted aerial-terrestrial communications via multi-task learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3035–3050, Oct. 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," *IEEE Trans. Wireless Commun.*, early access, Aug. 3, 2021, doi: 10.1109/TWC.2021.3100148.

[12] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.

[13] 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.

[14] 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 Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, Sep. 2020.

[15] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.

[16] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.

[17] *3GPP TR 38.901 V16.1.0—Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document ETSI TR 138 901-2018, Dec. 2019.

[18] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Quart., 2018.

[19] E. Basar and I. Yildirim, "Reconfigurable intelligent surfaces for future wireless networks: A channel modeling perspective," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 108–114, Jun. 2021.

[20] P. Nayeri, F. Yang, and A. Z. Elsherbeni, *Reflectarray Antennas: Theory, Designs, and Applications*. Hoboken, NJ, USA: Wiley, 2018.

[21] N. Docomo. (2016). *White on Paper 5G Channel Model for Bands up to 100 GHz*. [Online]. Available: http://www.5gworkshops.com/5GCMSIGWhite

[22] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[24] M. K. Samimi and T. S. Rappaport, "Statistical channel model with multi-frequency and arbitrary antenna beamwidth for millimeter-wave outdoor communications," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–7.