Multi-Task Learning Based Channel Estimation for Hybrid-Field STAR-RIS Systems

Jian Xiao*, Ji Wang*, Yuanwei Liu†, Wenwu Xie‡, Jun Wang§ and Shouyin Liu*
*Department of Electronics and Information Engineering, Central China Normal University, Wuhan, China
†School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K.
‡School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China
§CICT Mobile Communication Technology Co., Ltd., Wuhan, China
Email: jianx@mails.ccnu.edu.cn, {jiwang, syliu}@ccnu.edu.cn, yuanwei.liu@qmul.ac.uk, jwang@cictmobile.com

GLOBECOM 2023 - 2023 IEEE Global Communications Conference | 979-8-3503-1090-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/GLOBECOM54140.2023.10437505

***Abstract*—A joint cascaded channel estimation framework is proposed for simultaneously transmitting and reflecting reconfigurable intelligent surfaces (STAR-RIS) systems with hardware imperfection, in which practical the hybrid-field electromagnetic wave radiation with spatial non-stationarity is investigated. By exploiting the cascaded channel correlations in user domain and STAR-RIS element domain, we propose a multi-task network (MTN) with multi-expert branches to simultaneously reconstruct the high-dimensional transmitting and reflecting channels from the observed mixture channel with noise. In the proposed MTN architecture, a learnable shrinkage module is exploited to constrict the communication noise, and self-attention mechanism-based Transformer layers are utilized to extract the non-local feature of the non-stationary cascaded channel. Numerical results show that the proposed MTN achieves superior channel estimation accuracy with less training overhead compared with existing state-of-the-art benchmarks, in terms of required pilots, computations, and network parameters.**

I. Introduction

Metasurface-based communication paradigm has been regarded as a promising multiple input multiple output (MIMO) candidate to construct *smart radio environments* (SREs) in the sixth-generation wireless networks [1], i.e., reconfigurable intelligent surface (RIS) enabled extremely large-scale antenna array (ELAA) communications. The typical reflection-only RISs only reflect the incident signal to desired user equipments at the same side (referred to as UE^r), which only forms a *half-space* SRE. To break the limitation of reflection-only RISs and achieve the *full-space* SREs, the novel concept of *simultaneously transmitting and reflecting* RISs (STAR-RISs) has attracted increasing attention [2]. The signal imping on the STAR-RIS is divided into two parts with the law of energy conservation. One part electromagnetic wave is reflected to the UE^r at the same side as the incident wave, while the other part is transmitted to the users at the opposite side (referred to as UE^t). The dual functionality of STAR-RISs provides greater potentiality to extend the wireless signal coverage [2], [3].

The accurate channel estimation is vital to the RIS beamforming optimization, while it is also a crucial challenge due to the high-dimensionality caused by extensive passive RIS elements. In STAR-RIS systems, the channel estimation design is related to the dedicated operating protocol of STAR-RIS [4], i.e., time switching (TS) and energy splitting (ES) protocols. In the TS protocol, all elements of STAR-RIS are switched periodically between the transmitting and reflecting mode in orthogonal time slots, and hence the channel estimation in STAR-RIS systems is similar to that in reflecting-only RIS. In the ES protocol, the incident signal on each element of the STAR-RIS can be reflected and transmitted with an ES ratio at the same time slots, which can provide higher communication degree of freedom. Since the ES strategy reduces the received signal strength at UE^f ($\forall f \in \{t, r\}$), the channel estimation accuracy is significantly decreased than the TS protocol [5].

As the number of STAR-RIS elements grows large in ELAA systems, the far-field radiation assumptions are no longer valid for ELAA systems, while the near-field propagation is likely to happen due to the increase of array aperture [6]. In near-field communications, more complex channel characteristics need to be studied compared with the far-field channel, e.g., the spherical wavefront, variations angle of arrival/departure (AoA/AoD) across array elements, and spatial non-stationarity caused by visibility regions (VRs) [7]. Besides, a practical case of radiation field will happen in ELAA systems, which is the hybrid far- and near-field (hybrid-field) communication. Note that the boundary of near-field region in RIS systems is more strict compared with conventional extremely large-scale MIMO (XL-MIMO) systems [6], and the hybrid-field radiation and spatial non-stationarity effect are also more complex [8].

Compared with the channel estimation in RIS systems, the design of channel estimation schemes in STAR-RIS systems is at a preliminary stage, especially for the hybrid-field communications. In [5], a least square (LS)-based channel estimation scheme was derived for STAR-RIS systems, which applies to both the TS and ES protocol. However, as a classic linear estimator, the performance of the LS estimation is limited under the non-linear noise, and the pilot overhead of LS is expensive for the extremely large-scale STAR-RIS. Specifically, the minimum pilot overhead is $K^p N$ in [5], where N and K^p denote the number of STAR-RIS elements and UEs for a paired user group, respectively. Besides, in hybrid-field STAR-RIS systems, the hybrid-field radiation and spatial non-stationarity of the cascaded channel restrict the efficient application of the existing compressed sensing (CS) algorithms [7], where the pure sparse representation is hard to obtain by designing a specific transform domain [8].

To reduce the training overhead and improve the channel

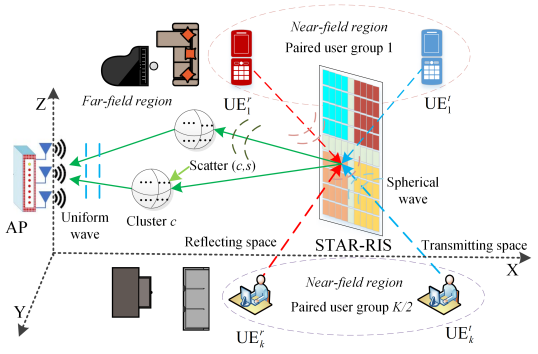

Fig. 1. STAR-RIS assisted indoor mmWave communications.

estimation accuracy for hybrid-field STAR-RIS systems, we propose a multi-task learning (MTL)-based joint cascaded channel estimation scheme. Firstly, we formulate the hybrid-field non-stationary channel modeling and the practical signal model with hardware imperfection. Then, we exploit an effective multi-task network (MTN) to estimate the transmitting and reflecting cascaded channel simultaneously, in which the required pilot overhead can be reduced to N/Γ and $\Gamma \geq 1$ is a sampling interval in STAR-RIS element domain. Moreover, the proposed MTN significantly reduces the network training overhead compared with the single-task learning (STL)-based channel estimation framework.

II. System Model and Problem Formulation

As shown in Fig. 1, a STAR-RIS with $N^s = N_1^s \times N_2^s$ uniform planar array (UPA) elements is deployed for indoor communications. The wireless access point (AP) with $M = M_1 \times M_2$ UPA antennas communicates K single-antenna UEs, aided by a STAR-RIS with the ES operation protocol. By adopting the typical elements-grouping strategy in RIS systems [5], [9], N^s elements is divided into $N = N_1 \times N_2$ sub-surfaces, each of which consists of $\nu = (N_1^s/N_1) \times (N_2^s/N_2)$ adjacent elements. We assume that K users are equally located in transmitting and reflecting space, and a UE_k^r and a $\mathrm{UE}_k^t, (k = 1, 2, \cdots, K/2)$ constitute a user group (UG).

To alleviate the severe multiplicative fading effect of the cascaded link, the STAR-RIS is deployed closed to UEs [10], and hence UEs are likely communicating in the near-field region of the STAR-RIS, which is determined by the Rayleigh distance Z. According to the near-field criterion in [6], the near-field region for metasurface-aided systems is given by

$$\frac{d_{c,s}^{\mathrm{R}} d_k^{\mathrm{UR}}}{d_{c,s}^{\mathrm{R}} + d_k^{\mathrm{UR}}} < Z = \frac{2D^2}{\lambda}, \tag{1}$$

where $d_{c,s}^{\mathrm{R}}$ and d_k^{UR} denote the distance from the STAR-RIS to the scatter (c, s) and the distance from the UE_k^f to the STAR-RIS, respectively. Parameter λ is the carrier wavelength and D is the equivalent array aperture of STAR-RIS systems. It can be further implied that as long as any of $d_{c,s}^{\mathrm{R}}$ and d_k^{UR} is shorter than the Rayleigh distance Z, the communication link is operating in the near-field region. On the other hand, from the perspective of AP with medium-size antennas, the environmental scatters are distributed in the far-field region of AP. Consequently, the far-field and near-field wireless signal will coexist in this system, which constitutes the hybrid-field STAR-RIS communications.

A. Hybrid-Field Channel Model

Following the clustered statistical MIMO channel modeling framework for millimeter-wave (mmWave) communications, the scatters are grouped into C_{s} clusters and each cluster is composed of $S_c, (c = 1, 2, \cdots, C_{\mathrm{s}})$ scatters. The STAR-RIS→AP channel $\mathbf{G} \in \mathbb{C}^{M \times N}$ can be expressed as

$$\mathbf{G} = \gamma \sum_{c=1}^{C_{\mathrm{s}}} \sum_{s=1}^{S_c} \varsigma_{c,s} \sqrt{R_{c,s}^{G_{\mathrm{r}}} L_{c,s}^{G_{\mathrm{r}}}} \mathbf{a}_{c,s} \mathbf{b}_{c,s}^T, \tag{2}$$

where $\gamma = \sqrt{\frac{1}{\sum_{c=1}^{C_{\mathrm{s}}} S_c}}$ is a normalization factor, $\varsigma_{c,s} \sim \mathcal{CN}(0, 1)$. Parameter $R_{c,s}$, and $L_{c,s}^{G_{\mathrm{r}}}$ are the complex gain, the STAR-RIS element pattern and the path loss for the scatter (c, s), respectively. $\mathbf{b} \in \mathbb{C}^{N \times 1}$ denotes the transmitting array response at the STAR-RIS, and $\mathbf{a} \in \mathbb{C}^{M \times 1}$ represents the receiving response at the AP. In conventional far-field radiation, the signals is approximated as uniform plane wave, and hence the array response only depends on the identical AoA/AoD. The receiving far-field response $\mathbf{a}$ can be represented as [10]

$$\mathbf{a}\left(\phi_{c,s}^{\mathrm{A}}, \varphi_{c,s}^{\mathrm{A}}\right) = \Big[1 \cdots e^{j2\pi d(x sin\varphi_{c,s}^{\mathrm{A}} + y \sin\phi_{c,s}^{\mathrm{A}} \cos\varphi_{c,s}^{\mathrm{A}})/\lambda} \cdots e^{j2\pi d((M_1-1) sin\varphi_{c,s}^{\mathrm{A}} + (M_2-1) \sin\phi_{c,s}^{\mathrm{A}} \cos\varphi_{c,s}^{\mathrm{A}})/\lambda}\Big], \tag{3}$$

where $0 \leq x \leq M_1 - 1$, $0 \leq y \leq M_2 - 1$, and d is the antenna spacing. $\phi_{c,s}^{\mathrm{A}}$ and $\varphi_{c,s}^{\mathrm{A}}$ denotes the azimuth and elevation of AoA for the (c, s)-th scatter path at the AP, respectively. In near-field communications, generic non-uniform spherical wave characteristics will be taken into account. The near-field array response $\mathbf{b}^n$ at the STAR-RIS can be expressed as [8]

$$\mathbf{b}^{\mathrm{n}}\left(d_{c,s}^{\mathrm{R}}\right) = \Big[e^{j2\pi d_{c,s}^{\mathrm{R}}(1,1)/\lambda}, \cdots, e^{j2\pi d_{c,s}^{\mathrm{R}}(1,N_2)/\lambda}, \cdots, e^{j2\pi d_{c,s}^{\mathrm{R}}(N_1,1)/\lambda}, \cdots, e^{j2\pi d_{c,s}^{\mathrm{R}}(N_1,N_2)/\lambda}\Big], \tag{4}$$

where $d_{c,s}^{\mathrm{R}}(n_1, n_2)$ denotes the distance from the scatter (c, s) to the (n_1, n_2)-th STAR-RIS element.

In ELAA systems, different parts of the STAR-RIS elements may view different scatters (terminals) due to the limitation of VRs, and hence the energy distribution across STAR-RIS elements is unequal. Specifically, we consider the cluster VR Ω_c [8] and user VR Ψ_k [7] for the STAR-RIS→scatters link and the UE_k→STAR-RIS link, respectively. The cluster VR Ω_c of STAR-RIS is identified by the center $(V_c^{\mathrm{x}}, V_c^{\mathrm{y}})$ and length $(V_l^{\mathrm{x}}, V_l^{\mathrm{y}})$ of Ω_c, in which the VR lengths V_l follows the Lognormal distribution $V_l \sim \mathcal{LN}(\mu_l, \sigma_l)$. The VR cover vector $v(\Omega_c) \in \mathbb{C}^{N \times 1}$ for the c-th cluster can be expressed as

$$[v(\Omega_c)]_n = \begin{cases} 1, & \text{if } n \in \Omega_c, \\ 0, & \text{else.} \end{cases} \tag{5}$$

Hence, the equivalent near-field array response with spatial non-stationarity is given by $\mathbf{b} = \mathbf{b}^{\mathrm{n}} \odot v(\Omega_c)$, in which $\odot$ represents the Hadamard product.

For the UE$_k$→STAR-RIS communication link, the receiving array response $\mathbf{u}_k$ at the STAR-RIS is related to the distance $d_k^{\mathrm{UR}}(n_1, n_2)$ from the UE$_k$ to the (n_1, n_2)-th STAR-RIS element. For the definition of user VR Ψ_k, we follow the line-of-sight (LOS) VR modeling method in [7]. The UE$_k$→STAR-RIS channel $\mathbf{h}_k$ can be represented as

$$\mathbf{h}_k = \sqrt{R_k^{\mathrm{h}} L_k^{\mathrm{h}}}\mathbf{u}_k \odot v(\Psi_k), \tag{6}$$

where R_k^{h} represents the radiation gain of STAR-RIS, L_k^{h} is the path loss, $v(\Psi_k) \in \mathbb{C}^{N\times 1}$ denotes the UE$_k$'s VR cover vector.

B. Problem Formulation

We focus on the UE$_k^f$ →STAR-RIS→AP$(\forall f \in \{t, r\})$ cascaded channel estimation, and the orthogonal pilot transmission strategy is adopted for different UGs [5]. Let $\boldsymbol{\theta}^t = [\beta_1^t e^{j\theta_1^t}, \beta_2^t e^{j\theta_2^t}, \cdots, \beta_N^t e^{j\theta_N^t}]^T \in \mathbb{C}^{N\times 1}$ and $\boldsymbol{\theta}^r = [\beta_1^r e^{j\theta_1^r}, \beta_2^r e^{j\theta_2^r}, \cdots, \beta_N^r e^{j\theta_N^r}]^T \in \mathbb{C}^{N\times 1}$ denote the transmitting and reflecting vectors, respectively, in which the ES ratio β_n satisfies $\beta_n^t + \beta_n^r \le 1, n = 1, 2, \cdots, N$. The received pilot signal $\mathbf{y}_{k,q} \in \mathbb{C}^{M\times 1}$ in the q-th time slot at the AP for the UG$_k$ can be expressed as

$$\begin{aligned}\mathbf{y}_{k,q} &= \mathbf{G}\left(\mathrm{diag}(\boldsymbol{\theta}_q^t)\mathbf{h}_k^t s_{k,q}^t + \mathrm{diag}(\boldsymbol{\theta}_q^r)\mathbf{h}_k^r s_{k,q}^r\right) + \mathbf{w}_q \\ &= \sum_{f=t}^{r} \mathbf{H}_k^f \boldsymbol{\theta}_q^f s_{k,q}^f + \mathbf{w}_q,\end{aligned} \tag{7}$$

where $\mathbf{h}_k^f \in \mathbb{C}^{N\times 1}(\forall f \in \{t, r\})$ represents the UE$_k^f$→STAR-RIS channel, while $\mathbf{H}_k^f = \mathbf{G}\mathrm{diag}(\mathbf{h}_k^f) \in \mathbb{C}^{M\times N}$ is defined as the cascaded channel. $s_{k,q}^f$ denotes the transmitted pilot signal at UE$_k^f$. $\mathbf{w}_q \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_M)$ is complex Gaussian noise.

In this work, a practical STAR-RIS systems with hardware imperfection is considered. Firstly, the coupled phase-shifts θ_n^t and θ_n^r model for passive STAR-RIS hardware is given by

$$\cos(\theta_n^t - \theta_n^r) = 0, n = 1, 2, \cdots, N. \tag{8}$$

Then, the residual hardware impairments (HWIs) at the AP and the UE are modeled by the additive Gaussian distribution. Hence, we rewrite (7) as

$$\tilde{\mathbf{y}}_{k,q} = \sum_{f=t}^{r} \mathbf{H}_k^f \boldsymbol{\theta}_q^f p_k^f (s_{k,q}^f + \eta_{k,q}^f) + \mathbf{w}_q + \boldsymbol{\mu}_q, \tag{9}$$

where $\eta_{k,q}^f \sim \mathcal{CN}(0, \rho_{t,k}^2 \upsilon_k^f)$ represents the transmitted distortion at the UE$_k^f$ and $\upsilon_k^f = \mathbb{E}[s_{k,q}^f (s_{k,q}^f)^*]$. $\boldsymbol{\mu}_q \sim \mathcal{CN}(0, \rho_r^2 \mathbf{p}_r)$ represents the HWIs at the AP, in which $\mathbf{p}_r = \sum_{f=t}^{r} (\upsilon_k^f \mathbf{I}_M \odot (\mathbf{H}_k^f \boldsymbol{\theta}_q^f)(\mathbf{H}_k^f \boldsymbol{\theta}_q^f)^H)$. $\rho_{t,k}$ and ρ_r denote the error vector magnitude (EVM) at UE$_k^f$ and AP, respectively.

Remark 1: Suppose Q time slots are used for pilots transmission, we can received the pilot signal matrix $\mathbf{Y}_k = [\tilde{\mathbf{y}}_{k,1}, \tilde{\mathbf{y}}_{k,2}, \cdots, \tilde{\mathbf{y}}_{k,Q}] \in \mathbb{C}^{M\times Q}$ at the AP. In [5], the classic LS estimation is utilized for the cascaded channel estimation

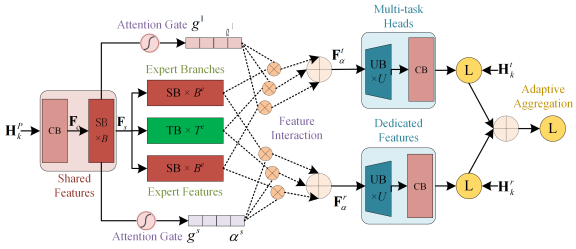

Fig. 2. The proposed multi-task learning (MTL) framework.

estimation in STAR-RIS systems, in which the pilot overhead Q is required to satisfy $Q \ge 2N$ due to the full-rank condition. Note that there are two cascaded links needed to be estimated compared with the conventional reflection-only RIS.

III. Proposed Methods

A. Channel Correlations in STAR-RIS Systems

In STAR-RIS systems, the transmitting user UE$_k^t$ and reflecting user UE$_k^r$ communicate with the AP via the same STAR-RIS, and hence the cascaded channels $\mathbf{H}_k^f(\forall f \in \{t, r\})$ associated with UE$_k^t$ and UE$_k^r$ shares the same STAR-RIS→AP channel $\mathbf{G}$. In [11], the multi-user channel correlations is explicitly characterized as a scalar $\mathcal{S}_k = \mathbf{H}_k^t/\mathbf{H}_k^r = \mathbf{h}_k^t/\mathbf{h}_k^r$, and then the channel estimation is converted to the estimation of scalar $\mathcal{S}_k$ for non-typical users. In this work, we leverage the MTL to implicitly exploit the multi-user correlations, and directly realize the joint cascaded channel estimation. The MTL-based channel estimation model avoid the additional training overhead caused by only supporting one-to-one mapping in the conventional STL framework. In addition, the MTL increases the diversity of training sample space, which can attain implicit data augmentation.

Furthermore, since the sub-wavelength units of the metasurface are integrated closely in hardware implementation, the channels at the neighboring elements of STAR-RIS are highly correlated, which motivates us to design a channel extrapolation strategy to reduce the pilot overhead. Specifically, we assume that a LS pre-estimator is used to obtain the partial channel $\mathbf{H}_k^P \in \mathbb{C}^{M\times P}$ with a few pilot slots P. Specifically, we select P STAR-RIS elements as a subset $\mathcal{P}$ of whole STAR-RIS elements, satisfying $\mathcal{P} = \{1, \Gamma + 1, \cdots, (P-1)\times\Gamma + 1\}$ with the sampling interval $\Gamma = 2^U (0 \le U \le \log_2 N)$. Then, a channel extrapolation network is constructed to realize the mapping from $\mathbf{H}_k^P$ to the complete channel matrix $\mathbf{H}_k^f \in \mathbb{C}^{M\times N}$. However, in contrast to the partial channel acquisition in conventional RIS systems [12], the single transmitting channel $\mathbf{H}_k^t$ and reflecting channel $\mathbf{H}_k^r$ are hard to obtain due to the superposed transmitting and reflecting signal at the AP, which results in larger channel reconstruction difficulty in STAR-RIS systems than conventional RIS systems.

B. Multi-task Learning for Joint Channel Estimation

In Fig. 2, we present the proposed MTL framework for joint cascaded channel estimation, which is a low-level shared MTL

framework and can be divided into three parts, i.e., shared features extraction in the bottom of network, features interaction in different expert branches, and multi-task heads in the network output layers. In the completely shared-bottom MTL with S tasks, the individual output $\mathbf{O}_s \in \mathbb{C}^{M\times N\times 2}, (1 \le s \le S)$ for the s-th subtask head can be represented as

$$\mathbf{O}_s = \omega^s(f(\mathbf{H}_k^P)), \tag{10}$$

where $\mathbf{H}_k^P$, function $f(\cdot)$ and function $\omega^s(\cdot)$ denote the input tensor, the shared-bottom module and the individual s-th task-specific head, respectively. In this framework, each subtask affects other subtasks by updating common weight parameters in the shared layers, while is also constructed in its own unique way on top of the shared low-level representations.

To model the task relationships and learns task-specific functionalities built upon shared representations, we leverage a multi-gate mixture-of-experts (MMoE) framework to design the proposed MTL framework [13], which is given by

$$\mathbf{O}_s = \omega^s\left(\sum_{i=1}^{I} g^s\left(f\left(\mathbf{H}_k^P\right)\right) \odot \varpi^i\left(f\left(\mathbf{H}_k^P\right)\right)\right), \tag{11}$$

where function $\varpi^i(\cdot)$ denotes the i-th expert module to capture shared task information for different perspectives. The function $g^s(\cdot)$ is a gating network for the s-th task, which is generated by utilizing the split attention mechanism. Specifically, for the feature map $\mathbf{F}_s$ obtained by the shared layers, the global average pooling (GAP) layer is used to obtain the feature vector $\mathbf{v} \in \mathbb{R}^C$. Then, we utilize a linear layer with weight $\mathbf{W}_\alpha \in \mathbb{R}^{C\times IC}$ to generate the feature tensor $\mathbf{v}_\alpha = \mathbf{v}\mathbf{W}_\alpha \in \mathbb{R}^{IC}$, and the dimension of $\mathbf{v}_\alpha$ is reshaped as $\mathbb{R}^{I\times C}$. Next, we utilize the Softmax function to generate the attention weight $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_c, \cdots, \boldsymbol{\alpha}_C] \in \mathbb{R}^{I\times C}$ and $\boldsymbol{\alpha}_c \in \mathbb{R}^I$, i.e., $\boldsymbol{\alpha}_c = \text{Softmax}(\mathbf{v}_c) = \frac{e^{v_c^i}}{\sum_{i=1}^I e^{v_c^i}}$, satisfying $\sum_{i=1}^I \boldsymbol{\alpha}_c^i = 1$. Based on the $\boldsymbol{\alpha}^s$ obtained by gating function $g^s(\cdot)$, the different expert branches are integrated with adaptive weights to generate the feature map $\mathbf{F}_\alpha^f (\forall f \in \{t, r\})$. Compared with the completely shared-bottom MTL framework, the proposed MTL can adaptively learn either shared information and task-specific information by the experts assembling [13].

In multi-task optimization process, the loss balancing strategy of different subtasks need to be carefully designed to alleviate task competition. For the channel estimation in STAR-RIS system, the ES ratio β^f will affect the estimation performance of the transmitting and reflecting cascaded channel, i.e., the corresponding channel estimation accuracy can be improved with larger β^f and vice versa. In this work, we aggregate the loss function of subtasks by an adaptive learning method. Specifically, we utilize the prior ES ratio β^f to balance the network training, and then the homoscedastic task uncertainty is used to obtain a leranable scalar σ_s for the subtask s [14], which is given by

$$\mathcal{L}_{\text{joint}}(\sigma_s) \approx \sum_{i=s}^{S} \frac{1}{2\sigma_s^2}(2-\beta_s)\mathcal{L}_s + \log\sigma_s, \tag{12}$$

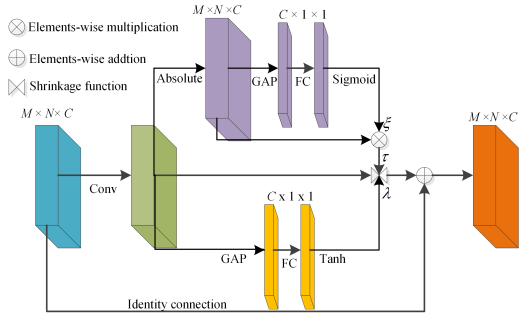

Fig. 3. The network architecture for proposed shrinkage block (SB).

where $\beta_1 = \beta^t, \beta_2 = \beta^r$, and ℓ_1-norm is used as the loss function of each subtask, i.e., $\mathcal{L}_s = |\mathbf{H}_k^f - \mathbf{O}_s|$.

C. Attention-based Multi-task Network architecture

As illustrated in Fig. 2., based on the proposed MTL framework, we further develop an efficient MTN backbone to realize the joint hybrid-field channel estimation in STAR-RIS systems. In the shared features extraction module of the MTN, we first use a convolutional block (CB) with C filters to increase the channel dimension of the input tensor $\bar{\mathbf{H}}_k^P = \{\text{Re}(\mathbf{H}_k^P), \text{Im}(\mathbf{H}_k^P)\} \in \mathbb{R}^{M\times P\times 2}$, i.e., $\bar{\mathbf{H}}_k^P$ is converted to $\mathbf{F}_c \in \mathbb{R}^{M\times P\times C}$. Considering the noise components of the input tensor introduced by the LS pre-estimation, we design a learnable denoising module by fusing the thresholding denoising and soft-attention mechanism. The shrinkage function of traditional thresholding denoising can be expressed as [15]

$$SF(\mathbf{x}, \boldsymbol{\tau}) = \begin{cases} \text{sgn}(\mathbf{x})(|\mathbf{x}| - a\boldsymbol{\tau}), & |\mathbf{x}| \ge \boldsymbol{\tau}, \\ 0, & |\mathbf{x}| \le \boldsymbol{\tau}, \end{cases} \tag{13}$$

where $\text{sgn}(\cdot)$ denotes the symbol function, $\mathbf{x}$, a and $\boldsymbol{\tau} \ge 0$ are the input signal, the given parameter and the threshold, respectively. When parameter a is set to $a = 1$, (13) will become the soft thresholding, while the selection of threshold $\boldsymbol{\tau}$ will significantly affect the denoising performance.

In [15], a deep residual shrinkage network is proposed to automatically learn the threshold $\boldsymbol{\tau}$ with specialized network layers by imitating the operations of soft thresholding. However, the signal $\mathbf{x}$ is completely eliminated when $|\mathbf{x}| < \boldsymbol{\tau}$ in (13), which may remove useful features except noises. In this work, we propose an improved shrinkage block (SB) by introducing a new learnable slope $\boldsymbol{\lambda}$, which is given by

$$SF(\mathbf{x}, \boldsymbol{\tau}, \boldsymbol{\lambda}) = \begin{cases} \text{sgn}(\mathbf{x})\left((\boldsymbol{\lambda}+1)|\mathbf{x}| - \boldsymbol{\tau}\right), & |\mathbf{x}| \ge \boldsymbol{\tau}, \\ \boldsymbol{\lambda}\mathbf{x}, & |\mathbf{x}| \le \boldsymbol{\tau}. \end{cases} \tag{14}$$

Fig. 3 shows the detailed architecture of the proposed SB, which is composed of three branches, i.e., the learning of threshold $\boldsymbol{\tau}$, slope $\boldsymbol{\lambda}$, and the identity connection. Note that the range scale of $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ is different due to the distinguishable functionalities. The threshold $\boldsymbol{\tau}$ is required to satisfy $\boldsymbol{\tau} \ge 0$ and is related to the input signal $\mathbf{x}$. Hence, we first compute the absolute value of feature map $\mathbf{F}_c$, and then a feature vector $\bar{\boldsymbol{\xi}}$ is obtained by similar methods with the attention weight $\boldsymbol{\alpha}$.

However, we adopt the Sigmoid function to activate the adaptive weight $\boldsymbol{\xi}$, satisfying $0 \leq \boldsymbol{\xi} = \text{Sigmoid}(\bar{\boldsymbol{\xi}}) = \frac{1}{1+e^{-\bar{\boldsymbol{\xi}}}} \leq 1$. Lastly, the threshold $\boldsymbol{\tau}$ is given by

$$\boldsymbol{\tau} = \boldsymbol{\xi} \odot |\mathbf{F}_c|. \tag{15}$$

In the design of slope $\boldsymbol{\lambda}$, the Tanh activation function is used to provided wider contractility, i.e., $-1 \leq \boldsymbol{\lambda} \leq 1$. The output of the proposed SB is given by

$$\mathbf{F}_s = SF(\mathbf{F}_c, \boldsymbol{\tau}, \boldsymbol{\lambda}) \cdot \mathbf{W}^l + \mathbf{F}_c, \tag{16}$$

where $\mathbf{W}^l \in \mathbb{R}^{C\times C}$ denotes the trainable weights of a linear layer. In the shared-bottom layers of MTN, B SB blocks are stacked to extract the shared features.

In the expert branches, we introduce two different network architecture to model the unique characteristics of hybrid-field cascaded channel. Firstly, the B^e SBs are used to realize the signal denoising and capture the spatial features of cascaded channel. Besides, the local spatial correlations of non-stationary cascaded channel will be partly lost due to the presence of VRs, which restricts the effective feature learning ability for local convolutional operations-based CNN. Hence, we design the self-attention mechanism-based Transformer block (TB) to model the long-range dependency of the non-stationary cascaded channel, which can obtain more effective global information than convolutional operations. Specifically, we flatten the feature map $\mathbf{F}_s$ along the spatial dimension at first, i.e., $\mathbf{F}_s \in \mathbb{R}^{M\times P\times C} \rightarrow \mathbf{F}_t \in \mathbb{R}^{L\times C}, (L = M \times P)$. Then, different linear transformations are applied to obtain the *Key* matrix $\mathbf{K} \in \mathbb{R}^{L\times D}$, *Query* matrix $\mathbf{Q} \in \mathbb{R}^{L\times D}$ and *Value* matrix $\mathbf{V} \in \mathbb{R}^{L\times D}$. According to the scaled dot-product attention [16], the output $\mathbf{A} \in \mathbb{R}^{L\times D}$ of self-attention module is given by

$$\mathbf{A} = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{D}}\right) \cdot \mathbf{V} = \mathbf{E} \cdot \mathbf{V}, \tag{17}$$

where $\mathbf{E}$ is termed as the attention matrix, the hyper-parameter $D = C$ in TB, and T^e TBs are stacked in the proposed MTN.

In the network design of subtask heads, we first use a SB to learn the initial subtask features, and then U upsampling blocks (UBs) is used to recover the complete spatial dimension $M \times N$ of the cascaded channel. Each UB is composed of a nearest interpolation layer with upscaling factor 2 and a convolutional layer. In the last layer of multi-task heads, a convolutional layer with 2 filters is used to match the two-channel dimension of the cascaded channel matrix.

Remark 2: In the proposed MTN architecture, we introduce three different attention mechanisms to learn common correlations between subtasks and unique characteristics of each subtasks, which can be divided into the split attention in the gating network, the self attention in Transformer layers, and the soft attention in the learnable shrinkage function.

IV. Numerical Results

In the simulation, we set $M = 4\times 8$, $N^s = 8\times 64$, $\nu = 2\times 2$, and hence we have $N = 4 \times 32$. The communication carrier frequency is set to $f_c = 73$ GHz, while the large-scale path loss parameters, array gains and the scatters distribution refer to the setting in [10]. We set $\rho = \rho_t = \rho_r = 0.1$ for the hardware impairments in the transmitter and receiver. The hyper-parameters of MTN are set to $B = 2$, $B^e = 4$, $T^e = 2$, $C = 64$, respectively. In the proposed MTN, the required overhead Q is equal to the number of selected STAR-RIS elements $P = N/\Gamma$ in LS pre-estimation. We adopt normalized mean squared error (NMSE) as the performance evaluation metric of channel estimation, i.e., $\text{NMSE} = \mathbb{E}\left\{||\hat{\mathbf{H}}_k^f - \mathbf{H}_k^f||_F^2 / ||\mathbf{H}_k^f||_F^2\right\}$, in which $\hat{\mathbf{H}}_k^f, \forall f \in \{t, r\}$ represents the estimated channel and $||\cdot||_F$ denotes the Frobenius norm. We compare the proposed MTN model with the existing channel estimation benchmarks from the perspective of channel estimation accuracy and network complexity. Specifically, we provided the LS estimator in [5], the polar domain-based CS estimator in [6], and enhanced super-resolution network (EDSR)-based DL estimator [12]. Moreover, we construct a STN model based on the proposed MTN backbone to shows the generalization of the proposed network in the case of STL framework, where the multi-task heads are reduced to the single-task head.

In Fig. 4, we provide the NMSE performance for different channel estimation schemes with the equal ES ratio, i.e., $\beta^t = \beta^r = 0.5$. As a linear estimator, the LS estimator provides sub-optimal channel estimation performance, and large amount of pilot overhead is required. Since the effective sparse representation is hard to obtain for the non-stationary hybrid-field cascaded channel, the estimation performance of the CS algorithm is limited. Thanks to the powerful non-linear mapping ability of DL model, the DL estimators can obtain better channel estimation accuracy than traditional estimators. Since the TS protocol avoids the power leakage in ES protocol, both EDSR and STN in TS protocol outperform channel estimation models with ES protocol, in which the proposed STN is superior to the EDSR model in terms of estimation accuracy and network complexity. Compared with STL-based estimators, the proposed MTN model has less pilot overhead and training overhead of neural network, and can achieve the estimation accuracy similar to the STN model in ES protocol.

Table I summarizes the required training overhead for DL-based channel estimation schemes. Since the transmitting and reflecting users need to send the pilots at different slots, and hence the required pilot overhead in TS protocol is twice of that in the ES protocol. For STL-based estimators, e.g., EDSR and STN, two independent networks need to be trained and saved for both transmitting and reflecting channels estimation, which results in more floating point of operations (FLOPs) and network parameters. On balance, the proposed MTL model has minimum training overhead for the STAR-RIS channel estimation, while the proposed STN version of the MTN model can provide better estimation accuracy in TS protocol.

Fig. 5 shows the transmitting (T) and reflecting (R) channel estimation performance of the proposed MTN model under different ES ratios, in which $r = \beta^t/\beta^r$ denotes the power ratio of between transmitting and reflecting modes. When larger r is allocated to the transmitting or reflecting modes, the received pilot signal will involve more transmitting or

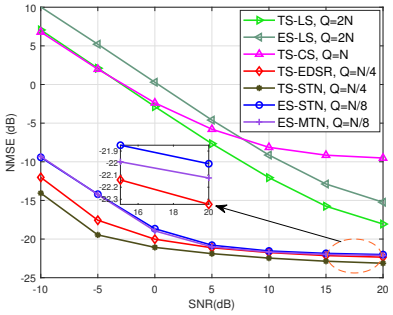

Fig. 4. NMSE for different algorithms

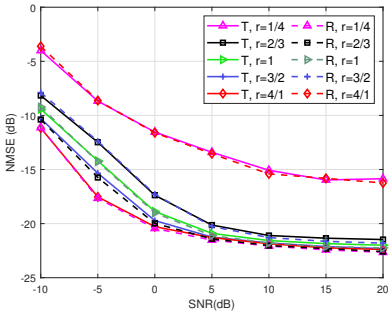

Fig. 5. NMSE for different ES ratios r.

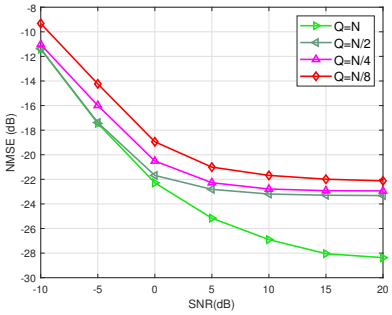

Fig. 6. NMSE for different pilot overhead Q.

TABLE I
TRAINING OVERHEAD FOR DIFFERENT DL MODELS

	FLOPs (G)	Parameters (M)	Pilots (P)
TS-EDSR	1.286×2	1.583×2	$N/8 \times 2$
TS-STN	1.270×2	1.180×2	$N/8 \times 2$
ES-STN	0.859×2	1.422×2	$N/8$
ES-MTN	1.321	1.662	$N/8$

reflecting signal components, and hence the corresponding cascaded channel estimation performance is also improved. In Fig. 6, the NMSE performance under different pilot overhead Q is showed. With the increase of Q, we can pre-estimate more unknown entries of the cascaded channel matrix by the LS algorithm, which reduces the required upscaling factor Γ of channel extrapolation for the MTN model, and hence the channel estimation accuracy is also improved.

V. CONCLUSIONS

By exploiting the ability to simultaneously tune transmission and reflection coefficients of metasurface, the STAR-RIS provide a promising paradigm to realize the *full-space* SREs. In this work, we proposed a MTL-based joint cascaded channel estimation model by utilizing the channel correlations in terms of user domain and spatial domain. In the proposed MTN architecture, the multi-attention mechanism is leveraged to model the shared features and task-specific information for hybrid-field cascaded channels. Compared with existing benchmarks, the proposed MTN can realize satisfactory channel estimation accuracy with less training overhead. In the future works, we will explore the joint optimization of channel estimation and beamforming in STAR-RIS systems.

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62101205, in part by the Natural Science Foundation of Hubei Province under Grant 2021CFB248, in part by the Key Research and Development Program of Hubei Province under Grants 2023BAB061, and in part by the Natural Science Foundation of Hunan Province under Grant 2023JJ50045.

REFERENCES

[1] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, 2020.

[2] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, 2021.

[3] X. Li, Y. Zheng, M. Zeng, Y. Liu, and O. A. Dobre, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2684–2688, 2023.

[4] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, 2022.

[5] C. Wu, C. You, Y. Liu, X. Gu, and Y. Cai, "Channel estimation for STAR-RIS-aided wireless communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 652–656, 2022.

[6] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, Jan. 2023.

[7] Y. Han, S. Jin, C.-K. Wen, and T. Q. Quek, "Localization and channel reconstruction for extra large RIS-assisted massive MIMO systems," *IEEE J. Sel. Top. Signal Process.*, 2022.

[8] J. Xiao, J. Wang, Z. Chen, and G. Huang, "U-MLP based hybrid-field channel estimation for XL-RIS assisted millimeter-wave MIMO systems," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2023.

[9] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, 2020.

[10] E. Basar, I. Yildirim, and F. Kilinc, "Indoor and outdoor physical channel modeling and efficient positioning for reconfigurable intelligent surfaces in mmWave bands," *IEEE Trans. Wireless Commun.*, vol. 69, no. 12, pp. 8600–8611, 2021.

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

[12] 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. 11 083–11 088, 2021.

[13] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *Proc. ACM SIGKDD*, 2018, pp. 1930–1939.

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

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

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. NeurIPS*, vol. 30, 2017.