SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Empowering Over-the-Air Personalized Federated Learning via RIS

Wei Shi[1,2], Jiacheng Yao[1,2], Jindan Xu[3*], Wei Xu[1,2*], Lexi Xu[4] & Chunming Zhao[1,2]

[1] *National Mobile Communications Research Laboratory, Southeast University, Nanjing* 210096*, China;*
[2] *Purple Mountain Laboratories, Nanjing* 211111*, China;*
[3] *School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore* 639798*, Singapore;*
[4] *Research Institute, China United Network Communications Corporation, Beijing* 100048*, China*

Appendix A Proof of Theorem 1

By substituting the RIS phase shifts in Theorem 1, the mean of $\ell_{m,k}$ for $k \in \mathcal{K}_m$ is calculated as

$$\mathbb{E}\left[\ell_{m,k}\right] = \frac{\sqrt{p_k}}{\lambda_m}\sum_{i=1}^{M}\beta_{i,k}\mathbb{E}\left[\Re\left\{\mathbf{h}_{p,i,m}^H\mathbf{\Theta}_i\mathbf{h}_{i,k}\right\}\right]. \tag{A1}$$

1) For $i = m$, we have

$$\begin{aligned}
\mathbb{E}\left[\Re\left\{\mathbf{h}_{p,m,m}^H\mathbf{\Theta}_m\mathbf{h}_{m,k}\right\}\right] &= \Re\left\{\mathbb{E}\left[\mathbf{h}_{p,m,m}^H\mathbf{\Theta}_m\mathbf{h}_{m,k}\right]\right\}\\
&= \Re\left\{\mathbb{E}\left[\sum_{n=1}^{N}\left|h_{p,m,m,n}^*\right| h_{m,k,n}\frac{\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*}{\left|\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*\right|}\right]\right\}\\
&\overset{(a)}{=} \frac{\sqrt{\pi}N}{2}\Re\left\{\mathbb{E}\left[h_{m,k,n}\frac{\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*}{\left|\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*\right|}\right]\right\}\\
&\overset{(b)}{=} \frac{\sqrt{\pi}N}{2\left|\mathcal{K}_m\right|}\Re\left\{\mathbb{E}\left[\sum_{k\in\mathcal{K}_m}h_{m,k,n}\frac{\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*}{\left|\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*\right|}\right]\right\}\\
&= \frac{\sqrt{\pi}N}{2\left|\mathcal{K}_m\right|}\Re\left\{\mathbb{E}\left[\left|\sum_{k\in\mathcal{K}_m}h_{m,k,n}\right|\right]\right\}\\
&\overset{(c)}{=} \frac{\sqrt{\pi}N}{2\left|\mathcal{K}_m\right|}\frac{\sqrt{\left|\mathcal{K}_m\right|\pi}}{2} = \frac{\pi N}{4\sqrt{\left|\mathcal{K}_m\right|}},
\end{aligned} \tag{A2}$$

where (a) is due to the independence between different channels and $\mathbb{E}\left[\left|h_{p,m,m,n}^*\right|\right] = \frac{\sqrt{\pi}}{2}$ [1], (b) follows from the identically distributed characteristic of the random variable (RV) $h_{m,k,n}\frac{\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*}{\left|\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*\right|}$ for different $k \in \mathcal{K}_m$, and (c) comes from the fact that $\sum_{k\in\mathcal{K}_m}h_{m,k,n} \sim \mathcal{CN}(0, |\mathcal{K}_m|)$ [2,3].

2) For $i \neq m$, we have

$$\begin{aligned}
\mathbb{E}\left[\Re\left\{\mathbf{h}_{p,i,m}^H\mathbf{\Theta}_i\mathbf{h}_{i,k}\right\}\right] &= \Re\left\{\mathbb{E}\left[\mathbf{h}_{p,i,m}^H\mathbf{\Theta}_i\mathbf{h}_{i,k}\right]\right\}\\
&= \Re\left\{\mathbb{E}\left[\sum_{n=1}^{N}h_{p,i,m,n}^*\mathrm{e}^{-j\angle h_{p,i,i,n}^*}h_{i,k,n}\frac{\sum_{k'\in\mathcal{K}_i}h_{i,k',n}^*}{\left|\sum_{k'\in\mathcal{K}_i}h_{i,k',n}^*\right|}\right]\right\}\\
&\overset{(d)}{=} N\cdot\Re\left\{\mathbb{E}\left[h_{p,i,m,n}^*\right]\mathbb{E}\left[\mathrm{e}^{-j\angle h_{p,i,i,n}^*}\right]\mathbb{E}\left[h_{i,k,n}\right]\mathbb{E}\left[\frac{\sum_{k'\in\mathcal{K}_i}h_{i,k',n}^*}{\left|\sum_{k'\in\mathcal{K}_i}h_{i,k',n}^*\right|}\right]\right\}\\
&= 0,
\end{aligned} \tag{A3}$$

where (d) exploits the independence of $\mathbf{h}_{p,i,m}$, $\mathbf{h}_{p,i,i}$, $\mathbf{h}_{i,k}$, and $\mathbf{h}_{i,k'}$, and the last equality comes from $h_{p,i,m,n} \sim \mathcal{CN}(0,1)$.

Then, by substituting (A2) and (A3) into (A1), it yields

$$\mathbb{E}\left[\ell_{m,k}\right] = \frac{\sqrt{p_k}\beta_{m,k}}{\lambda_m}\frac{\pi N}{4\sqrt{\left|\mathcal{K}_m\right|}} > 0. \tag{A4}$$

* Corresponding author (email: jindan.xu@ntu.edu.sg, wxu@seu.edu.cn)

In addition, the mean of $\ell_{m,k'}$ for $k' \in \mathcal{K}_{m'}$ and $m' \neq m$ is calculated as

$$\mathbb{E}\left[\ell_{m,k'}\right] = \frac{\sqrt{p_{k'}}}{\lambda_m} \sum_{i=1}^{M} \beta_{i,k'} \mathbb{E}\left[\Re\left\{\mathbf{h}_{p,i,m}^H \boldsymbol{\Theta}_i \mathbf{h}_{i,k'}\right\}\right]. \tag{A5}$$

1) For $i = m$, we have

$$\begin{aligned}\mathbb{E}\left[\Re\left\{\mathbf{h}_{p,m,m}^H \boldsymbol{\Theta}_m \mathbf{h}_{m,k'}\right\}\right] &= \Re\left\{\mathbb{E}\left[\mathbf{h}_{p,m,m}^H \boldsymbol{\Theta}_m \mathbf{h}_{m,k'}\right]\right\} \\ &= \Re\left\{\mathbb{E}\left[\sum_{n=1}^{N} \left|h_{p,m,m,n}^*\right| h_{m,k',n} \frac{\sum_{k\in\mathcal{K}_m} h_{m,k,n}^*}{\left|\sum_{k\in\mathcal{K}_m} h_{m,k,n}^*\right|}\right]\right\} \\ &= N \cdot \Re\left\{\mathbb{E}\left[\left|h_{p,m,m,n}^*\right|\right] \mathbb{E}\left[h_{m,k',n}\right] \mathbb{E}\left[\frac{\sum_{k\in\mathcal{K}_m} h_{m,k,n}^*}{\left|\sum_{k\in\mathcal{K}_m} h_{m,k,n}^*\right|}\right]\right\} = 0.\end{aligned} \tag{A6}$$

2) For $i = m'$, we have

$$\begin{aligned}\mathbb{E}\left[\Re\left\{\mathbf{h}_{p,m',m}^H \boldsymbol{\Theta}_{m'} \mathbf{h}_{m',k'}\right\}\right] &= \Re\left\{\mathbb{E}\left[\mathbf{h}_{p,m',m}^H \boldsymbol{\Theta}_{m'} \mathbf{h}_{m',k'}\right]\right\} \\ &= \Re\left\{\mathbb{E}\left[\sum_{n=1}^{N} h_{p,m',m,n}^* \mathrm{e}^{-j\angle h_{p,m',m',n}^*} h_{m',k',n} \frac{\sum_{k'\in\mathcal{K}_{m'}} h_{m',k',n}^*}{\left|\sum_{k'\in\mathcal{K}_{m'}} h_{m',k',n}^*\right|}\right]\right\} \\ &= N \cdot \Re\left\{\mathbb{E}\left[h_{p,m',m,n}^*\right] \mathbb{E}\left[\mathrm{e}^{-j\angle h_{p,m',m',n}^*}\right] \mathbb{E}\left[h_{m',k',n} \frac{\sum_{k'\in\mathcal{K}_{m'}} h_{m',k',n}^*}{\left|\sum_{k'\in\mathcal{K}_{m'}} h_{m',k',n}^*\right|}\right]\right\} = 0.\end{aligned} \tag{A7}$$

3) For $i \neq m$ and $i \neq m'$, we have

$$\begin{aligned}\mathbb{E}\left[\Re\left\{\mathbf{h}_{p,i,m}^H \boldsymbol{\Theta}_i \mathbf{h}_{i,k'}\right\}\right] &= \Re\left\{\mathbb{E}\left[\mathbf{h}_{p,i,m}^H \boldsymbol{\Theta}_i \mathbf{h}_{i,k'}\right]\right\} \\ &= \Re\left\{\mathbb{E}\left[\sum_{n=1}^{N} h_{p,i,m,n}^* \mathrm{e}^{-j\angle h_{p,i,i,n}^*} h_{i,k',n} \frac{\sum_{k''\in\mathcal{K}_i} h_{i,k'',n}^*}{\left|\sum_{k''\in\mathcal{K}_i} h_{i,k'',n}^*\right|}\right]\right\} \\ &= N \cdot \Re\left\{\mathbb{E}\left[h_{p,i,m,n}^*\right] \mathbb{E}\left[\mathrm{e}^{-j\angle h_{p,i,i,n}^*}\right] \mathbb{E}\left[h_{i,k',n}\right] \mathbb{E}\left[\frac{\sum_{k''\in\mathcal{K}_i} h_{i,k'',n}^*}{\left|\sum_{k''\in\mathcal{K}_i} h_{i,k'',n}^*\right|}\right]\right\} = 0.\end{aligned} \tag{A8}$$

Then, by substituting (A6), (A7) and (A8) into (A5), it yields

$$\mathbb{E}\left[\ell_{m,k'}\right] = 0. \tag{A9}$$

Appendix B Proof of Proposition 1

By directly applying Theorem 1 and substituting the parameters in Proposition 1 into the expectation of the estimated global gradient $\hat{\mathbf{g}}_{m,t}$, we have

$$\begin{aligned}\mathbb{E}\left[\hat{\mathbf{g}}_{m,t}\right] &= \sum_{k\in\mathcal{K}_m} \mathbb{E}\left[\ell_{m,k}\right] \bar{\mathbf{g}}_{m,t,k} + \sum_{k\in\mathcal{K}_m} \frac{u_{m,t,k}}{|\mathcal{K}_m|}\mathbf{1} + \sum_{\substack{1\leqslant m'\leqslant M\\ m'\neq m}} \sum_{k'\in\mathcal{K}_{m'}} \mathbb{E}\left[\ell_{m,k'}\right] \bar{\mathbf{g}}_{m',t,k'} + \mathbb{E}\left[\bar{\mathbf{z}}_{m,t}\right] \\ &= \sum_{k\in\mathcal{K}_m} \frac{\sigma_{m,t,k}}{|\mathcal{K}_m|} \frac{1}{\sigma_{m,t,k}} \left(\mathbf{g}_{m,t,k} - u_{m,t,k}\mathbf{1}\right) + \sum_{k\in\mathcal{K}_m} \frac{u_{m,t,k}}{|\mathcal{K}_m|}\mathbf{1} \\ &= \frac{1}{|\mathcal{K}_m|} \sum_{k\in\mathcal{K}_m} \mathbf{g}_{m,t,k} = \mathbf{g}_{m,t}.\end{aligned} \tag{B1}$$

Therefore, the expectation of the estimated global gradient $\hat{\mathbf{g}}_{m,t}$ is equal to the ground-truth global gradient $\mathbf{g}_{m,t}$ for $m \in [M]$, which ensures the unbiasedness of gradient transmission [4]. This completes the proof.

Appendix C Proof of Proposition 2

To begin with, we formulate the MSE of gradient estimation for cluster m as

$$\begin{aligned}\mathrm{MSE}_m &= \mathbb{E}\left[\left\|\mathbf{g}_{m,t} - \hat{\mathbf{g}}_{m,t}\right\|^2\right] \\ &= \mathbb{E}\left[\left\|\sum_{k\in\mathcal{K}_m} \ell_{m,k}\bar{\mathbf{g}}_{m,t,k} + \sum_{k\in\mathcal{K}_m} \frac{u_{m,t,k}}{|\mathcal{K}_m|}\mathbf{1} + \sum_{\substack{1\leqslant m'\leqslant M\\ m'\neq m}} \sum_{k'\in\mathcal{K}_{m'}} \ell_{m,k'}\bar{\mathbf{g}}_{m',t,k'} + \bar{\mathbf{z}}_{m,t} - \sum_{k\in\mathcal{K}_m} \frac{1}{|\mathcal{K}_m|}\mathbf{g}_{m,t,k}\right\|^2\right]\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{=} \mathbb{E}\left[\left\| \sum_{k\in\mathcal{K}_m}\left(\ell_{m,k}-\frac{\sigma_{m,t,k}}{|\mathcal{K}_m|}\right)\bar{\mathbf{g}}_{m,t,k} + \sum_{\substack{1\leqslant m'\leqslant M\\ m'\neq m}}\sum_{k'\in\mathcal{K}_{m'}}\ell_{m,k'}\bar{\mathbf{g}}_{m',t,k'}+\bar{\mathbf{z}}_{m,t}\right\|^2\right]\\
&\overset{(b)}{=} \sum_{k\in\mathcal{K}_m}\left(\ell_{m,k}-\frac{\sigma_{m,t,k}}{|\mathcal{K}_m|}\right)^2\sigma_{m,t,k}^2 D + \sum_{\substack{1\leqslant m'\leqslant M\\ m'\neq m}}\sum_{k'\in\mathcal{K}_{m'}}\ell_{m,k'}\sigma_{m',t,k'}^2 D+\frac{\sigma^2 D}{\lambda_m^2}\\
&\overset{(c)}{=} \left(\sum_{i=1}^{M}\sum_{k\in\mathcal{K}_i}p_k\bar{h}_{m,k}^2\sigma_{i,t,k}^2 D+\sigma^2 D\right)\frac{1}{\lambda_m^2}-2\left(\sum_{k\in\mathcal{K}_m}\frac{\sqrt{p_k}\bar{h}_{m,k}\sigma_{m,t,k}^3 D}{|\mathcal{K}_m|}\right)\frac{1}{\lambda_m}+\sum_{k\in\mathcal{K}_m}\frac{\sigma_{m,t,k}^4 D}{|\mathcal{K}_m|^2},
\end{aligned} \tag{C1}$$

where (a) is due to the definition of $\bar{g}_{m,t,k}$, (b) exploits the statistics of $\bar{\mathbf{g}}_{m,t,k}$ and $\bar{\mathbf{z}}_{m,t}$, and (c) comes from the definition of $\ell_{m,k}$. Note that the optimization of denoising factor is an unconstrained problem. For any given power control, we derive the optimal denoising factor, λ_m, by checking the following equality

$$\begin{aligned}
&\frac{\partial \mathrm{MSE}_m}{\partial\lambda_m} = -2\left(\sum_{i=1}^{M}\sum_{k\in\mathcal{K}_i}p_k\bar{h}_{m,k}^2\sigma_{i,t,k}^2 D+\sigma^2 D\right)\frac{1}{\lambda_m^3}+2\left(\sum_{k\in\mathcal{K}_m}\frac{\sqrt{p_k}\bar{h}_{m,k}\sigma_{m,t,k}^3 D}{|\mathcal{K}_m|}\right)\frac{1}{\lambda_m^2}=0\\
&\Rightarrow \lambda_m^* = |\mathcal{K}_m|\frac{\sum_{i=1}^{M}\sum_{k\in\mathcal{K}_i}p_k\bar{h}_{m,k}^2\sigma_{i,t,k}^2+\sigma^2}{\sum_{k\in\mathcal{K}_m}\sqrt{p_k}\bar{h}_{m,k}\sigma_{m,t,k}^3},
\end{aligned} \tag{C2}$$

and the proof completes.

As for the optimization of power control, substituting the optimal λ_m^* into (C1), we rewrite MSE_m as

$$\begin{aligned}
\mathrm{MSE}_m &= \sum_{k\in\mathcal{K}_m}\frac{\sigma_{m,t,k}^4 D}{|\mathcal{K}_m|^2}-\frac{\left(\sum_{k\in\mathcal{K}_m}\sqrt{p_k}\bar{h}_{m,k}\sigma_{m,t,k}^3\right)^2 D}{|\mathcal{K}_m|^2\left(\sum_{i=1}^{M}\sum_{k\in\mathcal{K}_i}p_k\bar{h}_{m,k}^2\sigma_{i,t,k}^2+\sigma^2\right)}\\
&= \sum_{k\in\mathcal{K}_m}\frac{\sigma_{m,t,k}^4 D}{|\mathcal{K}_m|^2}-\frac{\left(\mathbf{p}^T\mathbf{b}_m\right)^2 D}{\mathbf{p}^T\mathbf{A}_m\mathbf{p}+|\mathcal{K}_m|^2\sigma^2}\\
&= \sum_{k\in\mathcal{K}_m}\frac{\sigma_{m,t,k}^4 D}{|\mathcal{K}_m|^2}-D\frac{\mathbf{p}^T\mathbf{B}_m\mathbf{p}}{\mathbf{p}^T\mathbf{A}_m\mathbf{p}+|\mathcal{K}_m|^2\sigma^2},
\end{aligned} \tag{C3}$$

where $\mathbf{p}\triangleq[\sqrt{p_1},\sqrt{p_2},\cdots,\sqrt{p_K}]^T$, $\mathbf{b}_m\triangleq\sum_{k\in\mathcal{K}_m}\bar{h}_{m,k}\sigma_{m,t,k}^3\mathbf{e}_k$, $\mathbf{A}_m\triangleq|\mathcal{K}_m|^2\mathrm{diag}\left\{\bar{h}_{m,k}^2\sigma_{i,t,k}^2\right\}$, $\mathbf{B}_m\triangleq\mathbf{b}_m\mathbf{b}_m^T$, and $\mathbf{e}_k$ is the Kronecker delta vector with $[\mathbf{e}_k]_k=1$. Now, we formulate an equivalent power control optimization problem for minimizing the sum MSE as

$$\begin{aligned}
&\underset{\mathbf{p}}{\text{maximize}} \quad \sum_{m=1}^{M}\frac{\mathbf{p}^T\mathbf{B}_m\mathbf{p}}{\mathbf{p}^T\mathbf{A}_m\mathbf{p}+|\mathcal{K}_m|^2\sigma^2}\\
&\text{subject to} \quad [\mathbf{p}]_k\leqslant\sqrt{P_k},\ \ \forall k.
\end{aligned} \tag{C4}$$

It worth noting that the problem in (C4) is known as the sum of quadratic ratios maximization, which has been addressed in existing works via branch and bound [5], harmony search method [6] and semidefinite relaxation (SDR) technique [7–9].

Appendix D Details of the Proposed Approach

To summarize, we conclude the proposed RIS-enabled personalized AirFL approach in Algorithm D1.

Algorithm D1 Proposed RIS-enabled personalized AirFL approach

repeat
  The PS broadcasts the latest personalized models $\{\mathbf{w}_{m,t}\}_{m\in[M]}$ to each device.
  for each device $k=1,2,\cdots,K$ **do**
    Identify its cluster $\mathcal{K}_m$.
    Computes its local gradient $\mathbf{g}_{m,t,k}$ based on $\mathbf{w}_m$ and its local dataset $\mathcal{D}_k$.
    Normalize its local gradient as $\bar{\mathbf{g}}_{m,t,k}\triangleq\frac{1}{\sigma_{m,t,k}}(\mathbf{g}_{m,t,k}-u_{m,t,k}\mathbf{1})$.
    Upload the mean $u_{m,t,k}$, standard deviation $\sigma_{m,t,k}$, and its cluster identity m to the PS.
  end for
  The PS configures N RISs according to the specific clusters $\{\mathcal{K}_m\}_{m=1}^M$ and $\theta_{m,n}=-\angle h_{p,m,m,n}^*+\angle\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*$.
  The PS calculates the power control p_k for each device and the denoising factor λ_m for each cluster according to the unbiased strategy or the MMSE design, and then feedback the selected power $\{p_k\}_{k=1}^K$ to each device.
  Each device simultaneously upload its local gradient to the PS based on the predetermined transmit power p_k.
  Based on the received signal, the PS computes an estimated global gradient of cluster m as $\hat{\mathbf{g}}_{m,t}=\frac{\Re\{\mathbf{y}_{m,t}\}}{\lambda_m}+\sum_{k\in\mathcal{K}_m}\frac{u_{m,t,k}}{|\mathcal{K}_m|}\mathbf{1}$.
  The PS updates the personalized model for cluster m through $\mathbf{w}_{m,t+1}=\mathbf{w}_{m,t}-\eta_{m,t}\hat{\mathbf{g}}_{m,t}$.
  Set $t=t+1$.
until Convergence

References

1 Xu J D, Yuen C, Huang C W, et al. Reconfiguring wireless environment via intelligent surfaces for 6G: reflection, modulation, and security. Sci China Inf Sci, 2023, 66: 130304
2 Shi W, Xu J D, Xu W, et al. On secrecy performance of RIS-assisted MISO systems over Rician channels with spatially random eavesdroppers. IEEE Trans Wireless Commun, 2024, 23: 8357–8371
3 Shi W, Xu J D, Xu W, et al. Secure outage analysis of RIS-assisted communications with discrete phase control. IEEE Trans Veh Technol, 2023, 72: 5435–5440
4 Yao J C, Yang Z H, Xu W, et al. GoMORE: global model reuse for resource-constrained wireless federated learning. IEEE Wireless Commun Lett, 2023, 12: 1543–1547
5 Shen P P, Chen Y Q, Ma Y. Solving sum of quadratic ratios fractional programs via monotonic function. Appl Math Comput, 2009, 212: 234–244
6 Jaberipour M, Khorram E. Solving the sum-of-ratios problems by a harmony search algorithm. J Comput Appl Math, 2010, 234: 733–742
7 Yao J C, Xu J D, Xu W, et al. Superimposed RIS-phase modulation for MIMO communications: a novel paradigm of information transfer. IEEE Trans Wireless Commun, 2024, 23: 2978–2993
8 He Z Y, Xu W, Shen H, et al. Full-duplex communication for ISAC: joint beamforming and power optimization. IEEE J Sel Areas Commun, 2023, 41: 2920–2936
9 He Z Y, Shen H, Xu W, et al. MSE-based training and transmission optimization for MIMO ISAC systems. IEEE Trans Signal Process, 2024, 72: 3104–3121