Supplementary Material for Analysis and Optimizations

Appendix A
Proof of Theorem 1

Proof. We first decompose $\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^*\|_2^2\right]$ as

$$\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^*\|_2^2\right]=\mathbb{E}\left[\|\boldsymbol{\theta}_{t}-\boldsymbol{\theta}^*\|_2^2\right]+\underbrace{\mathbb{E}\left[\left\|-\frac{\eta}{K}\sum_{i\in\hat{\mathcal{S}}_t}K_{i,t}\mathbf{g}_{i,t}\right\|_2^2\right]}_{R_2}$$
$$+\underbrace{2\mathbb{E}\left[\left\langle-\frac{\eta}{K}\sum_{i\in\hat{\mathcal{S}}_t}K_{i,t}\mathbf{g}_{i,t},\boldsymbol{\theta}_t-\boldsymbol{\theta}^*\right\rangle\right]}_{R_1}$$
$$+\frac{\eta^2}{K^2}\left(\Phi_1\left(\{P_{1,i,t},\forall i\},\lambda_t,K\right)+\Phi_2\left(\{P_{2,i,t},\forall i\},\lambda_t\right)\right).$$

For R_1, we have

$$R_1=2\left\langle\mathbb{E}\left[-\frac{\eta}{K}\sum_{i\in\mathcal{S}_t}\mathbf{g}_{i,t}\right],\boldsymbol{\theta}_t-\boldsymbol{\theta}^*\right\rangle\overset{(a)}{=}-2\eta\left\langle\nabla F(\boldsymbol{\theta}_t),\boldsymbol{\theta}_t-\boldsymbol{\theta}^*\right\rangle$$
$$\overset{(b)}{\leq}-2\eta(F(\boldsymbol{\theta}_t)-F(\boldsymbol{\theta}^*)),\tag{29}$$

where (a) follows from Lemma 1, while (b) is derived by convexity. For R_2, we have

$$R_2=\mathbb{E}\left[\left\|-\frac{\eta}{K}\sum_{i\in\mathcal{S}_t}\mathbf{g}_{i,t}\right\|_2^2\right]=\frac{\eta^2}{K^2}\mathbb{E}\left[\left\|\sum_{i\in\mathcal{S}_t}\mathbf{g}_{i,t}\right\|_2^2\right],$$

where the expectation is taken with respect to the device sampling randomness. With K-iid multinomial sampling and Jensen's Inequality, we have

$$R_2\leq\frac{\eta^2}{K^2}\times K\sum_{k=1}^{K}\mathbb{E}_{s_{k,t}\sim\boldsymbol{\beta}}\left[\left\|\mathbf{g}_{s_{k,t},t}\right\|_2^2\right]$$
$$=\frac{\eta^2}{K^2}\times K^2\mathbb{E}_{s_{1,t}\sim\boldsymbol{\beta}}\left[\left\|\mathbf{g}_{s_{1,t},t}\right\|_2^2\right]=\eta^2\sum_{i=1}^{M}\beta_i\left\|\mathbf{g}_{i,t}\right\|_2^2.$$

By Assumptions 3 and 4, we have

$$R_2\leq\eta^2\sum_{i=1}^{M}\beta_i\left\|\nabla F(\boldsymbol{\theta}_t)\right\|_2^2\gamma^2\leq 2L\eta^2\gamma^2(F(\boldsymbol{\theta}_t)-F(\boldsymbol{\theta}^*)).\tag{30}$$

According to (29) and (30), we have

$$\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^*\|_2^2\right]\leq(2L\eta^2\gamma^2-2\eta)(F(\boldsymbol{\theta}_t)-F(\boldsymbol{\theta}^*))$$
$$+\mathbb{E}\left[\|\boldsymbol{\theta}_{t}-\boldsymbol{\theta}^*\|_2^2\right]+\frac{\eta^2}{K^2}\Phi_1\left(\{P_{1,i,t},\forall i\},\lambda_t,K\right)\tag{31}$$
$$+\frac{\eta^2}{K^2}\Phi_2\left(\{P_{2,i,t},\forall i\},\lambda_t\right).$$

Given $0<\eta<\frac{1}{L\gamma^2}$, we have $2\eta-2L\eta^2\gamma^2>0$. Then, according to (31), we have

$$F(\boldsymbol{\theta}_t)-F(\boldsymbol{\theta}^*)\leq\frac{\mathbb{E}\left[\|\boldsymbol{\theta}_{t}-\boldsymbol{\theta}^*\|_2^2\right]-\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}^*\|_2^2\right]}{2\eta-2L\eta^2\gamma^2}$$
$$+\frac{\frac{\eta^2}{K^2}\left(\Phi_1\left(\{P_{1,i,t},\forall i\},\lambda_t,K\right)+\Phi_2\left(\{P_{2,i,t},\forall i\},\lambda_t\right)\right)}{2\eta-2L\eta^2\gamma^2}.$$

By summing over the global model history and averaging, we get

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[F(\boldsymbol{\theta}_t)-F(\boldsymbol{\theta}^*)\right]\leq\frac{\mathbb{E}\left[\|\boldsymbol{\theta}_{0}-\boldsymbol{\theta}^*\|_2^2\right]}{T(2\eta-2L\eta^2\gamma^2)}$$
$$+\frac{\eta^2\sum_{t=0}^{T-1}\left(\Phi_1\left(\{P_{1,i,t},\forall i\},\lambda_t,K\right)+\Phi_2\left(\{P_{2,i,t},\forall i\},\lambda_t\right)\right)}{K^2T(2\eta-2L\eta^2\gamma^2)},$$

which completes the proof. □

Appendix B
Proof of Proposition 1

Proof. For each training round t, we see that

$$\Phi_1\left(\{P_{1,i,t}\},\lambda_t,K\right)+\Phi_2\left(\{P_{2,i,t}\},\lambda_t\right)\overset{(a)}{\geq}\Phi_2\left(\{P_{2,i,t}\},\lambda_t\right).$$

The equality in (a) holds if $P_{1,i,t}=\frac{C^2\lambda_t^2K_{i,t}^2}{Dh_i^2},\forall i\in\hat{\mathcal{S}}_t$. Then, by the peak transmit power constraint $P_{1,i,t}\leq P_i-P_{2,i,t},\forall i\in\hat{\mathcal{S}}_t$, we have

$$\lambda_t\leq\frac{\sqrt{D}}{C}\min_{i\in\hat{\mathcal{S}}_t}\frac{h_i\sqrt{P_i-P_{2,i,t}}}{K_{i,t}}.\tag{32}$$

With $\left\{P_{1,i,t}^*,\forall i,t\right\}$ in (21), the optimization problem (P1) over $\{\{P_{2,i,t},\forall i\},\lambda_t,\forall t\}$ is given by

$$\text{(P2)}\ \min_{\{\{P_{2,i,t}\},\lambda_t,\forall t\}}\ \frac{\eta^2}{K^2}\sum_{t=0}^{T-1}D\frac{\sum_{i=1}^{M}h_i^2P_{2,i,t}+\sigma_n^2}{\lambda_t^2}$$
$$\text{s.t.}\quad\lambda_t\leq\sqrt{\frac{\sum_{j=1}^{M}h_j^2P_{2,j,t}+\sigma_n^2}{8TC^2\max_{i\in\hat{\mathcal{S}}_t}K_{i,t}^2\rho_i}},\quad\forall t,$$
$$\lambda_t\leq\frac{\sqrt{D}}{C}\min_{i\in\hat{\mathcal{S}}_t}\frac{h_i\sqrt{P_i-P_{2,i,t}}}{K_{i,t}},\quad\forall t.$$

In the following, we solve the above problem in two cases.

- Case I: if $\sigma_n^2>8TD\left(\min_{i\in\hat{\mathcal{S}}_t}\frac{h_i^2P_i}{K_{i,t}^2}\right)\left(\max_{i\in\hat{\mathcal{S}}_t}K_{i,t}^2\rho_i\right)$, the optimization problem (P2) becomes

$$\text{(P2.1)}\ \min_{\{\{P_{2,i,t}\},\lambda_t,\forall t\}}\ \frac{\eta^2}{K^2}\sum_{t=0}^{T-1}D\frac{\sum_{j=1}^{M}h_j^2P_{2,j,t}+\sigma_n^2}{\lambda_t^2}$$
$$\text{s.t.}\quad\lambda_t\leq\frac{\sqrt{D}}{C}\min_{i\in\hat{\mathcal{S}}_t}\frac{h_i\sqrt{P_i-P_{2,i,t}}}{K_{i,t}},\forall t.$$

The objective function in (P2.1) is monotonically increasing with respect to $P_{2,i,t},\forall i,t$. Therefore, the optimal transmit power associated with artificial noises is

$P^*_{2,i,t} = 0, \forall i, t$. With the optimal $\{P_{2,i,t}, \forall i, t\}$, the objective function in (P2.1) is monotonically decreasing with respect to λ_t, which yields

$$\lambda_t^* = \frac{\sqrt{D}}{C} \min_{i\in\hat{\mathcal{S}}_t} \frac{h_i\sqrt{P_i}}{K_{i,t}}, \quad \forall t. \tag{33}$$

- Case II: if $\sigma_n^2 \leq 8TD\left(\min_{i\in\hat{\mathcal{S}}_t} \frac{h_i^2 P_i}{K_{i,t}^2}\right)\left(\max_{i\in\hat{\mathcal{S}}_t} K_{i,t}^2\rho_i\right)$, the optimization problem (P2) becomes

$$\text{(P2.2)} \min_{\{\{P_{2,i,t}\},\lambda_t,\forall t\}} \quad \frac{\eta^2}{K^2}\sum_{t=0}^{T-1} D\frac{\sum_{j=1}^M h_j^2 P_{2,j,t} + \sigma_n^2}{\lambda_t^2}$$
$$\text{s.t.} \quad \lambda_t \leq \sqrt{\frac{\sum_{j=1}^M h_j^2 P_{2,j,t} + \sigma_n^2}{8TC^2 \max_{i\in\hat{\mathcal{S}}_t} K_{i,t}^2\rho_i}}, \forall t.$$

Notice that

$$D\frac{\sum_{j=1}^M h_j^2 P_{2,j,t} + \sigma_n^2}{\lambda_t^2} \overset{(b)}{\geq} D\frac{\max_{i\in\hat{\mathcal{S}}_t} 8\lambda_t^2 K_{i,t}^2 C^2 T\rho_i}{\lambda_t^2},$$

To achieve the equality in (b), the optimal $\{\{P_{2,i,t}, \forall i\}, \lambda_t, \forall t\}$ satisfies the following equation,

$$\sum_{j=1}^M h_j^2 P^*_{2,j,t} + \sigma_n^2 = \max_{i\in\hat{\mathcal{S}}_t} 8\lambda_t^{*2} K_{i,t}^2 C^2 T\rho_i, \quad 1 \leq t \leq T.$$

Thus, Proposition 1 is proved. □

Appendix C
Proof of Proposition 2

Proof. Given the number of training rounds T, we analyze the long-term optimization problem for the device sampling variable K. According to the optimal transceiver design in Proposition 1, the objective function $g_1(K)$ with respect to K in Problem (P1) is given by

$$g_1(K) = \begin{cases} \frac{\mathbb{E}\left[C^2\eta\sigma_n^2/\min_{i\in\hat{\mathcal{S}}_t}\left(h_i^2P_i/K_{i,t}^2\right)\right]}{K^2(2-2L\eta\gamma^2)}, & \text{if } K\in[1,\tau_1], \\ \frac{8DC^2T\eta\mathbb{E}\left[\max_{i\in\hat{\mathcal{S}}_t} K_{i,t}^2\rho_i\right]}{K^2(2-2L\eta\gamma^2)}, & \text{if } K\in[\tau_2,\infty), \\ \frac{1}{T}\sum_{t=1}^T \frac{\eta\Phi_2(\{P^*_{2,i,t},\forall i\},\lambda_t^*)}{K^2(2-2L\eta\gamma^2)}, & \text{if } K\in(\tau_1,\tau_2). \end{cases}$$

To find the optimal device sampling variable K^*, we first discuss the optimization problems for different regions of K.

- Case I: $K \in [1, \tau_1]$

  The approximated problem in this case is

$$\text{(P3.1)} \quad \min_K \quad \frac{C^2\eta\sigma_n^2\mathbb{E}\left[\max_{i\in\mathcal{M}} K_{i,t}^2\right]}{K^2(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}} h_i^2 P_i} \tag{34}$$
$$\text{s.t.} \quad 1 \leq K \leq \tau_1.$$

  By Theorem 1 in [?] and the Law of Large Numbers,

$$\mathbb{E}\left[\max_{i\in\mathcal{M}} K_{i,t}^2\right] \approx \begin{cases} \left(\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)\right)^2, & \text{if } K < M\ln M, \\ K^2\beta_0^2, & \text{if } K \geq M\ln M. \end{cases} \tag{35}$$

  Accordingly, we further consider the following two subcases:

  - Subcase I-i: $M\ln M > \tau_1$

    Problem (P3.1) in this subcase is

$$\min_K \quad \frac{C^2\eta\sigma_n^2\left(\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)\right)^2}{K^2(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}} h_i^2 P_i} \tag{36}$$
$$\text{s.t.} \quad 1 \leq K \leq \tau_1.$$

    We first upperbound the objective function in (36) by

$$\frac{\left(\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)\right)^2}{K^2} \leq \frac{\left(\frac{\ln M}{A}\left(1+\frac{Z_1}{A}\right)\right)^2}{K^2},$$

    which is due to the fact that $\frac{\left(\frac{\ln M}{A}\left(1+\frac{1}{A}\right)\right)^2}{K^2} > 0, \forall K$. Here, $Z_1 = \ln^2(M\ln M)$. Then, by defining $Z_{power} \triangleq \frac{C^2\eta\sigma_n^2}{(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}} h_i^2 P_i}$, the problem in (36) becomes

$$\min_K \quad g_{power}(K) = \frac{Z_{power}\left(\frac{\ln M}{A}\left(1+\frac{Z_1}{A}\right)\right)^2}{K^2}$$
$$\text{s.t.} \quad 1 \leq K \leq \tau_1. \tag{37}$$

    The first derivative of $g_{power}(K)$ is

$$\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K} = \frac{2Z_{power}(\ln M)^2}{K^3A^5} \times\left(-A^3 - (2Z_1-1)A^2 - \left(Z_1^2 - 2Z_1\right)A + 2Z_1^2\right).$$

    To analyze the monotonicity of $\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K}$, we further calculate the second derivative as

$$\frac{\mathrm{d}^2g_{power}(K)}{\mathrm{d}K^2} = \frac{2Z_{power}(\ln M)^2}{K^4A^6}\left(A^4 + (2Z_1-3)A^3 + \left(Z_1^2 - 8Z_1 + 3\right)A^2 + \left(6Z_1^2 - 8Z_1\right)A + 10Z_1^2\right).$$

    Since $A > 0$, we have $\frac{\mathrm{d}^2g_{power}(K)}{\mathrm{d}K^2} > 0, \forall K$. Therefore, the first derivative $\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K}$ of the objective function in (37) is monotonically increasing. Moreover, we see that

$$\left.\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K}\right|_{K=1} < 0. \tag{38}$$

    Then, if $\left.\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K}\right|_{K=\tau_1} \geq 0$, the optimal solution K_1^* in this subcase is the unique solution to the equation $\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K} = 0$. Otherwise, $K_1^* = \tau_1$.

  - Subcase I-ii: $M\ln M \leq \tau_1$

    In this subcase, according to (35), we obtain the optimal solution to Problem (P3.1) by solving the following two subproblems. First, when $1 < K < M\ln M$, we have

$$\min_K \quad g_{power}(K) = \frac{Z_{power}\left(\frac{\ln M}{A}\left(1+\frac{Z_1}{A}\right)\right)^2}{K^2}$$
$$\text{s.t.} \quad 1 \leq K \leq M\ln M. \tag{39}$$

    By L'Hôpital's rule,

$$\lim_{K\to M\ln M} \frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K} = \infty. \tag{40}$$

Following the similar discussion in Subcase I-i, the optimal K for Problem (39) is the unique solution to the equation $\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K}=0$.
Then, when $K \in [M \ln M, \tau_1]$, we have

$$\begin{aligned} \min_{K} \quad & \frac{C^2\eta\sigma_n^2\beta_0^2}{(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}} h_i^2 P_i} \\ \text{s.t.} \quad & K \geq M\ln M, \end{aligned} \tag{41}$$

where the objective function does not depend on K. To find the optimal solution K_1^* in this subcase, we need to compare the optimal objective function values achieved by the above two subproblems. Notice that there exists a feasible solution $K_1 = \frac{M\ln M}{1.8} < M\ln M$ such that

$$g_{power}(K_1) < \frac{C^2\eta\sigma_n^2\beta_0^2}{(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}} h_i^2 P_i}.$$

Therefore, the optimal solution K_1^* in this subcase is the unique solution to the equation $\frac{\mathrm{d}g_{power}(K)}{\mathrm{d}K}=0$.

- Case II: $K \in [\tau_2, \infty)$
The approximated problem in this case is

$$\begin{aligned} \text{(P3.2)} \quad \min_{K} \quad & \frac{8DC^2TQ\eta\hat{\beta}_0\mathbb{E}\left[\max_{i\in\mathcal{M}} K_{i,t}^2\right]}{K^2(2-2L\eta\gamma^2)} \\ \text{s.t.} \quad & K \geq \tau_2. \end{aligned} \tag{42}$$

Similar to Case I, depending on the relationship between τ_2 and $M\ln M$, in the following we further discuss two subcases.

  - Subcase II-i: $M\ln M < \tau_2$
  Problem (P3.2) in this subcase is

$$\begin{aligned} \min_{K} \quad & \frac{8DC^2T\eta Q\hat{\beta}_0\beta_0^2}{2-2L\eta\gamma^2} \\ \text{s.t.} \quad & K \geq \tau_2. \end{aligned} \tag{43}$$

  Since the objective function in (43) is monotonically increasing with K, the optimal solution in this subcase is $K_2^* = \tau_2$.
  - Subcase II-ii: $M\ln M \geq \tau_2$
  According to (35), we solve Problem (P3.2) by considering the following two subproblems:
  First, when $K \in [\tau_2, M\ln M)$, we have

$$\begin{aligned} \min_{K} \quad & \frac{8DC^2TQ\eta\hat{\beta}_0\left(\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)\right)^2}{K^2(2-2L\eta\gamma^2)} \\ \text{s.t.} \quad & \tau_2 \leq K < M\ln M. \end{aligned} \tag{44}$$

  By defining $Z_{privacy} \triangleq \frac{8DC^2TQ\eta}{2-2L\eta\gamma^2}$ and upperbounding the objective function in (44), we have

$$\begin{aligned} \min_{K} \quad & g_{privacy}(K) = \frac{Z_{privacy}\hat{\beta}_0\left(\frac{\ln M}{A}\left(1+\frac{Z_1}{A}\right)\right)^2}{K^2} \\ \text{s.t.} \quad & \tau_2 \leq K < M\ln M. \end{aligned} \tag{45}$$

By taking the respective first derivative, we get

$$\begin{aligned} \frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K} = & \frac{Z_{privacy}(\ln M)^2 KA\ln(1-\beta_0)}{K^3A^5}\left(\hat{\beta}_0 \right. \\ & \left. -1\right)\left(A^2+2Z_1A+Z_1^2\right) - \frac{Z_{privacy}(\ln M)^2\hat{\beta}_0}{K^3A^5} \\ & \times \left(2A^3+(4Z_1-2)A^2+\left(2Z_1^2-6Z_1\right)A-4Z_1^2\right). \end{aligned}$$

Lemma 2: $\frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K}$ is monotonically increasing, i.e.,

$$\frac{\mathrm{d}^2 g_{privacy}(K)}{\mathrm{d}K^2} > 0, \quad \forall K < M\ln M. \tag{46}$$

Proof. Denote

$$\begin{aligned} h_1 = & (\ln M)^2 K^2A^2(\ln(1-\beta_0))^2\left(\hat{\beta}_0-1\right)\left(A^2 \right. \\ & \left. +2Z_1A+Z_1^2\right), \\ h_2 = & (\ln M)^2 KA\ln(1-\beta_0)\left(\hat{\beta}_0-1\right)\left(-4A^3\right. \\ & \left. -(8Z_1-4)A^2-\left(4Z_1^2-12Z_1\right)A+8Z_1^2\right), \\ h_3 = & (\ln M)^2\hat{\beta}_0\left(6A^4+(12Z_1-10)A^3+\left(6Z_1^2\right.\right. \\ & \left.\left. -30Z_1+6\right)A^2+\left(-20Z_1^2+24Z_1\right)A+20Z_1^2\right). \end{aligned}$$

We have $\frac{\mathrm{d}^2 g_{privacy}(K)}{\mathrm{d}K^2} = Z_{privacy}\frac{h_1+h_2+h_3}{K^4A^6}$.
Next, we give the lower bound on h_2 that

$$\begin{aligned} h_2 \geq & A(\ln M)^2\left(1-(1-\beta_0)^K\right)\left(-4A^3-(8Z_1\right. \\ & \left. -4)A^2-(4Z_1^2-12Z_1)A+8Z_1^2\right). \end{aligned}$$

The inequality follows from the fact that $A(\ln M)^2(-4A^3-(8Z_1-4)A^2-(4Z_1^2-12Z_1)A + 8Z_1^2) < 0, \forall K < M\ln M$ and the fundamental inequality $\ln(1-\beta_0)^{-K} \leq (1-\beta_0)^{-K}-1, \forall K$.
The lower bound on h_1 is given by

$$h_1 \geq -A^2(\ln M)^2\left(1-(1-\beta_0)^K\right)(A^2+2Z_1A+Z_1^2),$$

which follows from the fact that $-A^2(\ln M)^2(A^2+2Z_1A+Z_1^2) < 0, \forall K < M\ln M$ and a variant of the fundamental inequality $\ln^2(1-\beta_0)^{-K} \leq (1-\beta_0)^{-K}-1, \forall K$.
Summarize the above lower bounds and we get

$$\frac{\mathrm{d}^2 g_{privacy}(K)}{\mathrm{d}K^2} > 0, \quad \forall K < M\ln M. \tag{47}$$

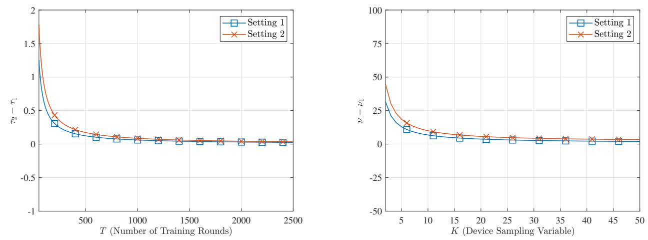
Thus, Lemma 2 is proved. □

By Lemma 2, as well as L'Hôpital's rule that

$$\lim_{K\to M\ln M}\frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K} = Z_{privacy}(\infty+\infty) = \infty,$$

if $\left.\frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K}\right|_{K=\tau_2} \leq 0$, the optimal K in this case is the unique solution to the equation $\frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K}=0$. Otherwise, the optimal K is τ_2.
Then, when $K \in [M\ln M, \infty)$, we have

$$\begin{aligned} \min_{K} \quad & g_3(K) = \frac{8DC^2T\eta Q\hat{\beta}_0\beta_0^2}{2-2L\eta\gamma^2} \\ \text{s.t.} \quad & K \geq M\ln M. \end{aligned} \tag{48}$$



(a) τ_1 and τ_2 (b) ν_1 and ν

Fig. 10. The differences in values in Setting 1, 2.

Since the objective function in (48) is monotonically increasing with K, the optimal K in this case is $M\ln M$.

To find the optimal solution K_2^* in this subcase, we need to compare the optimal objective function values of the above two subproblems (45) and (48). Notice that there exists a feasible solution $\frac{M\ln M}{1.8} < M\ln M$ such that

$$g_{privacy}(\frac{M\ln M}{1.8}) < g_3\left(M\ln M\right). \tag{49}$$

Therefore, the optimal solution K_2^* in this subcase is the unique solution to the equation $\frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K}=0$ if $\left.\frac{\mathrm{d}g_{privacy}(K)}{\mathrm{d}K}\right|_{K=\tau_2}\leq 0$. Otherwise, $K_2^*=\tau_2$.

- Case III: $K\in(\tau_1,\tau_2)$

  The approximated problem in this case is

$$\begin{aligned}\text{(P3.3)}\quad \min_{K}\quad & \frac{1}{T}\sum_{t=1}^{T}\frac{\eta\Phi_2(\{P_{2,i,t}^*,\forall i\},\lambda_t^*)}{K^2(2-2L\eta\gamma^2)}\\ \text{s.t.}\quad & \tau_1<K<\tau_2.\end{aligned} \tag{50}$$

  We can see from Fig. 10 (a) that the values of τ_1, τ_2 are close to each other under Setting 1, 2 considered in our simulated tests. Further, the objective function is intractable in this case, due to the random decision makings for the transceiver design at each round. Therefore, we omit the analysis in this case.

By combining the above results in three cases, Proposition 2 is thus proved. □

Appendix D
Proof of Proposition 3

Proof. Given the number of sampling trials K, we analyze the long-term optimization problem for the FL training rounds T. According to the optimal transceiver design in Proposition 1, the objective function $g_2(T)$ with respect to T in Problem (P1) is given by

$$g_2(T)=\begin{cases} g_2'(T)+\frac{\mathbb{E}\left[C^2\eta\sigma_n^2/\min_{i\in\hat{\mathcal{S}}_t}\left(h_i^2P_i/K_{i,t}^2\right)\right]}{K^2(2-2L\eta\gamma^2)}, & \text{if } T\in[1,\nu_1],\\ g_2'(T)+T\frac{8DC^2\eta\mathbb{E}\left[\max_{i\in\hat{\mathcal{S}}_t}K_{i,t}^2\rho_i\right]}{K^2(2-2L\eta\gamma^2)}, & \text{if } T\in[\nu,\infty),\\ g_2'(T)+\frac{1}{T}\sum_{t=1}^{T}\frac{\eta\Phi_2(\{P_{2,i,t}^*,\forall i\},\lambda_t^*)}{K^2(2-2L\eta\gamma^2)}, & \text{if } T\in(\nu_1,\nu),\end{cases}$$

where $g_2'(T)=\frac{\mathbb{E}\left[\|\boldsymbol{\theta}_0-\boldsymbol{\theta}^*\|_2^2\right]}{T(2\eta-2L\eta^2\gamma^2)}$, $\nu_1\triangleq\frac{\sigma_n^2}{8D\left(\max_{i\in\mathcal{M}}h_i^2P_i\right)\max_{i\in\mathcal{M}}\rho_i}$.

To find the optimal number of training rounds T^*, we first discuss the optimization problems for different regions of T.

- Case I: $T\in[1,\nu_1]$

  The approximated problem in this case is

$$\begin{aligned}\text{(P4.1)}\quad \min_{T}\quad & g_2'(T)+\frac{C^2\eta\sigma_n^2\mathbb{E}\left[\max_{i\in\mathcal{M}}K_{i,t}^2\right]}{K^2(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}}h_i^2P_i}\\ \text{s.t.}\quad & 1\leq T\leq\nu_1.\end{aligned} \tag{51}$$

  According to (35), depending on the relationship between K and $M\ln M$, we further consider the following two subcases:

  - Subcase I-i: $M\ln M>K$

    Problem (P4.1) in this subcase is

$$\begin{aligned}\min_{T}\quad & g_2'(T)+\frac{C^2\eta\sigma_n^2\left(\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)\right)^2}{K^2(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}}h_i^2P_i}\\ \text{s.t.}\quad & 1\leq T\leq\nu_1.\end{aligned} \tag{52}$$

  - Subcase I-ii: $M\ln M\leq K$

    Problem (P4.1) in this subcase is

$$\begin{aligned}\min_{T}\quad & g_2'(T)+\frac{C^2\eta\sigma_n^2\beta_0^2}{(2-2L\eta\gamma^2)\min_{i\in\mathcal{M}}h_i^2P_i}\\ \text{s.t.}\quad & 1\leq T\leq\nu_1.\end{aligned} \tag{53}$$

  Since objective functions in both subcases are monotonically decreasing with T, the optimal solution T_1^* in this case is ν_1.

- Case II: $T\in[\nu,\infty)$

  The approximated problem in this case is

$$\begin{aligned}\text{(P4.2)}\quad \min_{T}\quad & g_2'(T)+T\frac{8DC^2\eta\mathbb{E}\left[\max_{i\in\hat{\mathcal{S}}_t}K_{i,t}^2\rho_i\right]}{K^2(2-2L\eta\gamma^2)}\\ \text{s.t.}\quad & T\geq\nu.\end{aligned} \tag{54}$$

  Similarly, depending on the relationship between K and $M\ln M$, in the following we further discuss two subcases.

  - Subcase II-i: $M\ln M>K$

    Problem (P4.2) in this subcase is

$$\begin{aligned}\min_{T}\quad & g_2'(T)+T\frac{8DC^2Q\eta\hat{\beta}_0\left(\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)\right)^2}{K^2\left(2-2L\eta\gamma^2\right)}\\ \text{s.t.}\quad & T\geq\nu.\end{aligned} \tag{55}$$

    Since the objective function of (55) is convex with respect to T, we calculate the local minimum in the closed form, i.e.,

$$\hat{T}_2=\frac{K}{2C\eta\frac{\ln M}{A}\left(1+\frac{\ln A}{A}\right)}\sqrt{\frac{\mathbb{E}\left[\|\boldsymbol{\theta}_0-\boldsymbol{\theta}^*\|_2^2\right]}{2D\max_{i\in\mathcal{M}}\rho_i}}.$$

    Then, the optimal solution T_2^* in this subcase is

$$T_2^*=\max\left(\hat{T}_2,\nu\right).$$

  - Subcase II-ii: $M\ln M\leq K$

    Problem (P4.2) in this subcase is

$$\begin{aligned}\min_{T}\quad & g_2'(T)+T\frac{8DC^2Q\eta\hat{\beta}_0\beta_0^2}{2-2L\eta\gamma^2}\\ \text{s.t.}\quad & T\geq\nu.\end{aligned} \tag{56}$$

Since the objective function of (56) is again convex with respect to T with the constraint $T \geq \nu$. The optimal solution T_3^* in this subcase is

$$T_3^* = \max\left(\frac{1}{2C\eta\beta_0}\sqrt{\frac{\mathbb{E}\left[\|\boldsymbol{\theta}_0-\boldsymbol{\theta}^*\|_2^2\right]}{2D\max_{i\in\mathcal{M}}\rho_i}}, \nu\right).$$

- Case III: $T \in (\nu_1, \nu)$
  The approximated problem in this case is

$$\begin{aligned}(\text{P4.3})\quad \min_{T}\quad & g_2'(T) + \frac{1}{T}\sum_{t=1}^{T}\frac{\eta\Phi_2(\{P_{2,i,t}^*, \forall i\}, \lambda_t^*)}{K^2(2-2L\eta\gamma^2)}\\ \text{s.t.}\quad & \nu_1 \leq T \leq \nu.\end{aligned} \tag{57}$$

  Similar to Case III in Appendix C, we can see from Fig. 10 (b) that the values of ν_1, ν are close to each other under Setting 1, 2 considered in our simulated tests. Further, the objective function is intractable in this case, due to the random decision makings for the transceiver design at each round. Therefore, we omit the analysis in this case.

By combining the above results in three cases, Proposition 3 is thus proved. □

APPENDIX E
PROOF OF THEOREM 3

Proof. By Assumption 3 in the manuscript, we have

$$\begin{aligned}\mathbb{E}\left[F(\boldsymbol{\theta}_{t+1}) - F(\boldsymbol{\theta}_t)\right] \leq \nabla F(\boldsymbol{\theta}_t)^\top \mathbb{E}\left[\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}_t\right]\\ + \frac{L}{2}\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}_t\|_2^2\right].\end{aligned}$$

Since $\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t = -\eta\hat{\mathbf{r}}_t$, by Eq. (17) in the manuscript, we have

$$\begin{aligned}&\mathbb{E}\left[F(\boldsymbol{\theta}_{t+1}) - F(\boldsymbol{\theta}_t)\right] \leq \frac{L}{2}\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}_t\|_2^2\right] + \nabla F(\boldsymbol{\theta}_t)^\top\\ &\quad\left(\frac{\eta}{K}\mathbb{E}\left[\sum_{i\in\hat{\mathcal{S}}_t}\left(K_{i,t} - \frac{h_i\sqrt{DP_{1,i,t}}}{\lambda_t C}\right)\mathbf{g}_{i,t}\right] - \frac{\eta}{K}\mathbb{E}\left[\frac{\overline{\mathbf{m}}_t}{\lambda_t}\right]\right.\\ &\quad\left. + \mathbb{E}\left[-\frac{\eta}{K}\sum_{i\in\hat{\mathcal{S}}_t}K_{i,t}\mathbf{g}_{i,t}\right]\right)\\ &\overset{(a)}{\leq} -\eta\|\nabla F(\boldsymbol{\theta}_t)\|_2^2 + \frac{L}{2}\mathbb{E}\left[\|\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}_t\|_2^2\right]\\ &\quad + \frac{\eta}{K}\mathbb{E}\|\nabla F(\boldsymbol{\theta}_t)\|_2\,\mathbb{E}\left[\left\|\sum_{i\in\hat{\mathcal{S}}_t}\left(K_{i,t} - \frac{h_i\sqrt{DP_{1,i,t}}}{\lambda_t C}\right)\mathbf{g}_{i,t}\right\|_2\right]\\ &\overset{(b)}{=} \left(-\eta + \frac{L}{2}\eta^2\gamma^2\right)\|\nabla F(\boldsymbol{\theta}_t)\|_2^2\\ &\quad + \frac{L\eta^2}{2K^2}\left(\overline{\Phi}_1(\{P_{1,i,t}, \forall i\}, \lambda_t, K) + \Phi_2(\{P_{2,i,t}, \forall i\}, \lambda_t)\right),\end{aligned}$$

where inequality (a) follows from Lemma 1 in the manuscript, the Jensen's Inequality and the zero mean of the aggreagted noise $\overline{\mathbf{m}}_t$. Equality (b) follows from Eq. (30) in the manuscript as well as the definitions of $\overline{\Phi}_1(\cdot)$ and $\Phi_2(\cdot)$.

Given $0 < \eta < \frac{1}{L\gamma^2}$, we have $2\eta - L\eta^2\gamma^2 > 0$. Then, we have

$$\begin{aligned}\|\nabla F(\boldsymbol{\theta}_t)\|_2^2 \leq \frac{2\mathbb{E}\left[F(\boldsymbol{\theta}_t) - F(\boldsymbol{\theta}_{t+1})\right]}{2\eta - L\eta^2\gamma^2}\\ + \frac{L\eta\left(\overline{\Phi}_1(\{P_{1,i,t}, \forall i\}, \lambda_t, K) + \Phi_2(\{P_{2,i,t}, \forall i\}, \lambda_t)\right)}{K^2(2-L\gamma^2\eta)}.\end{aligned}$$

By summing over the global gradient history and averaging, we get

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\|\nabla F(\boldsymbol{\theta}_t)\|_2^2 \leq \frac{2\mathbb{E}\left[F(\boldsymbol{\theta}_0) - F(\boldsymbol{\theta}_T)\right]}{T(2\eta - L\gamma^2\eta^2)} + \frac{L\eta}{K^2T}\sum_{t=0}^{T-1}\\ \frac{\overline{\Phi}_1(\{P_{1,i,t}, \forall i\}, \lambda_t, K) + \Phi_2(\{P_{2,i,t}, \forall i\}, \lambda_t)}{2 - L\gamma^2\eta},\end{aligned}$$

which completes the proof. □

APPENDIX F
PROOF OF THEOREM 4

Proof. When $0 \leq e \leq E-1$, for any global iteration t, following Eq. (29)-(31) in the manuscript, we have

$$\begin{aligned}F(\overline{\boldsymbol{\theta}}_t^{(e)}) - F(\boldsymbol{\theta}^*) \leq \frac{\mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_t^{(e)} - \boldsymbol{\theta}^*\right\|_2^2\right] - \mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_t^{(e+1)} - \boldsymbol{\theta}^*\right\|_2^2\right]}{2\eta - 2L\eta^2\gamma^2}\\ + \frac{\eta^2\hat{\Phi}_1(\{P_{1,i,t}, \forall i\}, \lambda_t, K, E, e)}{K^2(2\eta - 2L\eta^2\gamma^2)}.\end{aligned}$$

On the other hand, when $e = E$, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_{t+1}^{(0)} - \boldsymbol{\theta}^*\right\|_2^2\right] &= \mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_t^{(E)} - \frac{\eta\overline{\mathbf{m}}_t}{K\lambda_t} - \boldsymbol{\theta}^*\right\|_2^2\right]\\ &= \mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_t^{(E)} - \boldsymbol{\theta}^*\right\|_2^2\right] + \underbrace{\mathbb{E}\left[\left\|\frac{\eta\overline{\mathbf{m}}_t}{K\lambda_t}\right\|_2^2\right]}_{\hat{R}_2}\\ &\quad + \underbrace{\mathbb{E}\left[\left\langle -\frac{\eta\overline{\mathbf{m}}_t}{K\lambda_t}, \overline{\boldsymbol{\theta}}_t^{(E)} - \boldsymbol{\theta}^*\right\rangle\right]}_{\hat{R}_1}.\end{aligned}$$

Since the aggregated noise $\overline{\mathbf{m}}_t$ has zero mean, $\hat{R}_1 = 0$. Then, given $\Phi_2(\cdot)$ in Theorem 1, by summing over the global model history and averaging, we get

$$\begin{aligned}\frac{1}{N}\sum_{t=0}^{T-1}\sum_{e=0}^{E-1}\mathbb{E}\left[F(\overline{\boldsymbol{\theta}}_t^{(e)}) - F(\boldsymbol{\theta}^*)\right] \leq \frac{\eta\sum_{t=0}^{T-1}\Phi_2(\{P_{2,i,t}, \forall i\}, \lambda_t)}{K^2N(2-2L\eta\gamma^2)}\\ + \frac{\mathbb{E}\left[\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\|_2^2\right] - \mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_T^{(0)} - \boldsymbol{\theta}^*\right\|_2^2\right]}{N(2\eta - 2L\eta^2\gamma^2)}\\ + \frac{\eta\sum_{t=0}^{T-1}\sum_{e=0}^{E-1}\hat{\Phi}_1(\{P_{1,i,t}, \forall i\}, \lambda_t, K, E, e)}{K^2N(2-2L\eta\gamma^2)}.\end{aligned}$$

We see that

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_T^{(0)} - \boldsymbol{\theta}^*\right\|_2^2\right] = \mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_T^{(0)} - \boldsymbol{\theta}_T\right\|_2^2\right] + \mathbb{E}\left[\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^*\|_2^2\right]\\ + \underbrace{2\left\langle\mathbb{E}\left[\overline{\boldsymbol{\theta}}_T^{(0)} - \boldsymbol{\theta}_T\right], \boldsymbol{\theta}_T - \boldsymbol{\theta}^*\right\rangle}_{\hat{R}_3},\end{aligned}$$

where $\hat{R}_3 = 0$ according to Lemma 1 in the manuscript. By Lemma 5 in [R4],

$$\mathbb{E}\left[\left\|\overline{\boldsymbol{\theta}}_T^{(0)} - \boldsymbol{\theta}_T\right\|_2^2\right] \leq \frac{4}{K}\eta^2 E^2 C^2.$$

Combine the results above, we have

$$\frac{1}{N}\sum_{t=0}^{T-1}\sum_{e=0}^{E-1}\mathbb{E}\left[F(\overline{\boldsymbol{\theta}}_t^{(e)}) - F(\boldsymbol{\theta}^*)\right] \leq \frac{\mathbb{E}\left[\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\|_2^2\right]}{N(2\eta - 2L\eta^2\gamma^2)} + \eta \times$$
$$\frac{\sum_{t=0}^{T-1}\sum_{e=0}^{E-1}\hat{\Phi}_1\left(\{P_{1,i,t}\}, \lambda_t, K, E, e\right) + \sum_{t=0}^{T-1}\Phi_2\left(\{P_{2,i,t}\}, \lambda_t\right)}{K^2 N(2 - 2L\eta\gamma^2)}$$
$$+ \frac{2\eta E^2 C^2}{KN\left(1 - L\eta\gamma^2\right)},$$

which completes the proof. $\square$