Appendix for the paper "Energy and Spectrum Efficient Federated Learning via High-Precision Over-the-Air Computation"

1 Proof of Theorem 1

Consider a non-convex FL model setting. Under the L-smoothness assumption of the global objective with the expectation taken, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_Q\left[\mathbb{E}_{\text{Air}}\left[f(\mathbf{w}^{r+1}) - f(\mathbf{w}^r)\right]\right]\right] \\
&\leq -\theta\eta\mathbb{E}\left[\mathbb{E}_Q\left[\mathbb{E}_{\text{Air}}\left[\langle \nabla f^r, \nabla F_Q^r \rangle\right]\right]\right] + \frac{\theta^2\eta^2 L}{2}\mathbb{E}\left[\mathbb{E}_Q\left[\mathbb{E}_{\text{Air}}\left[\|\nabla F_Q^r\|^2\right]\right]\right],
\end{aligned} \tag{1}$$

where we take the expectation over the sampling and operations. Next, the following lemmas are proposed to bound terms in the above inequality.

Lemma 1.1. *The inner product between the stochastic gradient ∇F_Q^r and full batch gradient ∇f^r can be bounded as*

$$\begin{aligned}
&\mathbb{E}_{\xi^{(r)}}\mathbb{E}_Q\mathbb{E}_{Air}\left[\langle \nabla f^r, \nabla F_Q^r \rangle\right] \\
&= \mathbb{E}_{\xi^{(r)}}\left[\left\langle \nabla f^r, \frac{1}{K}\sum_{k=1}^{K}\sum_{h=0}^{H-1}\nabla F_k^{r,h}\right\rangle\right] \\
&\leq \frac{1}{2K}\sum_{k=1}^{K}\sum_{h=0}^{H}\left[-\|\nabla f^r\|_2^2 - \|\nabla f_k^{r,h}\|_2^2 + L^2\|\mathbf{w}^r - \mathbf{w}_k^{r,h}\|_2^2\right].
\end{aligned} \tag{2}$$

Here, we set $\nabla F_k^r = \sum_{h=0}^{H-1}\nabla F_k^{r,h}$ and $\nabla F_{k,Q}^r = Q\left(\sum_{h=0}^{H-1}\nabla F_k^{(h,r)}\right)$. We further define $\nabla F_Q^r = Air_{\mathcal{K}}\left(Q\left(\sum_{h=0}^{H-1}\nabla F_k^{r,h}\right)\right)$.

Lemma 1.2. *Similar to the Lemma D.3 in [11], we can bound the distance between the global model and the local model at r-th communication round under Assumption 2 as follows:*

$$\mathbb{E}\left[\|\mathbf{w}^r - \mathbf{w}_k^{r,h}\|_2^2\right] \leq \eta^2 H\sigma^2 + \eta^2\sum_{h=0}^{H-1}H\|\nabla f_k^{r,h}\|_2^2 \tag{3}$$

Lemma 1.3. *The last term in (1) can be calculated as*

$$\begin{aligned}&\mathbb{E}_{\xi^{(r)}}\mathbb{E}_Q\mathbb{E}_{Air}\left[\|Air_{\mathcal{K}}\left(Q\left(\sum_{h=0}^{H-1}\nabla F_k^{r,h}\right)\right)\|^2\right]\leq\frac{\sigma_z^2}{{p_b}^2}\\&+\sum_{k=1}^{K}\frac{q+p_b}{K^2p_b}\operatorname{Var}(\nabla F_k^r)+\sum_{k=1}^{K}\frac{q(2-p_b)+Kp_b}{K^2p_b}\|\nabla f_k^r\|^2\end{aligned}\tag{4}$$

Proof.

$$\begin{aligned}&\mathbb{E}_{\xi^{(r)}}\mathbb{E}_Q\mathbb{E}_{Air}\left[\|Air_{\mathcal{K}}\left(Q\left(\sum_{h=0}^{H-1}\nabla F_k^{r,h}\right)\right)\|^2\right]\\
=&\mathbb{E}_{\xi^{(r)},Q}\left[\frac{1}{K^2}\left(\|\sum_{k=1}^{K}\nabla F_{k,Q}^r\|^2+(\frac{1}{p_b}-1)\sum_{k=1}^{K}\|\nabla F_{k,Q}^r\|^2\right)+\frac{\sigma_z^2}{{p_b}^2}\right]\\
=&\mathbb{E}_{\xi^{(r)}}\left[\mathbb{E}_Q\left[\|\frac{1}{K}\sum_{k=1}^{K}\nabla F_{k,Q}^r\|^2+\frac{1}{K^2}(\frac{1}{p_b}-1)\sum_{k=1}^{K}\|\nabla F_{k,Q}^r\|^2\right]\right]+\frac{\sigma_z^2}{{p_b}^2}\\
=&\mathbb{E}_{\xi^{(r)}}\left[\mathbb{E}_Q\left[\frac{1}{K^2}\sum_{k=1}^{K}\left[\|\nabla F_{k,Q}^r-\nabla F_k^r\|^2\right]\right]+\|\frac{1}{K}\sum_{k=1}^{K}\nabla F_k^r\|^2\right]+\\
&\mathbb{E}_{\xi^{(r)}}\left[\frac{1}{K^2}(\frac{1}{p_b}-1)\sum_{k=1}^{K}(\mathbb{E}_Q\left[\|\nabla F_{k,Q}^r-\nabla F_k^r\|^2\right]+\|\nabla F_k^r\|^2\right]+\frac{\sigma_z^2}{{p_b}^2}\\
=&\mathbb{E}_{\xi^{(r)}}\left[\mathbb{E}_Q\left[\frac{1}{K^2p_b}\sum_{k=1}^{K}\left[\|\nabla F_{k,Q}^r-\nabla F_k^r\|^2\right]\right]+\|\frac{1}{K}\sum_{k=1}^{K}\nabla F_k^r\|^2+\frac{1}{K}(\frac{1}{p_b}-1)\sum_{k=1}^{K}\|\nabla F_k^r\|^2\right]+\frac{\sigma_z^2}{{p_b}^2}\\
\leq&\mathbb{E}_{\xi^{(r)}}\left[\sum_{k=1}^{K}\frac{q}{K^2p}\|\nabla F_k^r\|^2+\|\frac{1}{K}\sum_{k=1}^{K}\nabla F_k^r\|^2+\frac{1}{K}(\frac{1}{p_b}-1)\sum_{k=1}^{K}\|\nabla F_k^r\|^2\right]+\frac{\sigma_z^2}{{p_b}^2}\\
=&\sum_{k=1}^{K}\frac{q}{K^2p_b}\left[\operatorname{Var}(\nabla F_k^r)+\|\nabla f_k^r\|^2\right]+\left[\frac{1}{K^2}\sum_{k=1}^{K}\operatorname{Var}(\nabla F_k^r)+\|\frac{1}{K}\sum_{k=1}^{K}\nabla f_k^r\|^2\right]+\\
&\frac{1}{K}(\frac{1}{p_b}-1)\sum_{k=1}^{K}\|\nabla f_k^r\|^2+\frac{\sigma_z^2}{{p_b}^2}\\
\leq&\sum_{k=1}^{K}\frac{q+p_b}{K^2p_b}\operatorname{Var}(\nabla F_k^r)+\sum_{k=1}^{K}\frac{q(2-p_b)+Kp_b}{K^2p_b}\|\nabla f_k^r\|^2+\frac{\sigma_z^2}{{p_b}^2}\end{aligned}\tag{5}$$

□

According to Assumption 2, we have $\operatorname{Var}(\nabla F_k^r)\leq H\sigma^2$. We further have

$\|\nabla f_k^r\|^2 = \| \sum_{h=0}^{H-1} \nabla f_k^{r,h}\|^2 \leq H \sum_{h=0}^{H-1} \|\nabla f_k^{r,h}\|^2$. Therefore, we have

$$\begin{aligned}
&\mathbb{E}_{\xi^{(r)}}\mathbb{E}_Q\mathbb{E}_{\text{Air}}\left[\|\text{Air}_{\mathcal{K}}\left(Q\left(\sum_{h=0}^{H-1}\nabla F_k^{r,h}\right)\right)\|^2\right] \\
\leq &\frac{q+p_b}{Kp_b}H\sigma^2 + H\frac{q(2-p_b)+Kp_b}{K^2p_b}\sum_{k=1}^{K}\sum_{h=0}^{H}\|\nabla f_k^{r,h}\|^2 + \frac{\sigma_z^2}{{p_b}^2}
\end{aligned} \tag{6}$$

Therefore, by integrating Lemma 1.1, 1.2, and 1.3 into (1), we will have:

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_Q\left[\mathbb{E}_{\text{Air}}\left[f(\mathbf{w}^{r+1}) - f(\mathbf{w}^r)\right]\right]\right] \\
\leq &\frac{\eta\theta}{2K}\sum_{k=1}^{K}\sum_{h=0}^{H}\left[-\|\nabla f^r\|_2^2 - \|\nabla f_k^{r,h}\|_2^2 + L^2\eta^2 H\left[\sigma^2 + H\|\nabla f_k^{r,h}\|_2^2\right]\right] + \\
&\frac{(q+p_b)\theta^2\eta^2 L}{2Kp_b}H\sigma^2 + HL\theta^2\eta^2\frac{q(2-p_b)+Kp_b}{2K^2p_b}\sum_{k=1}^{K}\sum_{h=0}^{H}\|\nabla f_k^{r,h}\|^2 + \frac{\theta^2\eta^2\sigma_z^2 L}{2{p_b}^2} \\
= &-\frac{\eta\theta H}{2}\|\nabla f^r\|_2^2 - \frac{\eta\theta}{2K}(1 - L^2\eta^2H^2 - HL\theta\eta\frac{q(2-p_b)+Kp_b}{Kp_b})\sum_{k=1}^{K}\sum_{h=0}^{H}\|\nabla f_k^{r,h}\|_2^2 \\
&+ \frac{\theta\eta^2 LH}{2K}(\eta LHK + \frac{(p_b+q)\theta}{p_b})\sigma^2 + \frac{\theta^2\eta^2\sigma_z^2 L}{2{p_b}^2}
\end{aligned} \tag{7}$$

If we set $1 - L^2\eta^2H^2 - HL\theta\eta\frac{q(2-p_b)+Kp_b}{Kp_b} \geq 0$, we can get

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_Q\left[\mathbb{E}_{\text{Air}}\left[f(\mathbf{w}^{r+1}) - f(\mathbf{w}^r)\right]\right]\right] \leq -\frac{\eta\theta H}{2}\|\nabla f^r\|_2^2 \\
&+\frac{\theta\eta^2 LH}{2K}(\eta LHK + \frac{(p_b+q)\theta}{p_b})\sigma^2 + \frac{\theta^2\eta^2\sigma_z^2 L}{2{p_b}^2}
\end{aligned} \tag{8}$$

Next, we sum up the above equation over all R communication rounds and get

$$\begin{aligned}
&\frac{1}{R}\sum_{r=0}^{R-1}\|\nabla f^r\|_2^2 \leq \frac{2(f(\mathbf{w}^0) - f(\mathbf{w}^*))}{\eta\theta HR} \\
&+ \frac{\eta L}{K}(\eta LHK + \frac{(p_b+q)\theta}{p_b})\sigma^2 + \frac{\theta\eta L}{H{p_b}^2}\sigma_z^2 \\
&= \frac{2(f(\mathbf{w}^0) - f(\mathbf{w}^*))}{\eta\theta HR} + \frac{\eta\theta L}{K}\frac{(p_b+q)}{p_b}\sigma^2 + \eta^2L^2H\sigma^2 + \frac{\theta\eta L\sigma_z^2}{H{p_b}^2}
\end{aligned} \tag{9}$$

2 Proof of Lemma 1

The second-order partial derivative of functions Θ_1 and Θ_2 can be calculated as:

$$
\begin{aligned}
\frac{\partial\Theta_1}{\partial p_b} &= -\frac{A_0 q}{{p_b}^2 H} - \frac{B_0 q}{2{p_b}^{\frac{3}{2}} H^{\frac{1}{2}} (p_b+q)^{\frac{1}{2}}} \\
\frac{\partial^2\Theta_1}{\partial {p_b}^2} &= \frac{2q\left(A_0\sqrt{H}\left(q\sqrt{p_b}+{p_b}^{3/2}\right)\sqrt{p_b+q}+\frac{1}{2}Hp_bB_0\left(p_b+\frac{3}{4}q\right)\right)}{H^{3/2}{p_b}^{7/2}(p_b+q)^{3/2}} \\
\frac{\partial^2\Theta_1}{\partial p_b \partial H} &= \frac{q\left(A_0\sqrt{p_b}\sqrt{H}\sqrt{p_b+q}+\frac{1}{4}B_0p_bH\right)}{H^{5/2}\sqrt{p_b+q}{p_b}^{5/2}} \\
\frac{\partial\Theta_1}{\partial H} &= -\frac{A_0(p_b+q)}{p_bH^2} - \frac{1}{2}\frac{B_0\sqrt{p_b+q}p_b}{(p_bH)^{3/2}} \\
\frac{\partial^2\Theta_1}{\partial H^2} &= \frac{2\left(A_0(p_b+q)\sqrt{p_bH}+\frac{3}{8}B_0\sqrt{p_b+q}p_bH\right)}{\sqrt{p_bH}p_bH^3} \\
\frac{\partial^2\Theta_1}{\partial H\partial p_b} &= \frac{1}{4}\frac{q(4A_0\sqrt{p_b+q\sqrt{p_b}H+B_0p_bH)}}{{p_b}^2H^2\sqrt{p_b+q}\sqrt{p_bH}} \\
\frac{\partial^2\Theta_2}{\partial {p_b}^2} &= \lambda\frac{\ln^2 p_b - 3\ln p_b + 1}{p_b\ln^2 p_b(\ln p_b - 1)^2}T^{comm} \\
\frac{\partial^2\Theta_2}{\partial H^2} &= \frac{\partial^2\Theta_2}{\partial H\partial p_b} = \frac{\partial^2\Theta_2}{\partial p_b\partial H} = 0
\end{aligned}
\tag{10}
$$

We further have $\frac{\partial^2\Theta_1}{\partial {p_b}^2} \geq 0$ and $\frac{\partial^2\Theta_1}{\partial {p_b}^2} \times \frac{\partial^2\Theta_1}{\partial H^2} - \frac{\partial^2\Theta_1}{\partial p_b\partial H} \times \frac{\partial^2\Theta_1}{\partial H\partial p_b} \geq 0$. Thus, both function $\Theta_1(\boldsymbol{\phi})$ and $\Theta_2(\boldsymbol{\phi})$ are positive and convex.

The Hessian matrix of the function Θ_1 can be described as $\begin{bmatrix} \frac{\partial^2\Theta_1}{\partial {p_b}^2} & \frac{\partial^2\Theta_1}{\partial p_b\partial H} \\ \frac{\partial^2\Theta_1}{\partial H\partial p_b} & \frac{\partial^2\Theta_1}{\partial H^2} \end{bmatrix}$, where we can find that both $\frac{\partial^2\Theta_1}{\partial {p_b}^2}$ and $\begin{vmatrix} \frac{\partial^2\Theta_1}{\partial {p_b}^2} & \frac{\partial^2\Theta_1}{\partial p_b\partial H} \\ \frac{\partial^2\Theta_1}{\partial H\partial p_b} & \frac{\partial^2\Theta_1}{\partial H^2} \end{vmatrix}$ are positive. In addition, the Hessian matrix of the function Θ_1 is positive-definite, and we can also easily obtain that the Hessian matrix of the function Θ_2 is positive semi-definite. Therefore, both function $\Theta_1(\boldsymbol{\phi})$ and $\Theta_2(\boldsymbol{\phi})$ are positive and convex.