VIII. Proof of Theorem

A. Additional Notation

For simplicity of notations, we denote the error of weight quantization $\boldsymbol{r}_n^k \triangleq Q_w\left(\boldsymbol{w}_n^{k+1}\right) - \boldsymbol{w}_n^{k+1}$, and the local "gradient" with weight quantization as $\hat{\boldsymbol{g}}_n^k \triangleq \nabla \widetilde{f}_n(\boldsymbol{w}_n^k) - \boldsymbol{r}_n^k/\eta$.

Inspired by the iterate analysis framework in we define the following virtual sequences:

$$\boldsymbol{u}_n^{k+1} = \boldsymbol{w}_n^k - \eta \hat{\boldsymbol{g}}_n^k, \tag{22}$$

$$\boldsymbol{w}_n^{k+1} = \begin{cases} \boldsymbol{u}_n^{k+1}, & k+1 \notin \mathcal{U}_H, \\ \boldsymbol{u}_n^{k'} - \sum_{n=1}^N p_n Q_g(\boldsymbol{\Delta}_n^{k'}), & k+1 \in \mathcal{U}_H. \end{cases} \tag{23}$$

Here, $k' = k+1-H$ is the last synchronization step and $\Delta_N^k = \boldsymbol{u}_n^{k'} - \boldsymbol{u}_n^{k+1}$ is the differences since the last synchronization. The following short-hand notation will be found useful in the convergence analysis of the proposed FL framework:

$$\bar{\boldsymbol{u}}^k = \sum_{n=1}^N p_n \boldsymbol{u}_n^k, \quad \bar{\boldsymbol{w}}_n^k = \sum_{n=1}^N p_n \boldsymbol{w}_n^k, \quad \hat{\boldsymbol{g}}^k = \sum_{n=1}^N p_n \hat{\boldsymbol{g}}_n^k \tag{24}$$

Thus, $\bar{\boldsymbol{u}}^{k+1} = \bar{\boldsymbol{w}}^k - \eta \hat{\boldsymbol{g}}^k$. Note that we can only obtain $\bar{\boldsymbol{w}}^{k+1}$ when $k+1 \in \mathcal{U}_H$. Further, due to the unbiased gradient quantization scheme, Q_g, no matter whether $k+1 \in \mathcal{U}_H$ or $k+1 \notin \mathcal{U}_H$, we always have $\mathbb{E}[\mathbb{E}_Q[\bar{\boldsymbol{w}}^{k+1}]] = \mathbb{E}[\bar{\boldsymbol{u}}^{k+1}]$.

B. Key Lemmas

Now, we give four important lemmas to convey our proof.

Lemma 1 (Bounding the weight quantization error)**.**

$$\mathbb{E}_Q\left[\left\|\boldsymbol{r}_n^k\right\|_2^2\right] \leq \eta\sqrt{d}\delta_w\tau. \tag{25}$$

Lemma 2. *According to the proposed algorithm the expected inner product between stochastic gradient and full batch gradient can be bounded with:*

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_Q\left[\langle\nabla F(\bar{\boldsymbol{w}}^k), \bar{\boldsymbol{u}}^{k+1} - \bar{\boldsymbol{w}}^k\rangle\right]\right] \\
&\leq + \frac{\eta L^2}{2}\sum_{n=1}^N p_n \mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^k - \boldsymbol{w}_n^k\right\|_2^2\right]\right] \\
&\quad - \frac{\eta}{2}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] - \frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{n=1}^N p_n \nabla F_n(\boldsymbol{w}_n^k)\right\|_2^2\right]
\end{aligned} \tag{26}$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_Q\left[\langle\nabla F(\bar{\boldsymbol{w}}^k), \bar{\boldsymbol{u}}^{k+1} - \bar{\boldsymbol{w}}^k\rangle\right]\right] \\
&= -\eta\mathbb{E}\left[\left\langle \nabla F(\bar{\boldsymbol{w}}^k), \mathbb{E}_Q[\hat{\boldsymbol{g}}^k]\right\rangle\right] \\
&= -\eta\mathbb{E}\left[\left\langle \nabla F(\bar{\boldsymbol{w}}^k), \sum_{n=1}^N p_n \nabla F_n(\boldsymbol{w}_n^k)\right\rangle\right] \\
&\overset{(a)}{=} -\frac{\eta}{2}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] - \frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{n=1}^N p_n \nabla F_n(\boldsymbol{w}_n^k)\right\|_2^2\right] \\
&\quad + \frac{\eta}{2}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k) - \sum_{n=1}^N p_n \nabla F_n(\boldsymbol{w}_n^k)\right\|_2^2\right] \\
&\overset{(b)}{\leq} -\frac{\eta}{2}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] - \frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{n=1}^N p_n \nabla F_n(\boldsymbol{w}_n^k)\right\|_2^2\right] \\
&\quad + \frac{\eta L^2}{2}\sum_{n=1}^N p_n \mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^k - \boldsymbol{w}_n^k\right\|_2^2\right]\right]
\end{aligned} \tag{27}$$

where (a) is due to $2 < a, b >= ||a||^2 + ||b||^2 + ||a-b||^2$ and $\mathbb{E}[\hat{\boldsymbol{g}}_n^k] = \nabla F_n(\boldsymbol{w}_n^k)$, and (b) follows from L-smoothness assumption and. ☐

Lemma 3 (Bounding the divergence)**.**

$$\begin{aligned}
&\sum_{k=0}^{K-1}\sum_{n=1}^N p_n \mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^k - \boldsymbol{w}_n^k\right\|_2^2\right]\right] \\
&\leq \frac{\eta^2 K H \sigma^2/M + 3\eta^2 K H^2 G^2 + \eta K H\sqrt{d}\delta_w\tau}{1 - 3\eta^2 L^2 H^2} \\
&\quad + \frac{3\eta^2 H^2}{1 - 3\eta^2 L^2 H^2}\sum_{i=0}^{K-1}\left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2
\end{aligned} \tag{28}$$

Proof. Recalling that at the synchronization step $k' \in \mathcal{U}_H$, $\boldsymbol{w}_n^{k'} = \bar{\boldsymbol{w}}^{k'}$ for all $n \in \mathcal{N}$. Therefore, for any $k \geq 0$, such that $k' \leq k \leq k' + H$, we get,

$$\begin{aligned}
A1_k &:= \sum_{n=1}^N p_n \mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^k - \boldsymbol{w}_n^k\right\|_2^2\right]\right] \\
&= \sum_{n=1}^N p_n \mathbb{E}\left[\mathbb{E}_Q\left[||(\bar{\boldsymbol{w}}^k - \bar{\boldsymbol{w}}^{k'}) - (\boldsymbol{w}_n^k - \bar{\boldsymbol{w}}^{k'})||_2^2\right]\right] \\
&\overset{(a)}{\leq} \sum_{n=1}^N p_n \mathbb{E}\left[\mathbb{E}_Q\left[||\boldsymbol{w}_n^k - \boldsymbol{w}_n^{k'}||_2^2\right]\right] \\
&= \sum_{n=1}^N p_n \mathbb{E}\left[\left\|\sum_{i=k'}^k\left(\eta\nabla\widetilde{f}_n(\boldsymbol{w}_n^i) - \boldsymbol{r}_n^i\right)\right\|_2^2\right] \\
&= \sum_{n=1}^N p_n \mathbb{E}\left[\left\|\sum_{i=k'}^k \eta\nabla\widetilde{f}_n(\boldsymbol{w}_n^i)\right\|_2^2 + \sum_{i=k'}^k \mathbb{E}_Q\left[||\boldsymbol{r}_n^i||_2^2\right]\right] \\
&\leq \eta^2\sum_{n=1}^N p_n \mathbb{E}\left[\left\|\sum_{i=k'}^{k'_H}\nabla\widetilde{f}_n(\boldsymbol{w}_n^i)\right\|_2^2\right] + H\eta\sqrt{d}\delta_w\tau,
\end{aligned} \tag{29}$$

where $k'_H = k' + H - 1$, (a) holds due to $\mathbb{E}[\|\sum_{i=1}^n a_i\|_2^2] = \sum_{i=1}^n \mathbb{E}[\|a_i\|_2^2]$ if $\mathbb{E}[a_i] = 0$. The last equality is due to Lemma 1.

We generalize the result from [] to upper-bound the first term in RHS of (29), (see the of Theorem 3 and its proof in appendix for the special case of $p_n = \frac{1}{N}$):

$$\eta^2 \sum_{n=1}^{N} p_n \mathbb{E}\left[\left\|\sum_{i=k'}^{k'_H}(\nabla \widetilde{f}_n(\boldsymbol{w}_n^i) - \nabla F_n(\boldsymbol{w}_n^i) + \nabla F_n(\boldsymbol{w}_n^i))\right\|_2^2\right]$$
$$\leq \eta^2 H \frac{\sigma^2}{M} + 3\eta^2 H^2 G^2 + 3\eta^2 L^2 H \sum_{i=k'}^{k'_H} A1_i$$
$$+ 3\eta^2 H \sum_{i=k'}^{k'_H} \left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2. \tag{30}$$

It follows that

$$\sum_{i=0}^{K-1} A1_i \leq \eta^2 K H \frac{\sigma^2}{M} + 3\eta^2 K H^2 G^2 + 3\eta^2 L^2 H^2 \sum_{i=0}^{K-1} A1_i$$
$$+ 3\eta^2 H^2 \sum_{i=0}^{K-1} \left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2 + \eta K H \sqrt{d}\delta_w \tau. \tag{31}$$

Suppose $1 - 3\eta^2 L^2 H^2 \geq 0$, we have

$$\sum_{i=0}^{K-1} A1_i \leq \frac{\eta^2 K H \sigma^2/M + 3\eta^2 K H^2 G^2 + \eta K H \sqrt{d}\delta_w \tau}{1 - 3\eta^2 L^2 H^2}$$
$$+ \frac{3\eta^2 H^2}{1 - 3\eta^2 L^2 H^2} \sum_{i=0}^{K-1} \left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2, \tag{32}$$

and the proof complete. □

Lemma 4. *According to the proposed algorithm the expected inner product between stochastic gradient and full batch gradient can be bounded with:*

$$\frac{L}{2}\mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{w}}^k\right\|_2^2\right]\right]$$
$$\leq \left(\frac{\eta^2 L \sigma^2}{2M} + \frac{\eta L \sqrt{d}\delta_w \tau}{2}\right)(\delta_g H + 1)\sum_{n=1}^{N} p_n^2$$
$$+ \frac{3\eta^2 L^3 \delta_g H}{2}\sum_{i=k'}^{k'_H} A1_i + \frac{3\eta^2 L \delta_g}{2} H \sum_{i=k'}^{k'_H}\left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2$$
$$+ \frac{\eta^2 L}{2}\mathbb{E}\left[\left\|\sum_{n=1}^{N} p_n \nabla F_n(\boldsymbol{w}_n^k)\right\|_2^2\right] + \frac{3\eta^2 L H^2 G^2 \delta_g}{2} \tag{33}$$

Proof.

$$\frac{L}{2}\mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{w}}^k\right\|_2^2\right]\right]$$
$$= \frac{L}{2}\mathbb{E}\left[\mathbb{E}_Q\left[\left\|\bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{u}}^{k+1} + \bar{\boldsymbol{u}}^{k+1} - \bar{\boldsymbol{w}}^k\right\|_2^2\right]\right]$$
$$= \mathbb{E}\left[\frac{L}{2}\mathbb{E}_Q\left[\left\|\sum_{n=1}^{N} p_n(Q(\widetilde{\Delta}_n^k) - \widetilde{\Delta}_n^k)\right\|_2^2\right] + \frac{L}{2}||\eta\hat{\boldsymbol{g}}^k||_2^2\right]$$
$$\leq \frac{L}{2}\sum_{n=1}^{N} p_n^2 \delta_g \mathbb{E}\left[||\boldsymbol{w}_n^k - \boldsymbol{w}^{k'}||_2^2\right] + \frac{\eta^2 L}{2}\mathbb{E}\left[||\hat{\boldsymbol{g}}^k||_2^2\right]$$
$$\leq \left(\frac{\eta^2 L \sigma^2}{2M} + \frac{\eta L \sqrt{d}\delta_w \tau}{2}\right)(\delta_g H + 1)\sum_{n=1}^{N} p_n^2$$
$$+ \frac{3\eta^2 L^3 \delta_g}{2} H \sum_{i=k'}^{k'_H} A1_i + \frac{3\eta^2 L \delta_g}{2} H \sum_{i=k'}^{k'_H}\left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2$$
$$+ \frac{3\eta^2 L H^2 G^2 \delta_g}{2} + \frac{\eta^2 L}{2}\mathbb{E}\left[\left\|\sum_{n=1}^{N} p_n \nabla F_n(\boldsymbol{w}_n^k)\right\|_2^2\right] \tag{34}$$

□

C. Main Results

Under the L-smooth assumption of F, we have,

$$\mathbb{E}\left[F(\bar{\boldsymbol{w}}^{k+1}) - F(\bar{\boldsymbol{w}}^k)\right]$$
$$\leq \mathbb{E}\left[\langle \nabla F(\bar{\boldsymbol{w}}^k), \bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{w}}^k\rangle\right] + \frac{L}{2}\mathbb{E}\left[\left\|\bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{w}}^k\right\|_2^2\right]$$
$$= \mathbb{E}\left[\langle \nabla F(\bar{\boldsymbol{w}}^k), \bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{u}}^{k+1} + \bar{\boldsymbol{u}}^{k+1} - \bar{\boldsymbol{w}}^k\rangle\right]$$
$$+ \frac{L}{2}\mathbb{E}\left[\left\|\bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{w}}^k\right\|_2^2\right]$$
$$\overset{(a)}{\leq} \mathbb{E}\left[\langle \nabla F(\bar{\boldsymbol{w}}^k), \bar{\boldsymbol{u}}^{k+1} - \bar{\boldsymbol{w}}^k\rangle + \frac{L}{2}\left\|\bar{\boldsymbol{w}}^{k+1} - \bar{\boldsymbol{w}}^k\right\|_2^2\right] \tag{35}$$

where (a) is $\mathbb{E}[\mathbb{E}_Q[\bar{\boldsymbol{w}}^{k+1}]] = \mathbb{E}[\bar{\boldsymbol{u}}^{k+1}]$. We use Lemma 1-4 to upper bound the RHS of (35) and set $\eta L \leq 1$, which gets,

$$\mathbb{E}\left[F(\bar{\boldsymbol{w}}^{k+1}) - F(\bar{\boldsymbol{w}}^k)\right]$$
$$\leq -\frac{\eta}{2}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] + \frac{\eta L^2}{2} A1_k + \frac{3\eta^2 L^3 H \delta_g}{2}\sum_{i=k'}^{k'_H} A1_i$$
$$+ \left(\frac{\eta^2 L \sigma^2}{2M} + \frac{\eta L \sqrt{d}\delta_w \tau}{2}\right)(\delta_g H + 1)\sum_{n=1}^{N} p_n^2$$
$$+ \frac{3\eta^2 L H^2 G^2 \delta_g}{2} + \frac{3\eta^2 L \delta_g}{2} H \sum_{i=k'}^{k'_H}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^i)\right\|_2^2\right]. \tag{36}$$

Summing up for all K communication rounds and rearranging the terms gives,

$$\mathbb{E}\left[F(\bar{\boldsymbol{w}}^K) - F(\bar{\boldsymbol{w}}^0)\right]$$
$$\leq -\frac{\eta C_1}{2}\sum_{k=0}^{K-1}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] + \frac{\eta L^2 C_2}{2}\sum_{k=0}^{K-1} A1_k$$
$$+ \frac{\eta L K p}{2}\left(\frac{\eta\sigma^2}{M} + \sqrt{d}\delta_w \tau\right)(\delta_g H + 1) + \frac{3\eta^2 L K H \delta_g}{2} G^2 \tag{37}$$

where $p = \sum_{n=1}^{N} p_n^2$ and $C_1 = 1 - 3\eta LH\delta_g$ and $C_2 = 1 + 3\eta L\delta_g H$. Plugging Lemma 4 into (37), if $C_1' = C_1 - \frac{3\eta^2 H^2 (L^2 + 3\eta L^3 \delta_g H)}{1 - 3\eta^2 L^2 H^2} \geq 0$, we have,

$$\begin{aligned}
&\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] \\
&\leq \frac{2\mathbb{E}\left[F(\bar{\boldsymbol{w}}^0) - F(\bar{\boldsymbol{w}}^K)\right]}{\eta C_1' K} + \frac{\eta^2 L^2 C_2 H \sigma^2}{C_1' M(1 - 3\eta^2 L^2 H^2)} \\
&+ \frac{3\eta^2 L^2 C_2 H^2 G^2}{C_1'(1 - 3\eta^2 L^2 H^2)} + \frac{\eta L^2 C_2 H \sqrt{d}\delta_w \tau}{C_1'(1 - 3\eta^2 L^2 H^2)} + \frac{3\eta LKH}{C_1'} G^2 \\
&+ \frac{\eta L \sigma^2 (\delta_g H + 1)}{C_1' M} + \frac{L(\delta_g H + 1)\sqrt{d}\delta_w \tau}{C_1'}
\end{aligned} \tag{38}$$

If we set $\eta = \sqrt{MN/K}$ and

$$\eta LH\delta_g \geq \frac{\eta^2 H^2 (L^2 + 3\eta L^3 \delta_g H)}{1 - 3\eta^2 L^2 H^2}, \tag{39}$$

we can get the $1/C_1' \leq 2$. Thus,

$$\begin{aligned}
&\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\left[\left\|\nabla F(\bar{\boldsymbol{w}}^k)\right\|_2^2\right] \\
&\leq \frac{4\mathbb{E}\left[F(\bar{\boldsymbol{w}}^0) - F(\bar{\boldsymbol{w}}^K)\right]}{\sqrt{MNK}} + \frac{2L\sigma^2 H(2\delta_g + 1)}{\sqrt{MNK}} \\
&\quad + \frac{12MLH\delta_g G^2}{\sqrt{MNK}} + 2L(2\delta_g + 1)\sqrt{d}\delta_w \tau.
\end{aligned} \tag{40}$$