Energy-Efficient Hierarchical Collaborative Learning over LEO Satellite Constellations - Proof of Theorem 1

Long Luo, Chi Zhang, Hongfang Yu, Zonghang Li, Gang Sun and Shouxi Luo
University of Electronic Science and Technology of China, Southwest Jiaotong University.

Assumption 1 (Lipschitz smooth gradient)**.** *The loss function $\nabla F\left(\boldsymbol{\omega}\right)$ is differentiable and L-smooth, which ensures that*

$$\|\nabla F\left(\boldsymbol{\omega}_1\right)-\nabla F\left(\boldsymbol{\omega}_2\right)\| \leq L\|\boldsymbol{\omega}_1-\boldsymbol{\omega}_2\|,\ \forall \boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \tag{1}$$

Assumption 2 (Unbiased gradient estimator)**.** *Let ξ be a random sample from the local dataset and the estimated gradient is unbiased as*

$$\mathbb{E}\left[\nabla F\left(\boldsymbol{\omega};\xi\right)\right]=\nabla F\left(\boldsymbol{\omega}\right) \tag{2}$$

Assumption 3 (Bounded gradient variance)**.** *We assume the existence of a constant σ such that the variance of the stochastic gradient is bounded by*

$$\mathbb{E}\left[\|\nabla F\left(\boldsymbol{\omega};\xi\right)-\nabla F\left(\boldsymbol{\omega}\right)\|^2\right] \leq \sigma^2 \tag{3}$$

Before stating the proof of Theorem 1 of FedAAC, we first mention the following three lemmas.

Lemma 1. *According to Assumption 1, we define the global gradient based on the local stochastic gradient as $g^h=\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\cdot TopK(\sum_{t=1}^{k_1 I}\nabla F_j(\boldsymbol{\omega}_j^{t,r,h}))$, then the expected inner product between full batch gradient and stochastic gradient can be bounded with:*

$$\begin{aligned}
&\mathbb{E}\left\langle\nabla f\left(\boldsymbol{\omega}^h\right), g^h\right\rangle\\
&\leq \frac{L^2}{2N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\sum_{t=1}^{k_1 I}\|\boldsymbol{\omega}^h-\boldsymbol{\omega}_j^{t,r,h}\|^2\\
&-\frac{1}{2N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\sum_{t=1}^{k_1 I}\|\nabla f\left(\boldsymbol{\omega}^h\right)^2\|^2\\
&-\frac{1}{2N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\sum_{t=1}^{k_1 I}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2
\end{aligned} \tag{4}$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\left\langle \nabla f\left(\boldsymbol{\omega}^h\right), g^h\right\rangle \\
&= \frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\sum_{t=1}^{k_1 I}\left\langle \nabla f\left(\boldsymbol{\omega}^h\right), \nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right\rangle \\
&\overset{①}{=} \frac{1}{2N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\sum_{t=1}^{k_1 I}\left[-\|\nabla f\left(\boldsymbol{\omega}^h\right)\|^2 - \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 + \|\nabla f\left(\boldsymbol{\omega}^h\right) - \nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2\right] \\
&\overset{②}{\leq} \frac{1}{2N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\sum_{t=1}^{k_1 I}\left[-\|\nabla f\left(\boldsymbol{\omega}^h\right)\|^2 - \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 + L^2\|\boldsymbol{\omega}^h - \boldsymbol{\omega}_j^{t,r,h}\|^2\right]
\end{aligned}
\tag{5}
$$

where ① is due to $2\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2$, and ② based on Assumption 1. □

Lemma 2. *Under Assumption 2 and 3 and in accordance with the TopK sparsification property* $\mathbb{E}\|\boldsymbol{\omega} - TopK_\gamma(\boldsymbol{\omega})\|^2 \leq (1-\gamma)\|\boldsymbol{\omega}\|^2$*, we have the following bound:*

$$
\mathbb{E}_{TopK}\|g^h\|^2 \leq \frac{(2-\gamma)}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\sum_{t=1}^{k_1 I}\left(\sigma^2 + k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2\right) \tag{6}
$$

Proof. Let $\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)$ be $\tilde{F}_j$ and $TopK\left(\tilde{F}_j\right)$ be $\tilde{F}_K$. Then we have:

$$
\begin{aligned}
&\mathbb{E}_{TopK}\|\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h TopK\left(\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right)\|^2 \\
&\overset{①}{=} \mathbb{E}_{TopK}\|\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\nabla\tilde{F}_K - \frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\mathbb{E}_{TopK}\left[\nabla\tilde{F}_K\right]\|^2 \\
&+ \|\mathbb{E}_{TopK}\left[\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\nabla\tilde{F}_K\right]\|^2 \\
&\overset{②}{=} \frac{1}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\|\nabla\tilde{F}_K - \nabla\tilde{F}_j\|^2 + \|\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\nabla\tilde{F}_j\|^2 \\
&\overset{③}{\leq} \frac{1-\gamma}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\|\nabla\tilde{F}_j\|^2 + \|\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h\nabla\tilde{F}_j\|^2 \\
&\leq \frac{1-\gamma}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\|Var\left(\nabla\tilde{F}_j\right) + \nabla\tilde{F}_j\|^2 \\
&+ \frac{1}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2 Var\left(\nabla\tilde{F}_j\right) + \frac{1}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\|\nabla\tilde{F}_j\|^2
\end{aligned}
\tag{7}
$$

where ① is due to $\mathbb{E}\left[\|x\|^2\right] = Var(x) + \|\mathbb{E}[x]\|^2$ and follows Assumption 2, ② is due to $\mathbb{E}_{TopK}\left[\nabla\tilde{F}_K\right] = \nabla\tilde{F}_j$ and ③ follows the TopK sparsification property.

Then, based on $Var\left(\nabla\tilde{F}_j\right) = Var\left(\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right) = \sum_{t=1}^{k_1 I}Var\left(\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right) \leq k_1 I\sigma^2$, we have:

$$
\begin{aligned}
&\mathbb{E}_{TopK}\|\frac{1}{N}\sum_{i=1}^{M}\alpha_i^h\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_j^h TopK\left(\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right)\|^2 \\
&\leq \frac{1-\gamma}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\left[k_1 I\sigma^2+\|\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2\right] \\
&+\frac{1}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2 k_1 I\sigma^2+\frac{1}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\|\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 \\
&\overset{①}{\leq} \frac{2-\gamma}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2 k_1 I\sigma^2+\frac{2-\gamma}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\sum_{t=1}^{k_1 I}k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 \\
&= \frac{2-\gamma}{N^2}\sum_{i=1}^{M}\left(\alpha_i^h\right)^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\left(\alpha_j^h\right)^2\sum_{t=1}^{k_1 I}\left[\sigma^2+k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2\right]
\end{aligned}
\tag{8}
$$

where ① is because $\|\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 \leq k_1 I\sum_{t=1}^{k_1 I}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2$. □

Lemma 3. *Under Assumptions 3 and according to [1], we have the following bound:*

$$
\begin{aligned}
\mathbb{E}\|\boldsymbol{\omega}^h-\boldsymbol{\omega}_j^{t,r,h}\|^2 &\leq \eta^2\left[(2-\gamma)\left(\frac{k_2-1}{Nk_2}\right)^2+\frac{(k_1 I-1)^2}{Nk_1^2 I^2 k_2^2}\right] \\
&\cdot\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\sum_{t=1}^{k_1 I}\left[k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2+\sigma^2\right]
\end{aligned}
\tag{9}
$$

Proof. According to [1], we have the following relationship:

$$
\frac{1}{N}\sum_{j=1}^{N}\|\nabla F_j\left(\boldsymbol{\omega}\right)-\nabla f\left(\boldsymbol{\omega}\right)\|^2=\sum_{i=1}^{M}\frac{N_i}{N}\|\nabla f_i\left(\boldsymbol{\omega}\right)-\nabla f\left(\boldsymbol{\omega}\right)\|^2+\sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\nabla F_j\left(\boldsymbol{\omega}\right)-\nabla f_i\left(\boldsymbol{\omega}\right)\|^2 \tag{10}
$$

Under the Law of Large Numbers, we have $\mathbb{E}\left[\boldsymbol{\omega}^h-\eta\nabla f\left(\boldsymbol{\omega}^h\right)\right]=\mathbb{E}\left[\boldsymbol{\omega}_j^{t,r,h}-\eta\nabla f\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right]$, so that we have $\|\boldsymbol{\omega}^h-\boldsymbol{\omega}_j^{t,r,h}\|^2=\eta^2\|\nabla F_j\left(\boldsymbol{\omega}\right)-\nabla f\left(\boldsymbol{\omega}\right)\|^2$, and then we can get:

$$
\frac{1}{N}\sum_{j=1}^{N}\|\boldsymbol{\omega}^h-\boldsymbol{\omega}_j^{t,r,h}\|^2=\eta^2\sum_{i=1}^{M}\frac{N_i}{N}\|\nabla f_i\left(\boldsymbol{\omega}\right)-\nabla f\left(\boldsymbol{\omega}\right)\|^2+\eta^2\sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\nabla F_j\left(\boldsymbol{\omega}\right)-\nabla f_i\left(\boldsymbol{\omega}\right)\|^2 \tag{11}
$$

Thus, we can prove the Lemma 3 by

$$
\begin{aligned}
(1)\quad &\sum_{i=1}^{M}\frac{N_i}{N}\|\nabla f_i\left(\boldsymbol{\omega}\right)-\nabla f\left(\boldsymbol{\omega}\right)\|^2 \\
&=\sum_{i=1}^{M}\frac{N_i}{N}\|\frac{1}{N_i}\sum_{j\in\mathcal{K}_i}TopK\left(\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right)-\frac{1}{N}\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}TopK\left(\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right)\|^2 \\
&=\sum_{i=1}^{M}\frac{N_i}{N}\|\frac{k_2-1}{Nk_2}\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\sum_{t=1}^{k_1 I}TopK\left(\sum_{t=1}^{k_1 I}\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\right)\|^2
\end{aligned}
$$

$$\leq \sum_{i=1}^{M} \frac{N_i}{N} \left\{ (2-\gamma) \left(\frac{k_2-1}{Nk_2} \right)^2 \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \left[Var\left(\nabla \tilde{F}_j \right) + \| \nabla \tilde{F}_j \|^2 \right] \right\}$$

$$\leq (2-\gamma) \left(\frac{k_2-1}{Nk_2} \right)^2 \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \sum_{t=1}^{k_1 I} \left(\sigma^2 + k_1 I \| \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \|^2 \right)$$

(2) $$\sum_{i=1}^{M} \frac{1}{N} \sum_{j \in \mathcal{K}_i} \| \nabla F_j (\boldsymbol{\omega}) - \nabla f_i (\boldsymbol{\omega}) \|^2$$

$$= \sum_{i=1}^{M} \frac{1}{N} \sum_{j \in \mathcal{K}_i} \| \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) - \frac{1}{N_i} \sum_{j \in \mathcal{K}_i} TopK \left(\sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \|^2$$

$$= \sum_{i=1}^{M} \frac{1}{N} \sum_{j \in \mathcal{K}_i} \| \frac{1}{Nk_1 I k_2} \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) - \frac{k_1 I}{Nk_1 I k_2} \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} TopK \left(\sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \|^2$$

$$\leq \sum_{i=1}^{M} \frac{1}{N} \sum_{j \in \mathcal{K}_i} \| \frac{k_1 I - 1}{Nk_1 I k_2} \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} TopK \left(\sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \|^2$$

$$\leq \sum_{i=1}^{M} \frac{1}{N} (2-\gamma) \left(\frac{k_1 I - 1}{Nk_1 I k_2} \right)^2 \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \sum_{t=1}^{k_1 I} \left[k_1 I \| \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \|^2 + \sigma^2 \right]$$

$$\leq N \left(\frac{k_1 I - 1}{Nk_1 I k_2} \right)^2 \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \sum_{t=1}^{k_1 I} \left[k_1 I \| \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \|^2 + \sigma^2 \right] \tag{12}$$

Combining (1) and (2), we can obtain:

$$\frac{1}{N} \sum_{j=1}^{N} \| \boldsymbol{\omega}^h - \boldsymbol{\omega}_j^{t,r,h} \|^2 \leq \eta^2 \left[(2-\gamma) \left(\frac{k_2-1}{Nk_2} \right)^2 + \frac{(k_1 I - 1)^2}{N (k_1 I k_2)^2} \right] \sum_{i=1}^{M} \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \sum_{t=1}^{k_1 I} \left[k_1 I \| \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \|^2 + \sigma^2 \right] \tag{13}$$

□

Now, according to Assumption 1 and the SGD model update formula we can obtain:

$$\begin{aligned} \mathbb{E} \left[f \left(\boldsymbol{\omega}^{h+1} \right) \right] &= \mathbb{E} \left[f \left(\boldsymbol{\omega}^h - \eta \frac{1}{N} \sum_{i=1}^{M} \alpha_i^h \sum_{r=1}^{k_2^h} \sum_{j \in \mathcal{K}_i} \alpha_j^h TopK \left(\sum_{t=1}^{k_1^h I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \right) \right] \\ &\leq \mathbb{E} \left[f \left(\boldsymbol{\omega}^h - \eta \frac{1}{N} \sum_{i=1}^{M} \alpha_2 \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \alpha_1 TopK \left(\sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \right) \right] \\ &\leq \mathbb{E} \left[f \left(\boldsymbol{\omega}^h \right) \right] - \eta \mathbb{E} \left[\left\langle \nabla f \left(\boldsymbol{\omega}^h \right), \frac{1}{N} \sum_{i=1}^{M} \alpha_2 \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \alpha_1 TopK \left(\sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \right\rangle \right] \\ &+ \frac{\eta^2 L}{2} \mathbb{E} \left[\| \frac{1}{N} \sum_{i=1}^{M} \alpha_2 \sum_{r=1}^{k_2} \sum_{j \in \mathcal{K}_i} \alpha_1 TopK \left(\sum_{t=1}^{k_1 I} \nabla F_j \left(\boldsymbol{\omega}_j^{t,r,h} \right) \right) \|^2 \right] \end{aligned} \tag{14}$$

where $k_1 = \max\left\{k_1^h\right\}$, $k_2 = \max\left\{k_2^h\right\}$, $\alpha_1 = \max\left\{\alpha_j^h\right\}$, $\alpha_2 = \max\left\{\alpha_i^h\right\}$.

Then, based on the Lemma 1, Lemma 2, Lemma 3 and (14), we can obtain:

$$
\begin{aligned}
&\mathbb{E}\left[f\left(\boldsymbol{\omega}^{h+1}\right) - f\left(\boldsymbol{\omega}^{h}\right)\right] \\
&\leq \frac{\eta}{2N}\sum_{i=1}^{M}\alpha_2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_1\sum_{t=1}^{k_1 I}\left[-\|\nabla f\left(\boldsymbol{\omega}^{h}\right)\|^2 - \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 + L^2\|\boldsymbol{\omega}^{h} - \boldsymbol{\omega}_j^{t,r,h}\|^2\right] \\
&+ \frac{\eta^2 L}{2}\frac{(2-\gamma)}{N^2}\sum_{i=1}^{M}\alpha_2^2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_1^2\sum_{t=1}^{k_1 I}\left[k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 + \sigma^2\right] \\
&\leq \frac{\eta}{2N}\sum_{i=1}^{M}\alpha_2\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\alpha_1\sum_{t=1}^{k_1 I}\left[-\|\nabla f\left(\boldsymbol{\omega}^{h}\right)\|^2 - \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2\right] \\
&+ \frac{L^2\eta^3\alpha_1\alpha_2}{2N}\frac{\left[(2-\gamma)k_1^2 I^2 (k_2-1)^2 + N(k_1 I - 1)^2\right]}{N k_1 I k_2}\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\sum_{t=1}^{k_1 I}\left[k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 + \sigma^2\right] \\
&+ \frac{L\eta^2(2-\gamma)}{2N^2}\alpha_1^2\alpha_2^2\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\sum_{t=1}^{k_1 I}\left[k_1 I\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 + \sigma^2\right] \\
&= -\frac{\eta\alpha_1\alpha_2 k_1 I k_2}{2}\|\nabla f\left(\boldsymbol{\omega}^{h}\right)\|^2 + \frac{L^2\eta^3\alpha_1\alpha_2\left[(2-\gamma)k_1^2 I^2(k_2-1)^2 + N(k_1 I-1)^2\right]}{2N}\sigma^2 + \frac{L\eta^2\alpha_1^2\alpha_2^2(2-\gamma)k_1 I k_2}{2N}\sigma^2 \\
&- \left\{\frac{\eta\alpha_1\alpha_2}{2N} - \frac{L^2\eta^3\alpha_1\alpha_2\left[(2-\gamma)k_1^2 I^2(k_2-1)^2 + N(k_1 I-1)^2\right]}{2N^2 k_2} - \frac{L\eta^2\alpha_1^2\alpha_2^2(2-\gamma)k_1 I}{2N^2}\right\} \\
&\cdot\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\sum_{t=1}^{k_1 I}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 \\
&\leq -\frac{\eta\alpha_1\alpha_2 k_1 I k_2}{2}\|\nabla f\left(\boldsymbol{\omega}^{h}\right)\|^2 + \frac{L^2\eta^3\alpha_1\alpha_2(2-\gamma)k_1^2 I^2 k_2^2}{2N}\sigma^2 + \frac{L^2\eta^3\alpha_1\alpha_2 k_1^2 I^2}{2}\sigma^2 + \frac{L\eta^2\alpha_1^2\alpha_2^2(2-\gamma)k_1 I k_2}{2N}\sigma^2 \\
&- \frac{\eta\alpha_1\alpha_2}{2N}\left\{1 - \frac{L^2\eta^2\left[(2-\gamma)k_1^2 I^2 k_2^2 + N k_1^2 I^2\right]}{N k_2} - \frac{L\eta\alpha_1\alpha_2(2-\gamma)k_1 I}{N}\right\}\sum_{i=1}^{M}\sum_{r=1}^{k_2}\sum_{j\in\mathcal{K}_i}\sum_{t=1}^{k_1 I}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,r,h}\right)\|^2 \\
&\leq -\frac{\eta\alpha_1\alpha_2 k_1 I k_2}{2}\|\nabla f\left(\boldsymbol{\omega}^{h}\right)\|^2 + \frac{L^2\eta^3\alpha_1\alpha_2(2-\gamma)k_1^2 I^2 k_2^2}{2N}\sigma^2 + \frac{L^2\eta^3\alpha_1\alpha_2 k_1^2 I^2}{2}\sigma^2 + \frac{L\eta^2\alpha_1^2\alpha_2^2(2-\gamma)k_1 I k_2}{2N}\sigma^2
\end{aligned}
\tag{15}
$$

where k_1, k_2, α_1, α_2 and η satisfy:

$$
\frac{L^2\eta^2(2-\gamma)k_1^2 I^2 k_2}{N} + \frac{L^2\eta^2 k_1^2 I^2}{k_2} + \frac{L\eta\alpha_1\alpha_2(2-\gamma)k_1 I}{N} \leq 1 \tag{16}
$$

Next, based on (15), we can obtain:

$$
\frac{1}{H}\sum_{h=1}^{H}\|\nabla f\left(\boldsymbol{\omega}^{h}\right)\|^2 \leq \frac{2\left[f\left(\boldsymbol{\omega}^{0}\right) - f\left(\boldsymbol{\omega}^{*}\right)\right]}{\eta\alpha_1\alpha_2 k_1 I k_2 H} + \frac{L^2\eta^2(2-\gamma)k_1 I k_2}{N}\sigma^2 + \frac{L^2\eta^2 k_1 I}{k_2}\sigma^2 + \frac{L\eta\alpha_1\alpha_2(2-\gamma)}{N}\sigma^2 \tag{17}
$$

Thus, the proof is completed.

References

[1] Jiayi Wang, Shiqiang Wang, Rong-Rong Chen, et al. "Demystifying why local aggregation helps: Convergence analysis of hierarchical SGD". In: *AAAI*. Vol. 36. 8. 2022, pp. 8548–8556.