Appendix A

Proof of lemma 1: Firstly, we analyze the impact of quantized components on bitstream transmission. We omit the user subscript m when it does not cause ambiguity. Let δ_i be represented as a bit sequence $[b_1, b_2, \cdots, b_{s_{(b)}}]$, where the error probability of each bit is $P_e^1, P_e^2, \cdots, P_e^{s_{(b)}}$. Then, the transmission error probability of $q(\delta_i^m)$ satisfies

$$\begin{aligned}
&\mathbb{E}\left[\left(\hat{q}(\delta_i^m) - q(\delta_i^m)\right)^2\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\hat{q}(\delta_i^m) - q(\delta_i^m)\right)^2 \Big| e_{b_1}\right]\right]\\
&\le P_e^1 \cdot (1-(-1))^2 +\\
&\quad (1-P_e^1)\mathbb{E}\left[\left(\hat{q}(\delta_i^m) - q(\delta_i^m)\right)^2 \Big| e_{b_1} = 0\right]\\
&= P_e^1 \cdot 2^2 +\\
&\quad (1-P_e^1)\mathbb{E}\left[\mathbb{E}\left[\left(\hat{q}(\delta_i^m) - q(\delta_i^m)\right)^2 \Big| e_{b_1} = 0, e_{b_2}\right]\right]\\
&\le P_e^1 \cdot 2^2 + (1-P_e^1)\Big(P_e^2 \cdot (2^{-1})^2 +\\
&\quad \mathbb{E}\left[\left(\hat{q}(\delta_i^m) - q(\delta_i^m)\right)^2 \Big| e_{b_1} = 0, e_{b_2} = 0\right]\Big)\\
&\le P_e^1 \cdot 2^2 + (1-P_e^1) \cdot \sum_{i=2}^{s_{(b)}} P_e^i \cdot 2^{-2(i-1)},
\end{aligned} \tag{22}$$

where $e_{b_i} = 1$ indicates that b_i bit is transmitted incorrectly.

Next, we analyze the bitstream transmission error of the parameter update vector $\boldsymbol{\delta}_t$ during the t-th round of upload

$$\begin{aligned}
&\mathbb{E}\left\|\hat{\boldsymbol{\delta}}_t - \boldsymbol{\delta}_t\right\|^2\\
&= \mathbb{E}\left[\left(\hat{\boldsymbol{\delta}}_t - Q(\boldsymbol{\delta}_t)\right) + \left(Q(\boldsymbol{\delta}_t) - \boldsymbol{\delta}_t\right)\right]^2\\
&= \mathbb{E}\left[\left\|\hat{\boldsymbol{\delta}}_t - Q(\boldsymbol{\delta}_t)\right\|^2\right] + \mathbb{E}\left[\left\|Q(\boldsymbol{\delta}_t) - \boldsymbol{\delta}_t\right\|\right]^2,
\end{aligned} \tag{23}$$

where (23) comes from the unbiasedness of the quantizer. For the first term in (23), we have

$$\begin{aligned}
\mathbb{E}\left[\left\|\hat{\boldsymbol{\delta}}_t - Q(\boldsymbol{\delta}_t)\right\|^2\right] &= \|\boldsymbol{\delta}_t\|^2 \,\mathbb{E}\left[\|\hat{q}(\boldsymbol{\delta}_t) - q(\boldsymbol{\delta}_t)\|^2\right]\\
&= \|\boldsymbol{\delta}_t\|^2 \sum_{i=1}^{d} \mathbb{E}\left[\left(\hat{q}(\delta_{t,i}) - q(\delta_{t,i})\right)^2\right]\\
&\le \|\boldsymbol{\delta}_t\|^2 \, dD_t,
\end{aligned} \tag{24}$$

where (24) comes from (22). For the second term in (23), we have [11]

$$\mathbb{E}\left[\left\|Q(\boldsymbol{\delta}_t) - \boldsymbol{\delta}_t\right\|\right]^2 \le \|\boldsymbol{\delta}_t\|^2 \min\left(\frac{d}{s^2}, \frac{\sqrt{d}}{s}\right), \tag{25}$$

Plugging (24) and (25) into (23), we can obtain the result. ■

Appendix B

Proof of lemma 2: After the t-th communication, the local stochastic gradient of the client m at round $\tau = \{0, 1, \cdots, K\}$ is $\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)$, and we have

$$\begin{aligned}
f\left(\tilde{\boldsymbol{w}}_{t+1}\right) &= f\left(\sum_{m\in[M]} p_m \boldsymbol{w}_{t+1}^m\right)\\
&= f\left(\sum_{m\in[M]} p_m \left(\boldsymbol{w}_t - \eta \sum_{\tau=0}^{K-1} \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)\right)\right)\\
&= f\left(\boldsymbol{w}_t - \eta \sum_{m\in[M]} \sum_{\tau=0}^{K-1} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)\right)\\
&\le f\left(\boldsymbol{w}_t\right) - \left\langle \nabla f\left(\boldsymbol{w}_t\right), \eta \sum_{m\in[M]} \sum_{\tau=0}^{K-1} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)\right\rangle\\
&\quad + \frac{L\eta^2}{2}\left\|\sum_{m\in[M]} \sum_{\tau=0}^{K-1} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)\right\|^2,
\end{aligned} \tag{26}$$

where the inequality comes from the L-smooth property.

We consider the second term first and take the expectation of $\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)$. After transforming the coefficient, we have

$$\begin{aligned}
&\mathbb{E}\left\langle \nabla f\left(\boldsymbol{w}_t\right), \eta \sum_{m\in[M]} \sum_{\tau=0}^{K-1} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m, \boldsymbol{\xi}_{t,\tau+1}^m\right)\right\rangle\\
&= \eta \sum_{\tau=0}^{K-1} \mathbb{E}\left\langle \nabla f\left(\boldsymbol{w}_t\right), \sum_{m\in[M]} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau+1}^m\right)\right\rangle.
\end{aligned} \tag{27}$$

Each term inside the expectation satisfies

$$\begin{aligned}
&\left\langle \nabla f\left(\boldsymbol{w}_t\right), \sum_{m\in[M]} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m\right)\right\rangle\\
&= \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 + \frac{1}{2}\left\|\sum_{m\in[M]} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m\right)\right\|^2\\
&\quad - \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right) - \sum_{m\in[M]} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m\right)\right\|^2\\
&\ge \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 - \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right) - \sum_{m\in[M]} p_m \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m\right)\right\|^2\\
&= \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 -\\
&\quad \frac{1}{2}\left\|\sum_{m\in[M]} p_m \left(\nabla f_m\left(\boldsymbol{w}_t\right) - \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m\right)\right)\right\|^2\\
&\ge \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 -\\
&\quad \frac{1}{2}M \sum_{m\in[M]} p_m^2 \left\|\nabla f_m\left(\boldsymbol{w}_t\right) - \nabla f_m\left(\boldsymbol{w}_{t,\tau}^m\right)\right\|^2 \qquad (28)\\
&\ge \frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 - \frac{1}{2}ML^2 \sum_{m\in[M]} p_m^2 \left\|\boldsymbol{w}_t - \boldsymbol{w}_{t,\tau}^m\right\|^2, \qquad (29)
\end{aligned}$$

where (28) comes from the Cauchy inequality; (29) comes from the L-smooth condition in Assumption 1. After taking expectation over a part of the second term on (29), we have

$$\mathbb{E}\left[\left\|\boldsymbol{w}_t - \boldsymbol{w}_{t,\tau}^m\right\|^2\right]$$

$$= \mathbb{E}\left[\left\|\eta\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m,\boldsymbol{\xi}_{t,j+1}^m\right)\right\|^2\right]$$

$$= \eta^2\mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m,\boldsymbol{\xi}_{t,j+1}^m\right) - \nabla f_m\left(\boldsymbol{w}_{t,j}^m\right) + \nabla f_m\left(\boldsymbol{w}_{t,j}^m\right)\right\|^2\right]$$

$$= \eta^2\mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m,\boldsymbol{\xi}_{t,j+1}^m\right) - \nabla f_m\left(\boldsymbol{w}_{t,j}^m\right)\right\|^2\right] + \eta^2\mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m\right)\right\|^2\right]$$

$$\leq \eta^2\tau^2\sigma^2 + \eta^2\mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m\right)\right\|^2\right] \tag{30}$$

$$= \eta^2\tau^2\sigma^2 + \eta^2\mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m\right) - \nabla f_m\left(\boldsymbol{w}_t\right) + \nabla f_m\left(\boldsymbol{w}_t\right) - \nabla f\left(\boldsymbol{w}_t\right) + \nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]$$

$$\leq \eta^2\tau^2\sigma^2 + 3\eta^2\left(\mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_{t,j}^m\right) - \nabla f_m\left(\boldsymbol{w}_t\right)\right\|^2\right] + \mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f_m\left(\boldsymbol{w}_t\right) - \nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right] + \mathbb{E}\left[\left\|\sum_{j=0}^{\tau-1}\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right) \tag{31}$$

$$\leq \eta^2\tau^2\sigma^2 + 3\eta^2\left(L^2\tau\sum_{j=0}^{\tau-1}\mathbb{E}\left[\left\|\boldsymbol{w}_t - \boldsymbol{w}_{t,j}^m\right\|^2\right] + \tau^2G_t^2 + \tau^2\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right) \tag{32}$$

$$\leq \eta^2\tau^2\sigma^2 + 3\eta^2\left(L^2\tau A_t + \tau^2G_t^2 + \tau^2\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right). \tag{33}$$

Here, (30) is derived from the gradient variance bound in hypothesis 1, (31) is derived from Cauchy's inequality, (32) is derived from the heterogeneous data property in hypothesis 1, and (33) is derived from the definition $A_t \triangleq \sum_{\tau=0}^{K-1}\mathbb{E}\left[\left\|\boldsymbol{w}_t - \boldsymbol{w}_{t,\tau}^m\right\|^2\right]$. Adding both sides of inequality (33), we get

$$A_t \leq \sum_{\tau=0}^{K-1}\eta^2\tau^2\sigma^2 + 3\eta^2\cdot \left(L^2\tau A_t + \tau^2G_t^2 + \tau^2\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right)$$

$$= \eta^2\left(\frac{K(K-1)(2K-1)}{6}\left(\sigma^2 + 3G_t^2 + 3\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right) + \frac{3K(K-1)}{2}L^2A_t\right). \tag{34}$$

Set $0 < \eta < \frac{1}{\sqrt{3}KL}$, then we have

$$A_t \leq \frac{\eta^2}{1 - \frac{3K(K-1)}{2}L^2\eta^2}\frac{K(K-1)(2K-1)}{6}\left(\sigma^2 + 3G_t^2 + 3\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right)$$

$$\leq \frac{2\eta^2K^3}{3(2-3K^2L^2\eta^2)}\left(\sigma^2 + 3G_t^2 + 3\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right).$$

Substituting the above results into (27), we get

$$\mathbb{E}\left\langle \nabla f\left(\boldsymbol{w}_t\right), \eta\sum_{m\in[M]}\sum_{\tau=0}^{K-1}p_m\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m,\boldsymbol{\xi}_{t,\tau+1}^m\right)\right\rangle$$

$$= \eta\sum_{\tau=0}^{K-1}\mathbb{E}\left\langle \nabla f\left(\boldsymbol{w}_t\right), \sum_{m\in[M]}p_m\nabla f_m\left(\boldsymbol{w}_{t,\tau+1}^m\right)\right\rangle$$

$$\geq \eta\sum_{\tau=0}^{K-1}\frac{1}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 - \frac{1}{2}ML^2\sum_{m\in[M]}p_m^2\left\|\boldsymbol{w}_t - \boldsymbol{w}_{t,\tau}^m\right\|^2$$

$$= \frac{\eta K}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 - \frac{ML^2}{2}\bar{p}A_t, \tag{35}$$

where $\bar{p} \triangleq \sum_{m\in[M]}p_m^2$.

Then we consider the third term of (26) and obtain

$$\frac{L\eta^2}{2}\left\|\sum_{m\in[M]}\sum_{\tau=0}^{K-1}p_m\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m,\boldsymbol{\xi}_{t,\tau+1}^m\right)\right\|^2$$

$$= \frac{L\eta^2}{2}\left\|\sum_{m\in[M]}p_m\sum_{\tau=0}^{K-1}\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m,\boldsymbol{\xi}_{t,\tau+1}^m\right)\right\|^2$$

$$\leq \frac{L\eta^2M}{2}\sum_{m\in[M]}p_m^2\left\|\sum_{\tau=0}^{K-1}\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m,\boldsymbol{\xi}_{t,\tau+1}^m\right)\right\|^2$$

$$\leq \frac{LM}{2}\bar{p}A_t. \tag{36}$$

Substituting (36) and (35) into (26), we get

$$\mathbb{E}f\left(\tilde{\boldsymbol{w}}_{t+1}\right)$$

$$\leq \mathbb{E}f\left(\boldsymbol{w}_t\right) - \left(\frac{\eta K}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 - \frac{ML^2}{2}\bar{p}A_t\right) + \frac{LM}{2}\bar{p}A_t$$

$$= \mathbb{E}f\left(\boldsymbol{w}_t\right) - \frac{\eta K}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 + \frac{\bar{p}ML(L+1)}{2}A_t$$

$$\leq \mathbb{E}f\left(\boldsymbol{w}_t\right) - \frac{\eta K}{2}\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 + \frac{\eta^2\bar{p}ML(L+1)K^3}{3(2-3K^2L^2\eta^2)}\cdot \left(\sigma^2 + 3G_t^2 + 3\mathbb{E}\left[\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2\right]\right)$$

$$= \mathbb{E}f\left(\boldsymbol{w}_t\right) + \left(\frac{\eta^2\bar{p}ML(L+1)K^3}{2-3K^2L^2\eta^2} - \frac{\eta K}{2}\right)\left\|\nabla f\left(\boldsymbol{w}_t\right)\right\|^2 + \frac{\eta^2\bar{p}ML(L+1)K^3}{3(2-3K^2L^2\eta^2)}\left(\sigma^2 + 3G_t^2\right).$$

■

Appendix C

Proof of Lemma 3:

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{w}_{t+1}-\tilde{\boldsymbol{w}}_{t+1}\|^2\\
&=\mathbb{E}\left\|\boldsymbol{w}_t+\sum_{m\in[M]}p_m\hat{\boldsymbol{\delta}}_t^m-\sum_{m\in[M]}p_m\boldsymbol{w}_{t+1}^m\right\|^2\\
&=\mathbb{E}\left\|\sum_{m\in[M]}p_m\left[\hat{\boldsymbol{\delta}}_t^m-\left(\boldsymbol{w}_{t+1}^m-\boldsymbol{w}_t\right)\right]\right\|^2\\
&\le M\sum_{m\in[M]}p_m^2\mathbb{E}\left[\left\|\hat{\boldsymbol{\delta}}_t^m-\boldsymbol{\delta}_t^m\right\|^2\right]\\
&\le M\sum_{m\in[M]}p_m^2\left\|\eta\sum_{\tau=0}^{K-1}\nabla f_m\left(\boldsymbol{w}_{t,\tau}^m,\boldsymbol{\xi}_{t,\tau+1}^m\right)\right\|^2\alpha_t \qquad (37)\\
&\le M\bar{p}A_t\alpha_t \qquad (38)\\
&\le \frac{2\eta^2K^3M\bar{p}\alpha_t}{3(2-3K^2L^2\eta^2)}\left(\sigma^2+3G_t^2+3\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]\right).
\end{aligned}$$

Here, (37) is derived from Lemma 1 and (38) is derived from the definition of A_t.

■

Appendix D

Proof of Theorem 1: We use the L-smooth condition to reveal the effect of quantized transmission in the training process:

$$\begin{aligned}
\mathbb{E}f(\boldsymbol{w}_{t+1})&=\mathbb{E}f(\tilde{\boldsymbol{w}}_{t+1}+\boldsymbol{w}_{t+1}-\tilde{\boldsymbol{w}}_{t+1})\\
&\le\mathbb{E}\left[f(\tilde{\boldsymbol{w}}_{t+1})+\langle\nabla f(\tilde{\boldsymbol{w}}_{t+1}),\boldsymbol{w}_{t+1}-\tilde{\boldsymbol{w}}_{t+1}\rangle\right.\\
&\quad\left.+\frac{L}{2}\|\boldsymbol{w}_{t+1}-\tilde{\boldsymbol{w}}_{t+1}\|^2\right] \qquad (39)\\
&\approx\mathbb{E}f(\tilde{\boldsymbol{w}}_{t+1})+\frac{L}{2}\mathbb{E}\|\boldsymbol{w}_{t+1}-\tilde{\boldsymbol{w}}_{t+1}\|^2, \qquad (40)
\end{aligned}$$

where (39) comes from the quadratic upper bound of L-smooth functions $f(\boldsymbol{y})-f(\boldsymbol{x})-\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x})\le\frac{L}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$; (40) comes from the nearly unbiasedness of quantized transmission in Lemma 1. Using Lemma 2 and Lemma 3, we get

$$\begin{aligned}
&\mathbb{E}f(\boldsymbol{w}_{t+1})\\
&\le\mathbb{E}[f(\boldsymbol{w}_t)]+\left(\frac{\eta^2\bar{p}ML(L+1)K^3}{2-3K^2L^2\eta^2}-\frac{\eta K}{2}\right)\cdot\\
&\quad\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]+\frac{\eta^2\bar{p}ML(L+1)K^3}{3(2-3K^2L^2\eta^2)}\left(\sigma^2+3G_t^2\right)\\
&\quad+\frac{2\eta^2K^3M\bar{p}\alpha_t}{3(2-3K^2L^2\eta^2)}\left(\sigma^2+3G_t^2+3\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]\right)\\
&=\mathbb{E}[f(\boldsymbol{w}_t)]+\left(\left(\frac{MK^3\bar{p}\eta^2}{2-3K^2L^2\eta^2}\right)\left(L^2+L+2\alpha_t\right)-\right.\\
&\quad\left.\frac{\eta K}{2}\right)\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]+\frac{MK^3\bar{p}\eta^2}{3(2-3K^2L^2\eta^2)}\cdot\\
&\quad\left(L^2+L+2\alpha_t\right)\left(\sigma^2+3G_t^2\right)\\
&=\mathbb{E}[f(\boldsymbol{w}_t)]+\left(\kappa\left(L^2+L+2\alpha_t\right)-\frac{\eta K}{2}\right)\cdot\\
&\quad\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]+\kappa\left(L^2+L+2\alpha_t\right)\left(\sigma^2+3G_t^2\right), \qquad (41)
\end{aligned}$$

where (41) comes from the definition $\kappa\triangleq\frac{MK^3\bar{p}\eta^2}{3(2-3K^2L^2\eta^2)}$.

Accumulate over $t=0,\cdots,T-1$ and rearrange terms, yielding

$$\begin{aligned}
&\sum_{t=0}^{T-1}\left(\frac{\eta K}{2}-\kappa\left(L^2+L+2\alpha_t\right)\right)\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]\\
&\le f(\boldsymbol{w}_0)-\mathbb{E}f(\boldsymbol{w}_T)+\sum_{t=0}^{T-1}\kappa\left(L^2+L+2\alpha_t\right)\left(\sigma^2+3G_t^2\right).
\end{aligned}$$

We next study the lower bound of the coefficient of $\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]$. Take η such that $0<\eta<\frac{3}{4MK^2\bar{p}(L^2+L+4d+E)}$, then we have

$$\begin{aligned}
&\frac{\eta K}{2}-\kappa\left(L^2+L+2\alpha_t\right)\\
&=\frac{\eta K}{2}-\frac{MK^3\bar{p}\eta^2}{3(2-3K^2L^2\eta^2)}\left(L^2+L+2\alpha_t\right) \qquad (42)\\
&\ge\frac{\eta K}{2}-\frac{MK^3\bar{p}\eta^2}{3(2-3K^2L^2\eta^2)}\left(L^2+L+4d+E\right)\\
&>\frac{\eta K}{2}-\frac{MK^3\bar{p}\eta^2}{3}\left(L^2+L+4d+E\right) \qquad (43)\\
&>\eta K\left(\frac{1}{2}-\frac{1}{4}\right) \qquad (44)\\
&=\frac{\eta K}{4}, \qquad (45)
\end{aligned}$$

where (42) comes from the upper bound of transmission error α_t, i.e.,

$$\alpha_t=dD_t+E\le4d+E. \qquad (46)$$

(43) and (44) come from the constraint on the value range of η. Therefore, we have

$$\begin{aligned}
&\frac{\eta K}{4}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]\\
&<\sum_{t=0}^{T-1}\left(\frac{\eta K}{2}-\kappa\left(L^2+L+2\alpha_t\right)\right)\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]\\
&\le f(\boldsymbol{w}_0)-\mathbb{E}f(\boldsymbol{w}_T)+\sum_{t=0}^{T-1}\kappa\left(L^2+L+2\alpha_t\right)\left(\sigma^2+3G_t^2\right).
\end{aligned}$$

Rearranging, we get

$$\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla f(\boldsymbol{w}_t)\|^2\right]\le\frac{4}{\eta TK}\left(f(\boldsymbol{w}_0)-\mathbb{E}f(\boldsymbol{w}_T)+\right.\\
&\left.\sum_{t=0}^{T-1}\kappa\left(L^2+L+2\alpha_t\right)\left(\sigma^2+3G_t^2\right)\right).
\end{aligned}$$

■