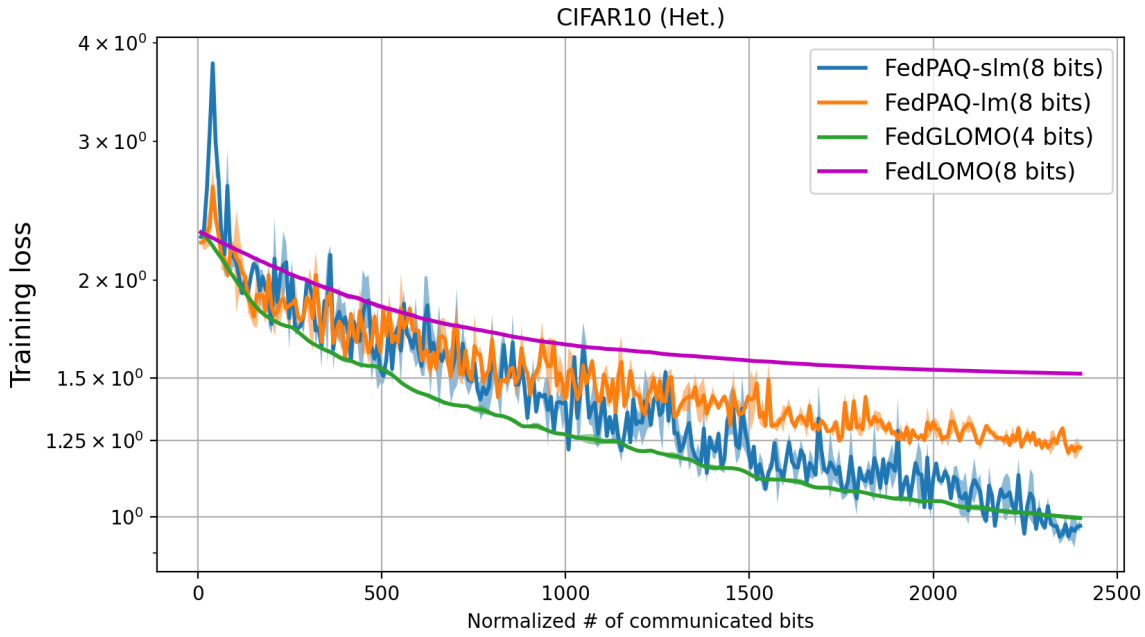

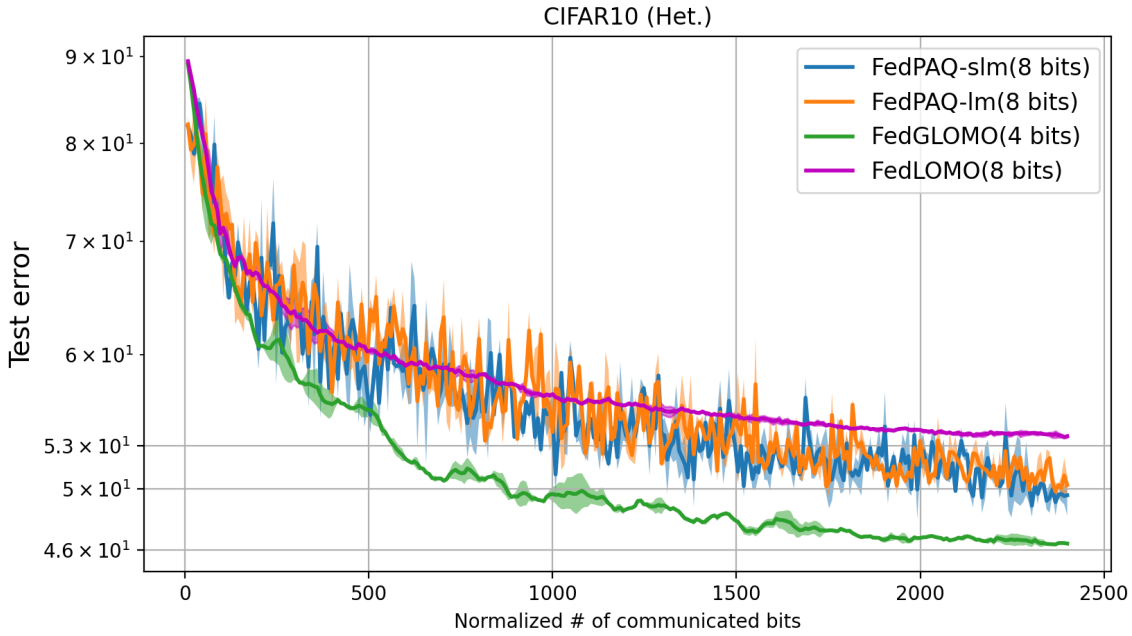

FedPAQ-slm stands for FedPAQ (which is FedAvg with compressed communication) with server as well as client-level momentum, where the momentum is the default momentum available in PyTorch. Similarly, FedPAQ-lm stands for FedPAQ with only local client-level (PyTorch) momentum; this is FedPAQ-m in our paper. The server-level and client-level momentum parameters are 0.5 and 0.9, respectively. From the plots, FedPAQ-slm is very noisy as expected due to the high heterogeneity. FedGLOMO has the best generalization out of all methods -- about a 3% difference.