

Extra Experiments

In this experiments, we compare our new algorithm **DASHA-PP** with previous baselines **MARINA** and **FRECON** in the partial participation setting (the comparisons of **DASHA-PP** with **DASHA** we present in Section A of the paper). We consider **MARINA** and **FRECON** because they are the previous SOTA methods in the *partial participation setting with compression*. We investigate the same optimization problem and setup as in Section A of the paper. All methods use the RandK compressor in these experiments.

Finite-Sum Setting

In Figures 1 and 2, we compare all three methods in the finite-sum setting on two different datasets: *real-sim* and *MNIST*. The parameter s is the number of clients participating in each round. We can see that **DASHA-PP** converges faster than **MARINA**. Since **FRECON** does not support variance reduction of stochastic gradients, it converges to less accurate solutions.

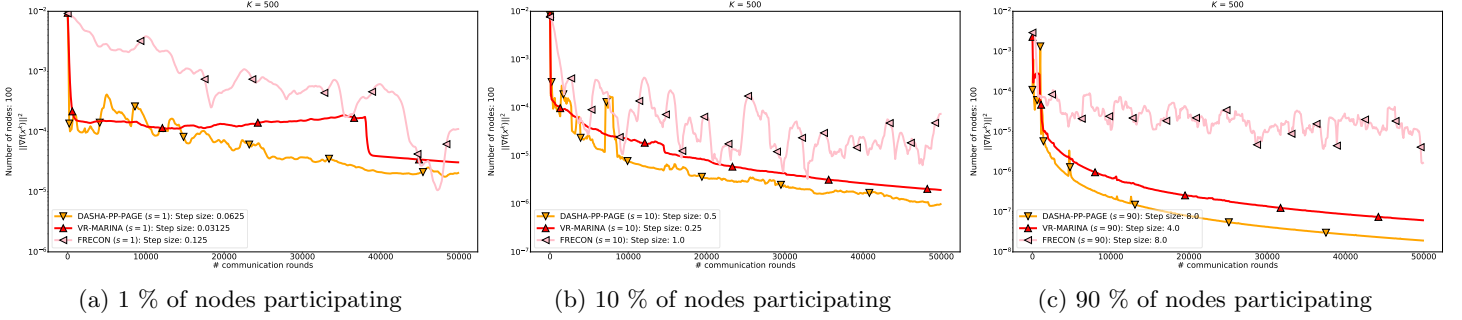


Figure 1: Classification task on *real-sim*

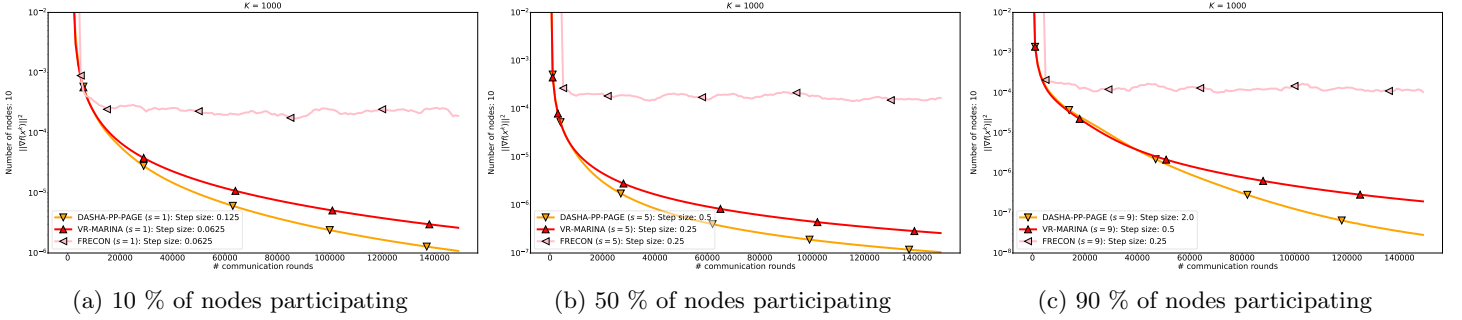


Figure 2: Classification task on *MNIST*

Stochastic Setting

In Figure 3, we consider the stochastic setting. We can see that **DASHA-PP** has better convergence rates than **FRECON**. **DASHA-PP** with $s \in \{5, 10\}$ even improves **MARINA** that works in the *full participation regime* (one should expect that convergence rate of **MARINA** with partial participation will be even worse).

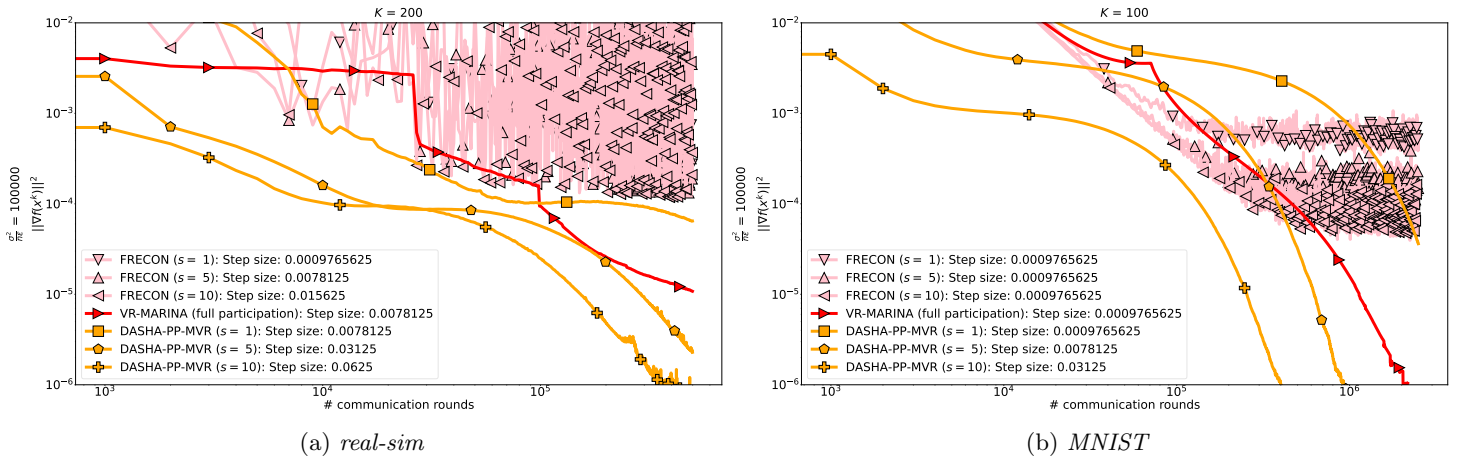


Figure 3: Classification task