

Figure R1: **Communication efficiency.** In this experiment ResNet model was trained on CIFAR-100. *Left:* model accuracy on the test set in the end of each communication round. *Right:* accuracy against communication costs (measured as the number of sent parameters). The full model trained with FedAvg algorithm is significantly less efficient than with spectral sharding-based training. In contrast, slim model trained with FedAvg is more efficient but the size of the final model is restricted by the capabilities of the weakest client.

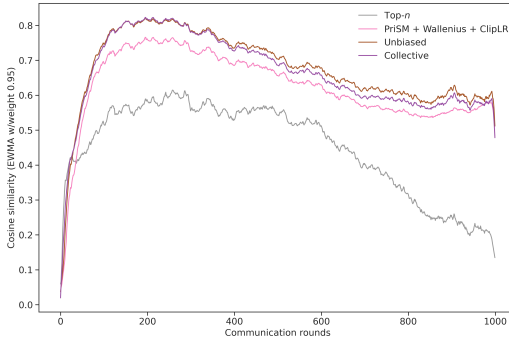


Figure R2: **Comparison of FedAvg weight updates against ours.** For ResNet model trained on CIFAR-10, updates provided by the Top- $n$  strategy significantly deviate from those of FedAvg method. This correlates with the worse performance in this experiment.

Table R1: **FedHM-like aggregation.** In this experiment during training we averaged on the server side the re-materialized clients' weight matrices instead of updated singular vectors. In this case, Collective strategy demonstrates the best performance. Unbiased strategy lags behind, probably because of large auxiliary multipliers affecting the re-materialized matrices.

STRATEGY	RESNET	CCT
TOP- $n$	$50.58 \pm 1.80$	$46.35 \pm 0.69$
UNBIASED	$46.04 \pm 1.24$	$46.24 \pm 0.25$
COLLECTIVE	$52.30 \pm 1.49$	$47.98 \pm 1.26$

Table R2: **Faster approximated SVD.** In this experiment we trained models on CIFAR-100 data with keep ratio of 0.05 and used fast randomized approximation of SVD algorithm which estimates only half of the singular values.

STRATEGY	RESNET	CCT
UNBIASED	$36.89 \pm 1.06$	$36.81 \pm 0.87$
COLLECTIVE	$41.99 \pm 0.73$	$42.09 \pm 0.70$