

Fig. 1. Selection of hyperparameters: mini-batch  $b$  (left) and local iterations  $K$  (right) on the Shakespeare dataset for our FedCBG algorithm.

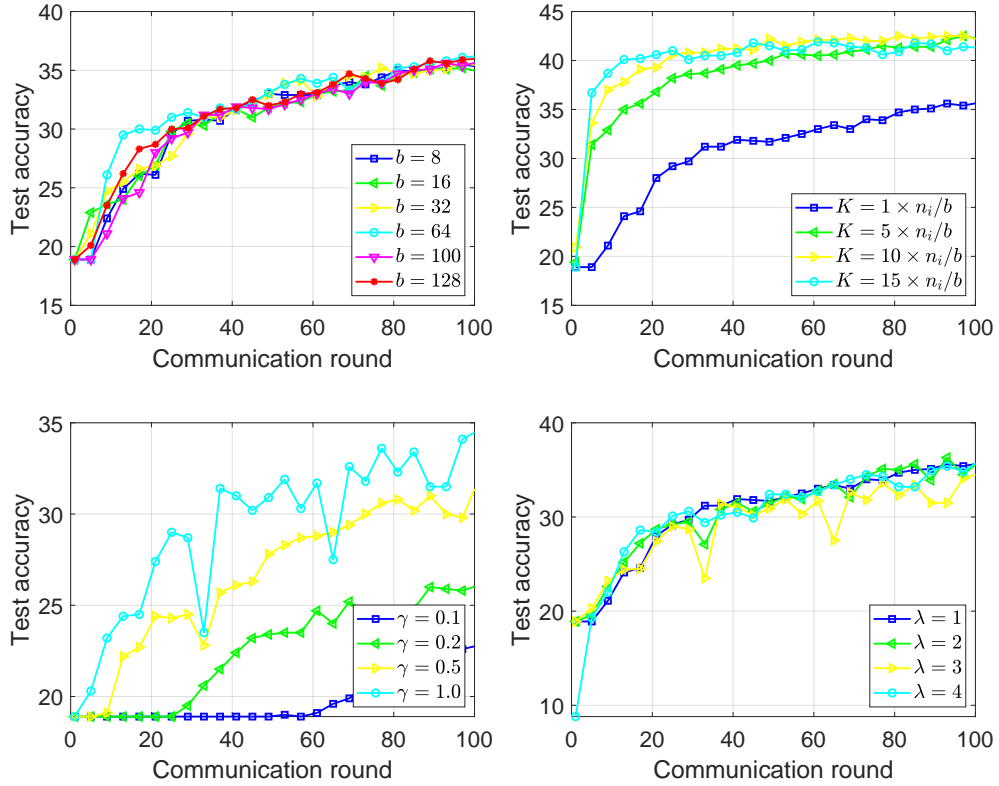


Fig. 2. Selection of hyperparameters: mini-batch  $b$ , local iterations  $K$ , global learning rate  $\gamma$  and clipping parameter  $\lambda$  on the Shakespeare dataset for our FAT-clipping-PI algorithm.

TABLE I  
THE HYPERPARAMETER SETTINGS FOR EACH ALGORITHM ARE BASED ON THE FINE-TUNING EXPERIMENTS IN FIGS. 1, 2 AND 3.

Algorithms	hyper-parameters			
	$b$	$K$	$\gamma$	$\lambda$
FAT-Clipping-PR	100	$10 \times n_i/b$	1.0	3.0
FAT-Clipping-PI	100	$10 \times n_i/b$	1.0	4.0
FedCBG (Ours)	100	$10 \times n_i/b$	0.3	3.0

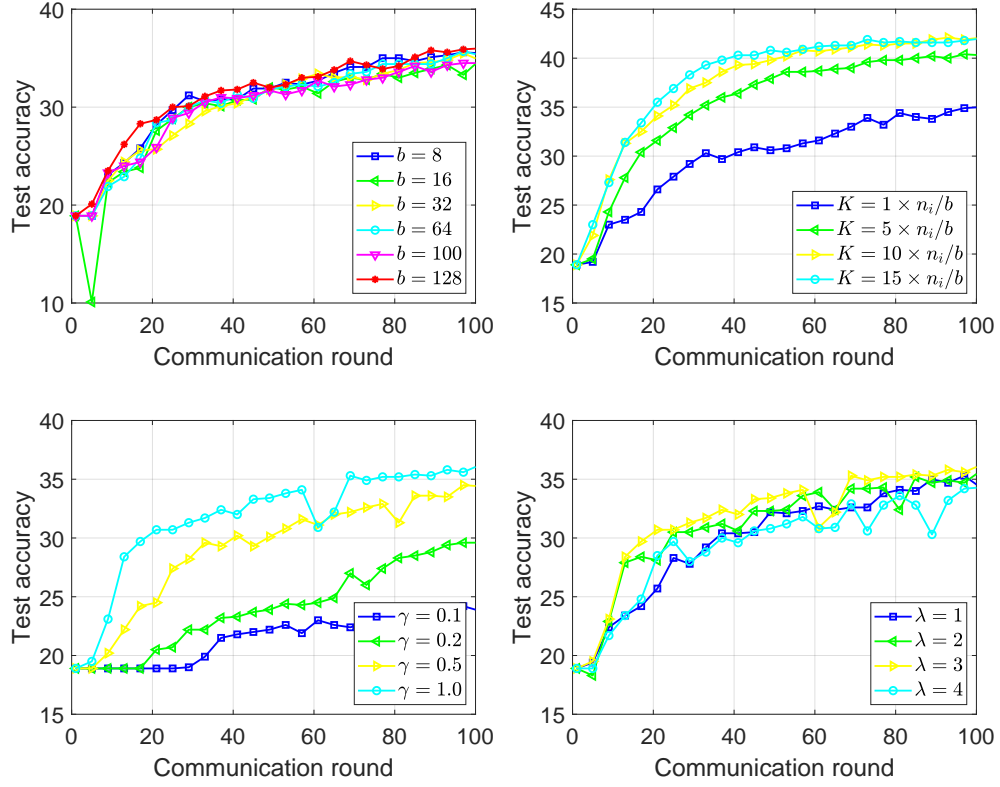


Fig. 3. Selection of hyperparameters: mini-batch  $b$ , local iterations  $K$ , global learning rate  $\gamma$  and clipping parameter  $\lambda$  on the Shakespeare dataset for our FAT-clipping-PR algorithm.

TABLE II  
COMPARISON OF THE TRAINING LOSS (TLOSS.), TESTING CLASSIFICATION ACCURACY (TACC.) AND COMMUNICATION ROUND TO REACH TARGET TEST ACCURACY IN FL WITH HEAVY-TAILED NOISE UNDER HYPER-PARAMETER IN TABLE I. “ROUND” REFERS TO THE NUMBER OF COMMUNICATION ROUNDS TO ACHIEVE TEST ACCURACY 84.5% FOR CIFAR-10, 46.0% FOR CIFAR-100 AND 42.0% FOR SHAKESPEARE DATASETS.

Datasets	Evaluation	CIFAR-10	CIFAR-100	Shakespeare
FAT-Clipping-PR	TLoss	0.15	3.05	2.08
	TAcc. (%)	83.6	43.8	42.2
	Round	245 (2.9 $\times$ )	296 (1.8 $\times$ )	64 (1.4 $\times$ )
FAT-Clipping-PI	TLoss	0.08	2.54	2.06
	TAcc. (%)	85.1	44.2	42.1
	Round	171 (2.0 $\times$ )	235 (1.4 $\times$ )	58 (1.3 $\times$ )
FedCBG (Ours)	TLoss	<b>0.06</b>	<b>1.70</b>	<b>2.03</b>
	TAcc. (%)	<b>85.9</b>	<b>45.8</b>	<b>42.7</b>
	Round	<b>85</b>	<b>166</b>	<b>46</b>