A APPENDIX

The appendix contains supplementary experimental results. In Table 5 we show the ablation study of Non-IID experiments. Compared to results of IID experiments shown in Table 4, Table .5 shows the ablation study Maksed CrossEntropy which is shown beneficial for balanced Non-IID data partition. In Table 6, we show the hyperparameters adopted in our experiments. In Figure 3 and 4, we show the *Fix* complexity assignments of MNIST and WikiText with iid data partition experiments. From Figure 6 to 8, we show the *Fix* complexity assignments for all balanced Non-IID data partition experiments. Figure 9 to 11, we show the learning curve of experiments with *Dyanmic* complexity assignments. The complete results for all experiments with *Dyanmic* complexity assignments can be found in Table 7, 8, and 9.

Model	Normalization	Scaler	Masked CrossEntropy	Accuracy Non-IID			
				MNIST		CIFAR10	
				Local	Global	Local	Global
a	None	N/A	✗	97.4	97.4	42.6	42.8
	sBN			99.4	**99.4**	53.4	53.7
	None	N/A	✓	99.7	95.6	91.7	58.5
	IN			99.8	98.7	88.4	43.7
	GN			99.7	98.3	91.2	58.2
	LN			99.8	98.3	89.9	54.2
	sBN			**99.9**	98.6	**92.1**	**59.2**
e	None	N/A	✗	96.2	96.0	38.9	38.2
	sBN			90.1	90.1	40.7	40.4
	None	N/A	✓	**99.5**	**96.5**	86.6	48.9
	IN			98.5	89.8	83.7	37.0
	GN			99.2	92.2	83.5	36.7
	LN			99.3	94.0	82.6	40.0
	sBN			99.3	94.2	**90.1**	**52.9**
a-e	None	✗	✗	96.8	96.9	37.8	37.4
	sBN			99.2	99.2	41.0	41.3
	None	✗	✓	99.4	95.7	89.1	52.8
	sBN			99.8	98.0	90.7	57.7
	None	✓	✗	97.3	97.3	34.6	34.4
	sBN			99.3	**99.3**	46.0	46.7
	None	✓	✓	99.5	95.8	90.3	55.6
	IN			99.8	98.7	87.0	34.4
	GN			99.5	96.2	88.7	49.7
	LN			99.5	96.2	78.5	25.0
	sBN			**99.8**	98.2	**92.8**	**60.4**

Table 5: Ablation Study of Non-IID scenarios. The Masked CrossEntropy is used for Non-IID experiments. It significantly improves the local performance and moderately improves global performance. Single letter model 'a' and 'e' are FedAvg equipped with various normalization methods. The sBN significantly outperforms other existing normalization methods, including InstanceNorm (IN), GroupNorm (GN) (the number of group G=4), and LayerNorm (LN). Scaler is used for HeteroFL to train models of different sizes and moderately improve the results.

Data		MNIST	CIFAR10	WikiText2
Model		CNN	PreResNet18	Transformer
Hidden size		[64, 128, 256, 512]	[64, 128, 256, 512]	[512, 512, 512, 512]
Local Epoch E		5	5	1
Local Batch Size B		10	10	100
Optimizer			SGD	
Momentum			0.9	
Weight decay			5.00E-04	
Learning rate η		0.01	0.1	0.1
Communication rounds	IID	200	400	100
	non-IID	400	800	200
Decay schedule (0.1)	IID	[100]	[150, 250]	[25, 50]
	non-IID	[200]	[300, 500]	[50, 100]
Embedding Size				256
Number of heads			N/A	8
Dropout				0.2
Sequence length				64

Table 6: Hyperparameters and model architecture used in our experiments.

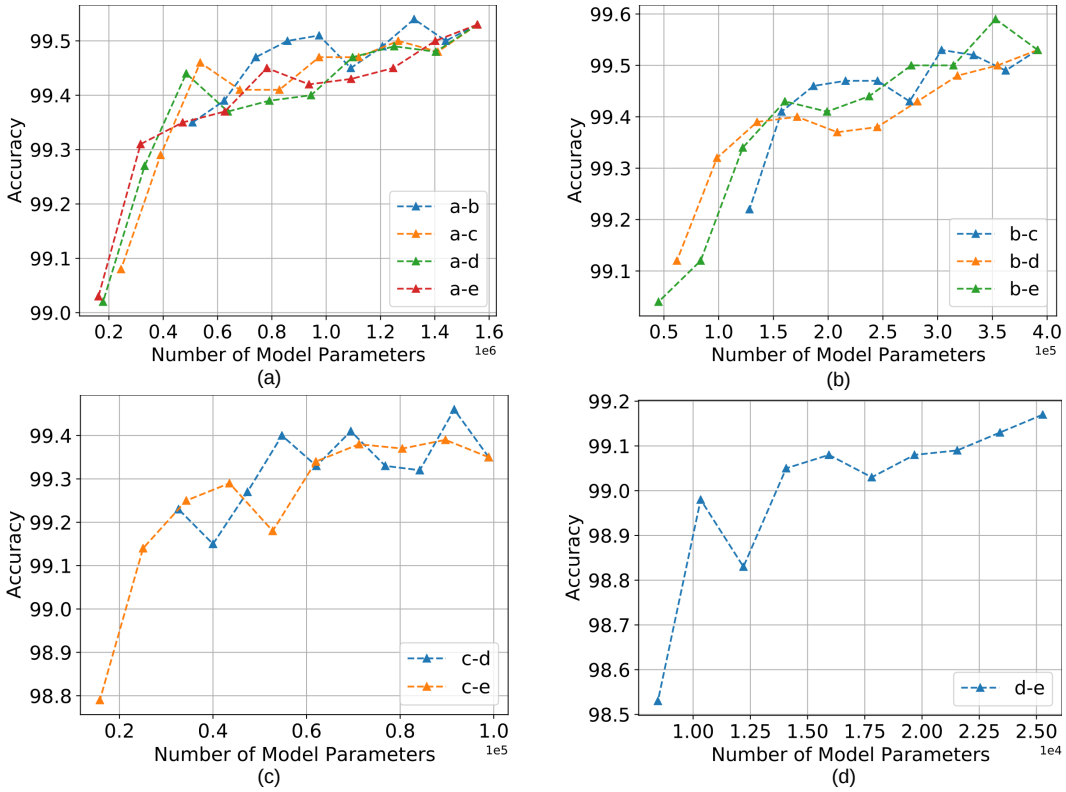

Figure 3: Interpolation experimental results for MNIST (IID) dataset between global model complexity ((a) a, (b) b, (c) c, (d) d) and various smaller model complexities.

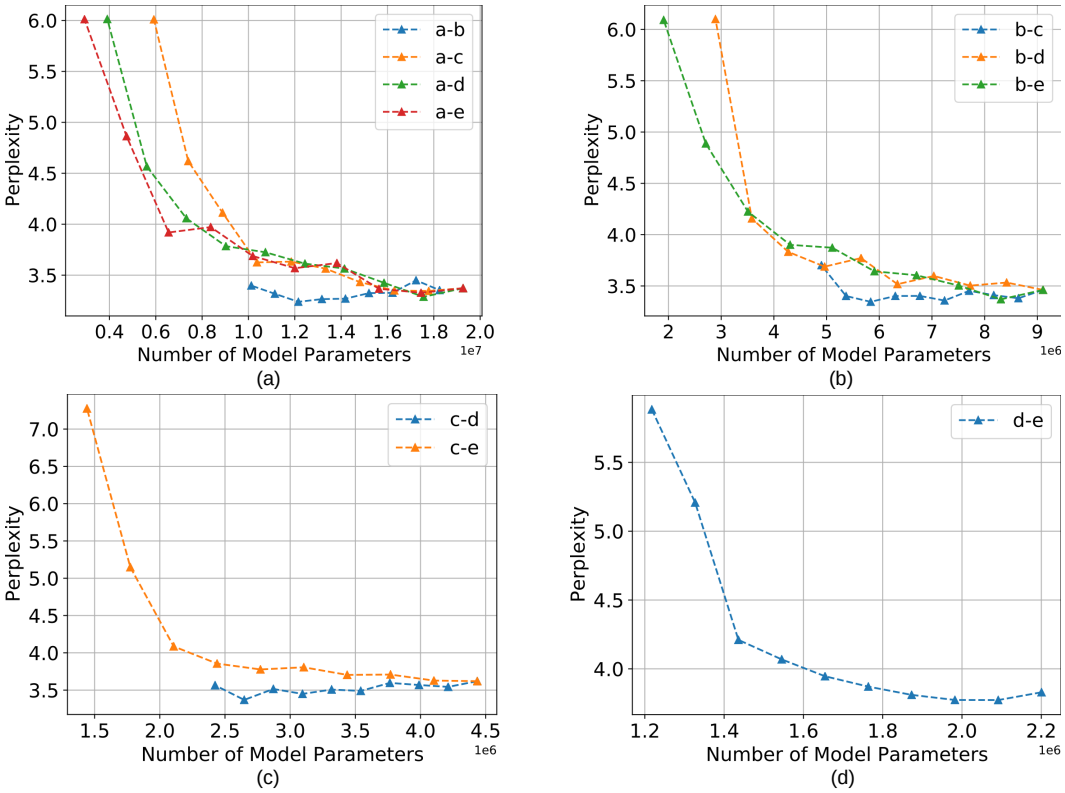

Figure 4: Interpolation experimental results for WikiText2 (IID) dataset between global model complexity ((a) a, (b) b, (c) c, (d) d) and various smaller model complexities.

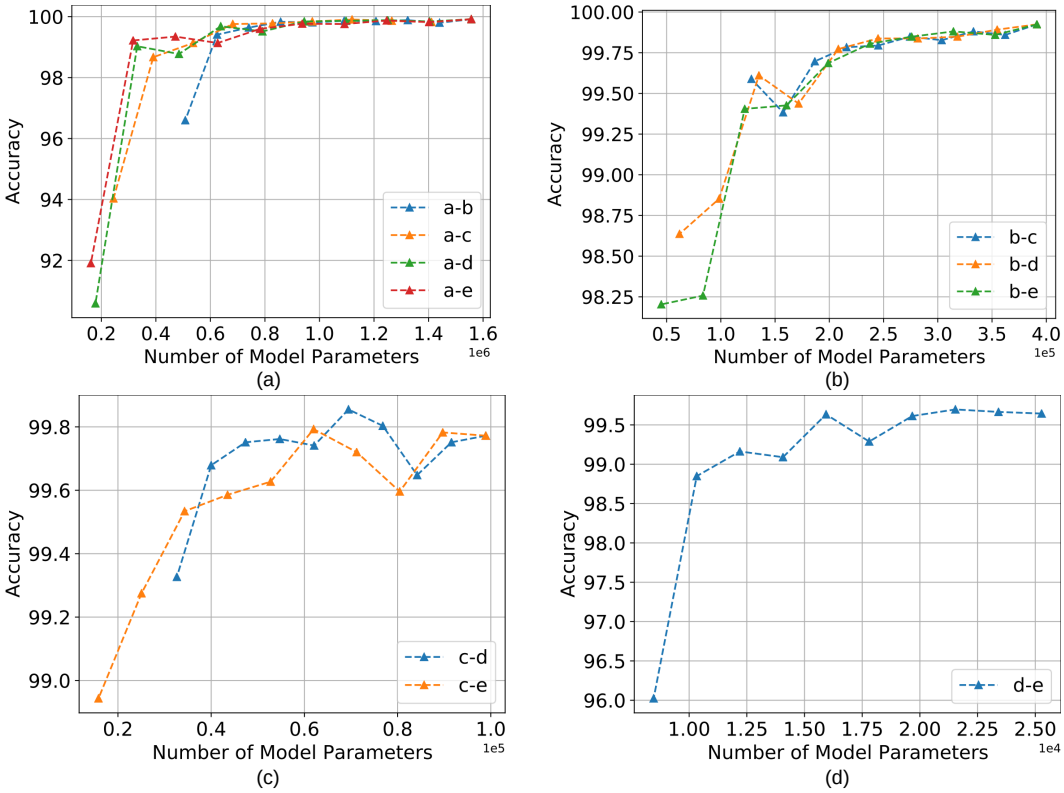

Figure 5: Interpolation experimental results for MNIST (non-IID, Local) dataset between global model complexity ((a) a, (b) b, (c) c, (d) d) and various smaller model complexities.

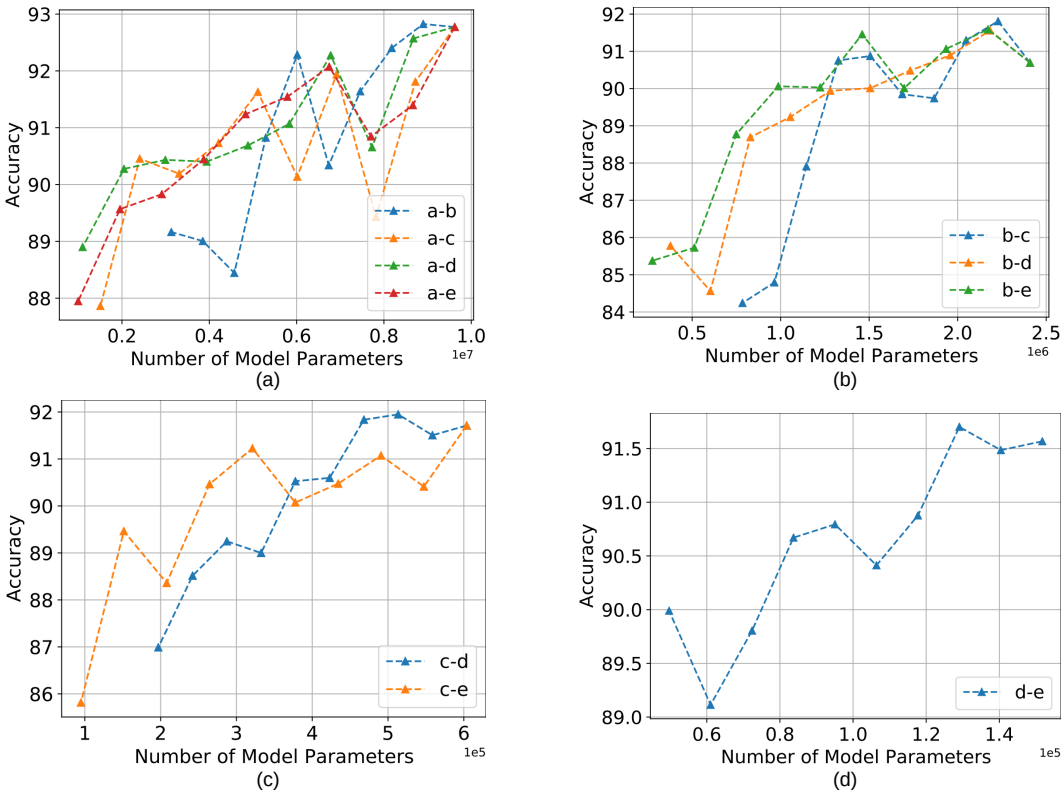

Figure 6: Interpolation experimental results for CIFAR10 (non-IID, Local) dataset between global model complexity ((a) a, (b) b, (c) c, (d) d) and various smaller model complexities.

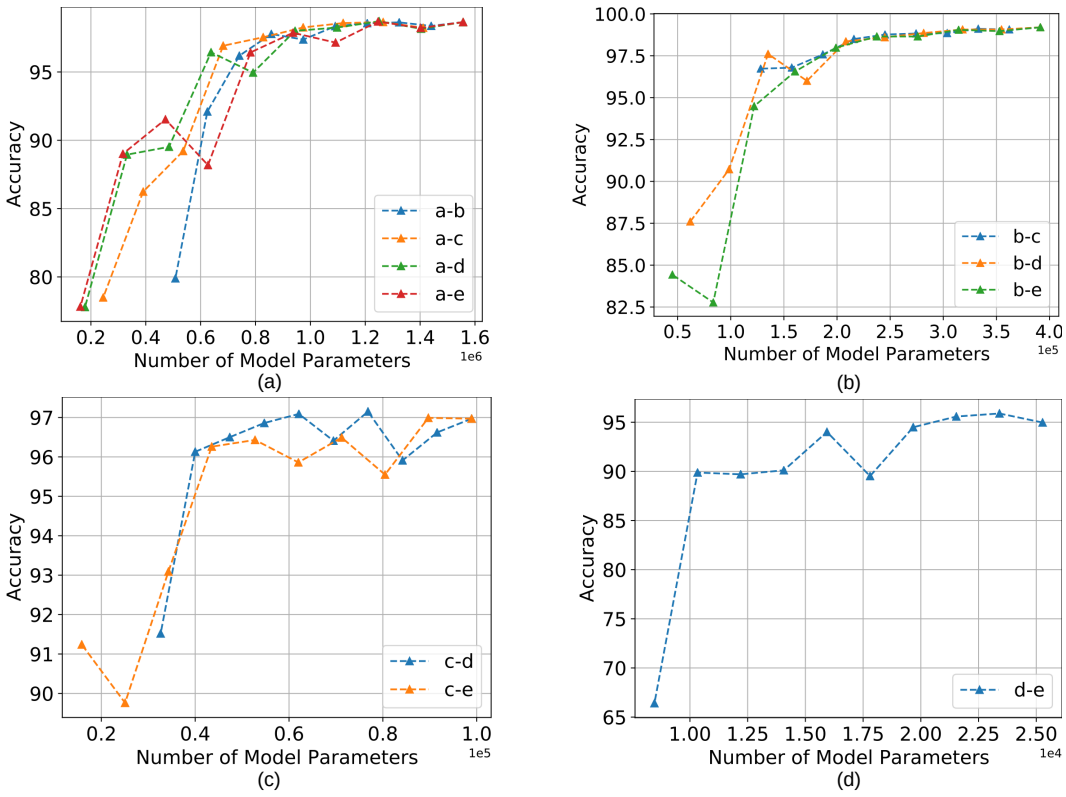

Figure 7: Interpolation experimental results for MNIST (non-IID, Global) dataset between global model complexity ((a) a, (b) b, (c) c, (d) d) and various smaller model complexities.

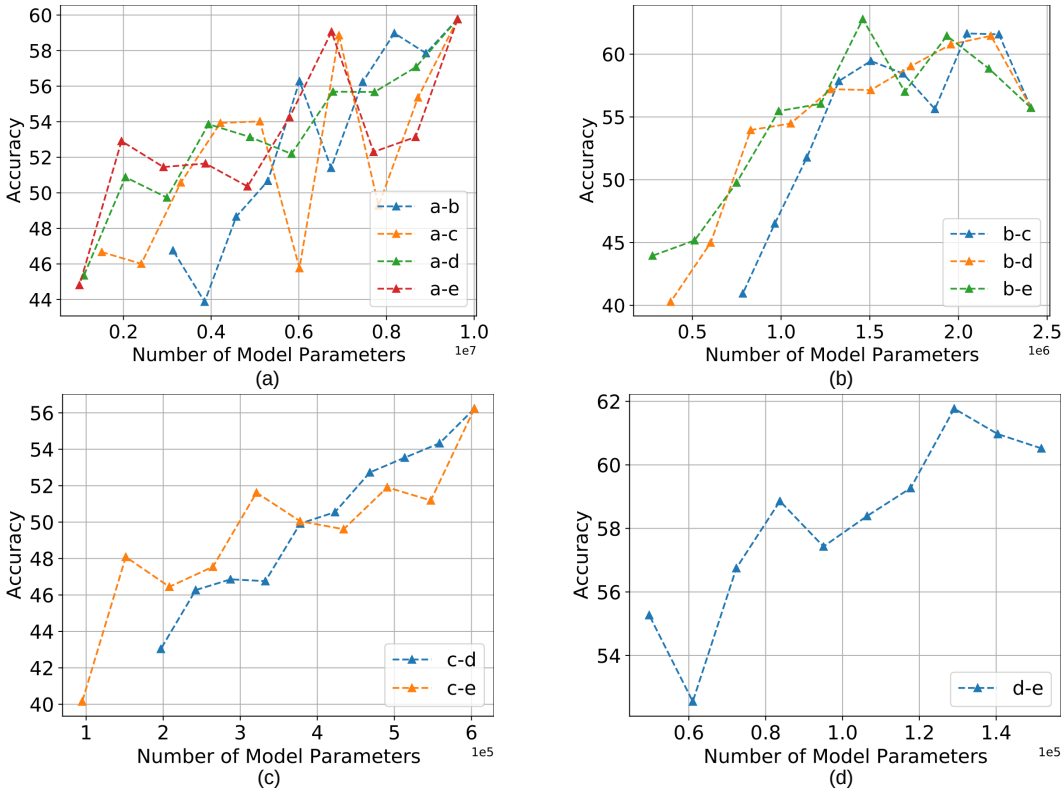

Figure 8: Interpolation experimental results for CIFAR10 (non-IID, Global) dataset between global model complexity ((a) a, (b) b, (c) c, (d) d) and various smaller model complexities.

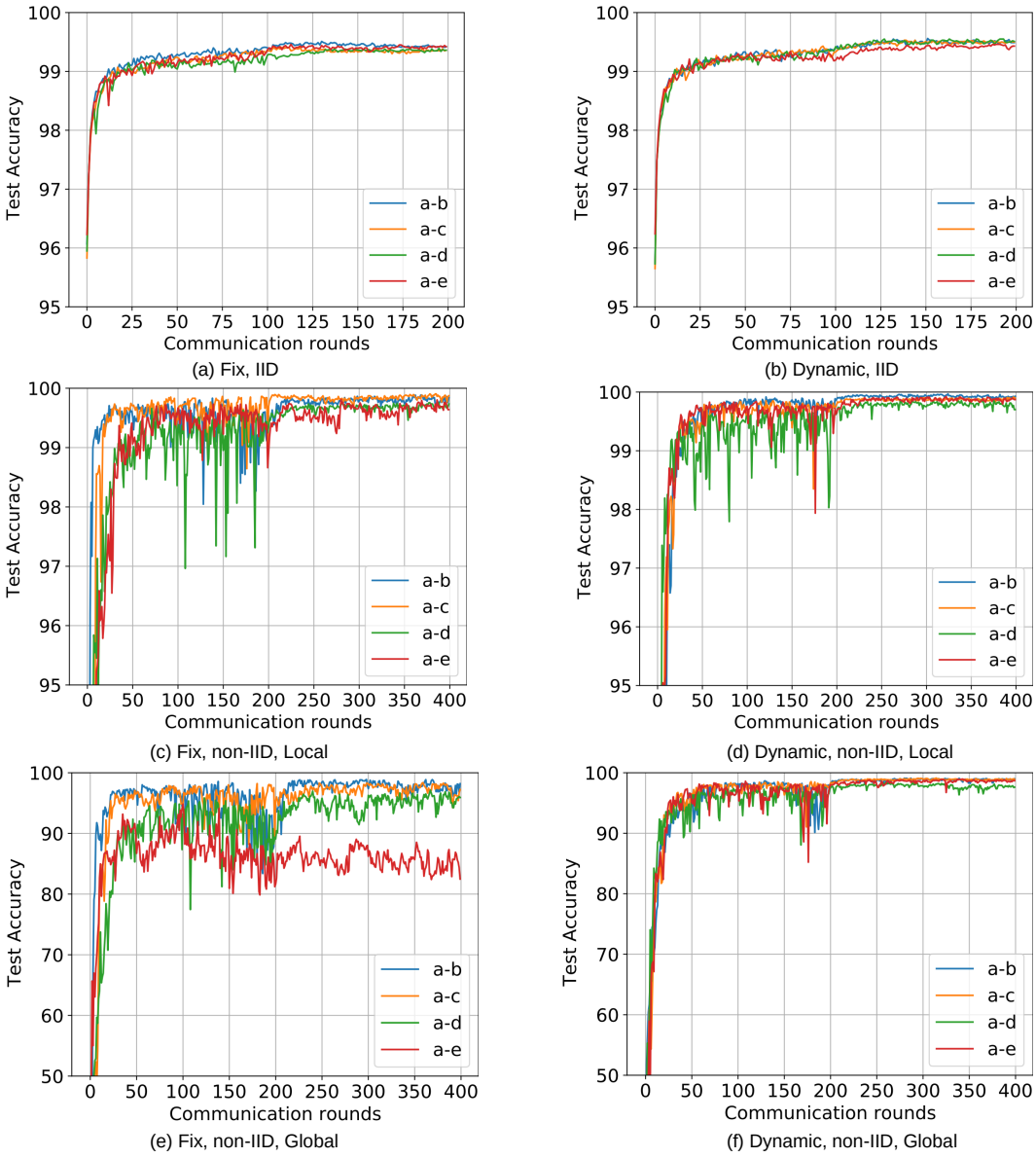

Figure 9: Learning curves of MNIST datasets with 50% Fix and Dynamic computation complexity assignments.

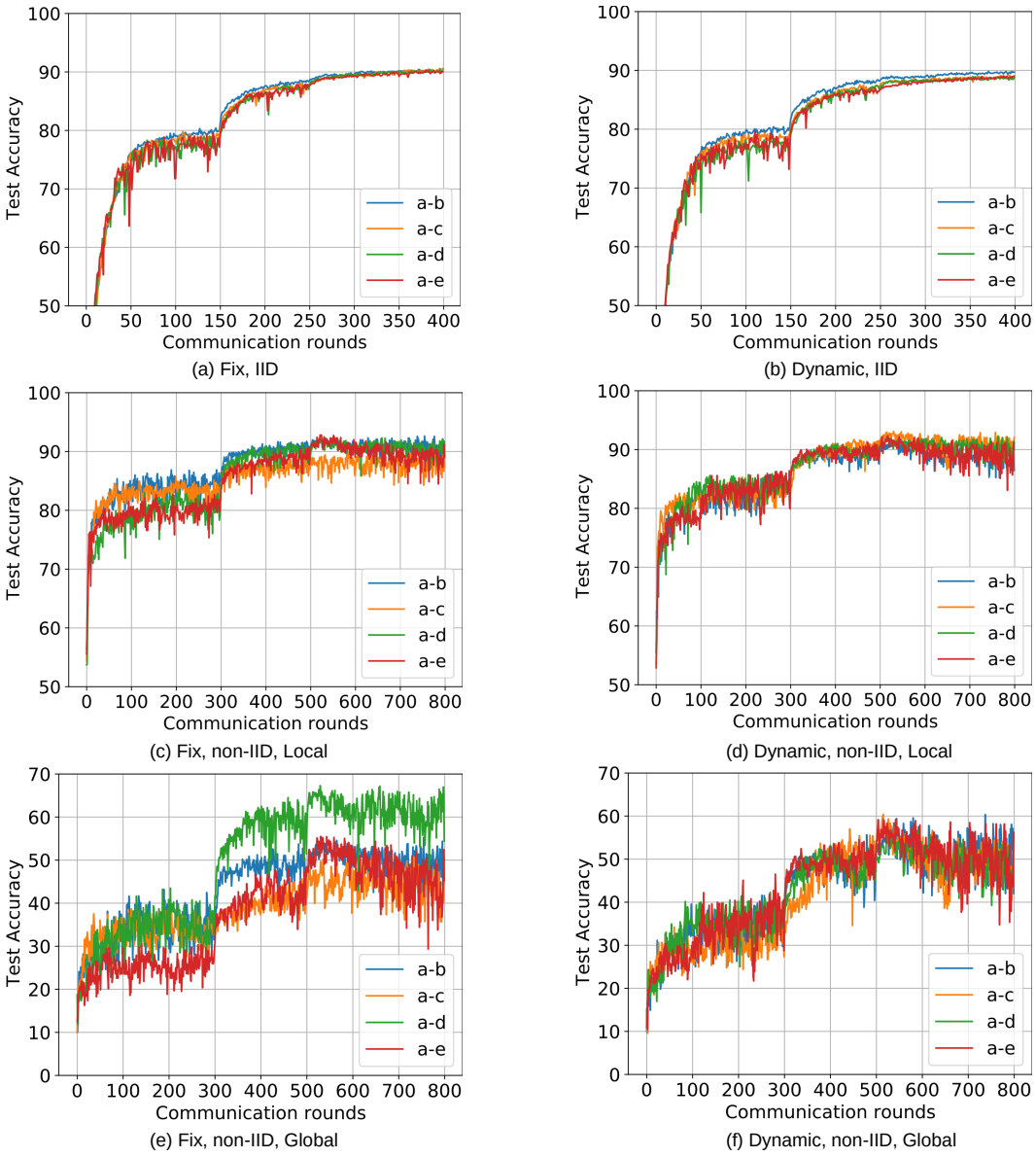

Figure 10: Learning curves of CIFAR10 datasets with 50% Fix and Dynamic computation complexity assignments.

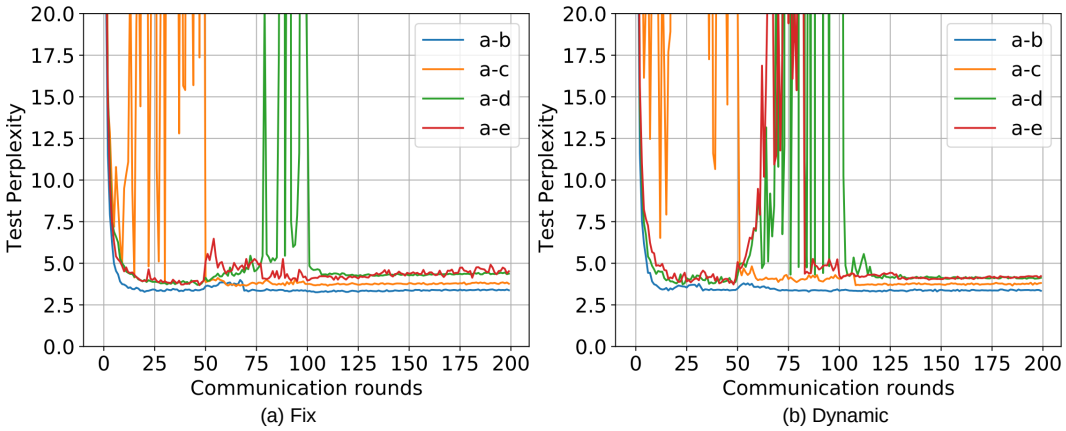

Figure 11: Learning curves of WikiText2 datasets with 50% Fix and Dynamic computation complexity assignments.

Model	Ratio	Parameters	FLOPs	Space (MB)	Accuracy		
					IID	Non-IID	
						Local	Global
a	1.00	1.6 M	80.5 M	5.94	99.53	99.85	98.92
a-b	0.63	974.1 K	50.5 M	3.72	99.54	99.96	99.10
a-c	0.53	827.9 K	42.9 M	3.16	99.52	99.89	99.12
a-d	0.51	791.1 K	41.0 M	3.02	99.54	99.81	98.37
a-e	0.50	781.7 K	40.5 M	2.98	99.46	99.89	98.96
a-b-c	0.44	682.4 K	35.4 M	2.60	99.53	99.90	98.72
a-b-d	0.42	657.8 K	34.1 M	2.51	99.52	99.78	98.02
a-b-e	0.42	651.6 K	33.8 M	2.49	99.54	99.95	98.92
a-c-d	0.36	560.4 K	29.1 M	2.14	99.57	99.95	99.34
a-c-e	0.36	554.1 K	28.7 M	2.11	99.52	99.94	98.43
a-d-e	0.34	529.6 K	27.4 M	2.02	99.57	99.80	98.92
a-b-c-d	0.33	518.1 K	26.9 M	1.98	99.54	99.80	99.03
a-b-c-e	0.33	513.4 K	26.7 M	1.96	99.46	99.69	97.53
a-b-d-e	0.32	495.0 K	25.7 M	1.89	99.49	99.85	98.66
a-c-d-e	0.27	421.9 K	21.9 M	1.61	99.54	99.84	98.80
a-b-c-d-e	0.27	415.8 K	21.6 M	1.59	99.46	99.85	98.29
b	1.00	391.4 K	20.5 M	1.49	99.53	99.87	99.10
b-c	0.63	245.1 K	12.9 M	0.94	99.49	99.87	99.05
b-d	0.53	208.3 K	11.0 M	0.79	99.44	99.85	98.95
b-e	0.51	199.0 K	10.4 M	0.76	99.51	99.67	98.51
b-c-d	0.44	171.9 K	9.1 M	0.66	99.54	99.84	98.98
b-c-e	0.42	165.6 K	8.7 M	0.63	99.51	99.85	98.20
b-d-e	0.36	141.1 K	7.4 M	0.54	99.48	99.89	98.72
b-c-d-e	0.33	130.5 K	6.9 M	0.50	99.52	99.88	98.99
c	1.00	98.9 K	5.3 M	0.38	99.35	99.56	96.34
c-d	0.63	62.1 K	3.4 M	0.24	99.38	99.92	99.05
c-e	0.53	52.8 K	2.9 M	0.20	99.39	99.79	97.27
c-d-e	0.44	43.6 K	2.4 M	0.17	99.31	99.76	97.85
d	1.00	25.3 K	1.4 M	0.10	99.17	99.86	97.86
d-e	0.63	15.9 K	909.5 K	0.06	99.19	99.63	97.70
e	1.00	6.6 K	400.5 K	0.03	98.66	99.07	92.84
Standalone (Liang et al., 2020)	1.00	633.2 K	1.3 M	2.42	86.24	98.72	30.41
FedAvg (Liang et al., 2020)	1.00	633.2 K	1.3 M	2.42	97.93	98.20	98.20
LG-FedAvg (Liang et al., 2020)	1.00	633.2 K	1.3 M	2.42	97.93	98.54	98.17

Table 7: Results of combination of various computation complexity levels for MNIST dataset.

Model	Ratio	Parameters	FLOPs	Space (MB)	Accuracy		
					IID	Non-IID	
						Local	Global
a	1.00	9.6 M	330.2 M	36.71	91.19	92.38	56.88
a-b	0.63	6.0 M	206.8 M	22.95	90.60	91.35	59.93
a-c	0.53	5.1 M	175.7 M	19.50	90.59	92.83	60.25
a-d	0.51	4.9 M	167.9 M	18.64	90.28	91.78	56.54
a-e	0.50	4.8 M	165.9 M	18.43	90.29	92.10	59.11
a-b-c	0.44	4.2 M	144.9 M	16.07	89.70	90.41	54.16
a-b-d	0.42	4.1 M	139.7 M	15.49	89.98	90.29	51.79
a-b-e	0.42	4.0 M	138.4 M	15.35	89.79	90.79	62.17
a-c-d	0.36	3.5 M	119.0 M	13.20	89.47	89.82	53.13
a-c-e	0.36	3.4 M	117.6 M	13.05	89.35	93.59	57.30
a-d-e	0.34	3.3 M	112.4 M	12.48	88.75	91.11	56.74
a-b-c-d	0.33	3.2 M	110.1 M	12.19	89.33	91.32	54.50
a-b-c-e	0.33	3.2 M	109.1 M	12.09	89.37	92.52	61.56
a-b-d-e	0.32	3.1 M	105.1 M	11.65	89.40	91.80	56.78
a-c-d-e	0.27	2.6 M	89.6 M	9.93	88.42	91.50	62.15
a-b-c-d-e	0.27	2.6 M	88.4 M	9.78	88.83	92.49	61.64
b	1.00	2.4 M	83.3 M	9.19	89.82	93.83	55.45
b-c	0.63	1.5 M	52.3 M	5.75	89.00	89.96	52.29
b-d	0.53	1.3 M	44.4 M	4.88	89.18	91.78	51.07
b-e	0.51	1.2 M	42.4 M	4.67	89.10	90.68	59.81
b-c-d	0.44	1.1 M	36.7 M	4.02	88.35	92.79	58.09
b-c-e	0.42	1.0 M	35.3 M	3.88	87.98	91.98	58.28
b-d-e	0.36	866.3 K	30.1 M	3.30	88.06	91.94	54.02
b-c-d-e	0.33	800.7 K	27.9 M	3.05	87.92	91.90	59.10
c	1.00	603.8 K	21.2 M	2.30	87.55	91.09	55.12
c-d	0.63	377.8 K	13.4 M	1.44	86.75	91.58	54.61
c-e	0.53	321.1 K	11.3 M	1.22	86.88	91.83	63.47
c-d-e	0.44	264.6 K	9.4 M	1.01	85.79	91.49	55.42
d	1.00	151.8 K	5.5 M	0.58	84.21	90.77	61.13
d-e	0.63	95.1 K	3.5 M	0.36	82.93	90.89	56.16
e	1.00	38.4 K	1.5 M	0.15	77.09	89.62	54.16
Standalone (Liang et al., 2020)	1.00	1.8 M	3.6 M	6.88	16.90	87.93	10.03
FedAvg (Liang et al., 2020)	1.00	1.8 M	3.6 M	6.88	67.74	58.99	58.99
LG-FedAvg (Liang et al., 2020)	1.00	1.8 M	3.6 M	6.88	69.76	91.77	60.79

Table 8: Results of combination of various computation complexity levels for CIFAR10 dataset.

Model	Ratio	Parameters	FLOPs	Space (MB)	Perplexity
a	1.00	19.3 M	1.4 B	73.49	3.37
a-b	0.74	14.2 M	991.4 M	54.12	3.31
a-c	0.62	11.8 M	829.0 M	45.20	3.71
a-b-c	0.57	10.9 M	757.6 M	41.72	3.42
a-d	0.56	10.7 M	754.3 M	40.94	3.74
a-b-d	0.53	10.2 M	707.8 M	38.87	3.53
a-e	0.53	10.2 M	718.6 M	38.86	3.75
a-b-e	0.51	9.8 M	684.0 M	37.49	3.47
b	1.00	9.1 M	614.8 M	34.74	3.46
a-b-c-d	0.45	8.8 M	603.4 M	33.39	3.61
a-c-d	0.45	8.6 M	599.6 M	32.93	4.08
a-b-c-e	0.44	8.5 M	585.5 M	32.34	3.50
a-c-e	0.43	8.3 M	575.8 M	31.54	3.65
a-b-d-e	0.41	7.9 M	548.2 M	30.21	3.64
a-d-e	0.39	7.5 M	526.0 M	28.70	4.02
a-b-c-d-e	0.37	7.2 M	496.6 M	27.55	3.55
b-c	0.74	6.8 M	452.4 M	25.83	3.45
a-c-d-e	0.35	6.8 M	467.0 M	25.76	3.92
b-d	0.62	5.7 M	377.7 M	21.57	3.70
b-c-d	0.58	5.2 M	348.5 M	20.02	3.47
b-e	0.56	5.1 M	342.0 M	19.49	3.90
b-c-e	0.54	4.9 M	324.7 M	18.63	3.46
c	1.00	4.4 M	290.1 M	16.92	3.62
b-c-d-e	0.46	4.2 M	278.7 M	16.07	3.64
b-d-e	0.45	4.1 M	274.9 M	15.79	3.92
c-d	0.75	3.3 M	215.4 M	12.66	3.46
c-e	0.62	2.8 M	179.7 M	10.57	3.89
c-d-e	0.58	2.6 M	166.7 M	9.85	3.66
d	1.00	2.2 M	140.7 M	8.39	3.83
d-e	0.75	1.7 M	105.0 M	6.31	3.90
e	1.00	1.1 M	69.3 M	4.23	7.41

Table 9: Results of combination of various computation complexity levels for WikiText2 dataset.