

**Table 4:** Hyperparameters for all experiments presented in the paper

Experiment	LeNet-300-100 on MNIST	WRN-28-2 on CIFAR10	Resnet-50 on Imagenet
Hyperparameters for training			
Number of training epochs	100	200	100
Mini-batch size	100	100	256
Learning rate schedule (epoch range: learning rate)	1 - 25: 0.100 26 - 50: 0.020 51 - 75: 0.040 76 - 100: 0.008	1 - 60: 0.100 61 - 120: 0.020 121 - 160: 0.040 161 - 200: 0.008	1 - 30: 0.1000 31 - 60: 0.0100 61 - 90: 0.0010 91 - 100: 0.0001
Momentum (Nesterov)	0.9	0.9	0.9
$L^1$ regularization multiplier	0.0001	0.0	0.0
$L^2$ regularization multiplier	0.0	0.0005	0.0001
Hyperparameters for sparse compression ( <i>compressed sparse</i> ) (Zhu & Gupta, 2017)			
Number of pruning iterations ( $T$ )	10	20	20
Number of training epochs between pruning iterations	2	2	2
Number of training epochs post-pruning	20	10	10
Number of epochs during pruning	40	50	50
Learning rate schedule during pruning (epoch range: learning rate)	1 - 20: 0.0200 21 - 30: 0.0040 31 - 40: 0.0008	1 - 25: 0.0200 25 - 35: 0.0040 36 - 50: 0.0008	1 - 25: 0.0100 26 - 35: 0.0010 36 - 50: 0.0001
Hyperparameters for dynamic sparse reparameterization ( <i>dynamic sparse</i> ) (ours)			
Number of parameters to prune ( $K$ )	600	20,000	200,000
Fractional tolerance of $K$ ( $\delta$ )	0.1	0.1	0.1
Initial pruning threshold ( $H^{(0)}$ )	0.001	0.001	0.001
Reparameterization period ( $P$ ) schedule (epoch range: $P$ )	1 - 25: 100 26 - 50: 200 51 - 75: 400 76 - 100: 800	1 - 25: 100 26 - 80: 200 81 - 140: 400 141 - 200: 800	1 - 25: 1000 26 - 50: 2000 51 - 75: 4000 76 - 100: 8000
Hyperparameters for Sparse Evolutionary Training ( <i>SET</i> ) (Mocanu et al., 2018)			
Number of parameters to prune at each re-parameterization step	-	20,000	200,000
Reparameterization period ( $P$ ) schedule (epoch range: $P$ )	-	1 - 25: 100 26 - 80: 200 81 - 140: 400 141 - 200: 800	1 - 25: 1000 26 - 50: 2000 51 - 75: 4000 76 - 100: 8000
Hyperparameters for Deep Rewiring ( <i>DeepR</i> ) (Bellec et al., 2017)			
$L^1$ regularization multiplier ( $\alpha$ )	-	$10^{-5}$	$10^{-5}$
Temperature ( $T$ ) schedule (epoch range: $T$ )	-	1 - 25: $10^{-5}$ 26 - 80: $10^{-8}$ 81 - 140: $10^{-12}$ 141 - 200: $10^{-15}$	1 - 25: $10^{-5}$ 26 - 50: $10^{-8}$ 51 - 75: $10^{-12}$ 76 - 100: $10^{-15}$