Table 3: Test error rates for networks with less than 40M parameters, sorted by CIFAR-100.

Method	# params	C10	C100	I32 Top-1 / Top-5
WRN 22-10 (Zagoruyko & Komodakis, 2016)	27M	4.44	20.75	-
1-bit weights WRN 20-10	**27M**	**4.72**	**19.35**	**44.27 / 21.09**
WRN 28-10 (Chrabaszcz et al., 2017)	37M	-	-	40.97 / 18.87
Full precision WRN 20-10	**27M**	**4.22**	**18.76**	**39.96 / 17.89**
1-bit weights WRN 20-10 + cutout	**27M**	**3.92**	**18.51**	-
WRN 28-10 + cutout (Devries & Taylor, 2017)	34M	3.08	18.41	-
WRN 28-10 + dropout (Zagoruyko & Komodakis, 2016)	37M	3.80	18.30	-
ResNeXt-29, 8×64d Xie et al. (2016)	36M	3.65	17.77	-
Full precision WRN 20-10 + cutout	**27M**	**3.46**	**17.19**	-
DenseNets (Huang et al., 2016)	26M	3.46	17.18	-
ResNext + EraseReLU	36M?	3.56	16.53	-
1-bit weights ResNeXt-29, 8 ×64d + cutout	**36M**	**3.29**	**16.97**	-
Full precision ResNeXt-29, 8 ×64d + cutout	**36M**	**3.32**	**16.12**	-
Shake-shake regularization (Gastaldi, 2017)	26M	2.86	15.97	-
Shake-shake + cutout (Devries & Taylor, 2017)	26M	2.56	15.20	-

Table 4: Test error rates using 1-bit-per-weight at test time and propagation during training.

Method	C10	C100	SVHN	ImageNet top-1/top-5
BC (Courbariaux et al., 2015)	8.27	-	2.30	-
Weight binarization (Merolla et al., 2016)	8.25	-	-	-
BWN - Googlenet (Rastegari et al., 2016)	9.88	-	-	34.5 / 13.9
VGG+HWGQ (Cai et al., 2017)	7.49	-	-	-
BC with ResNet + ADAM (Li et al., 2017)	7.17	35.34	-	52.11
BW with VGG (Cai et al., 2017)	-	-	-	34.5
My approach	**3.29**	**16.97**	**1.93**	**26.04 / 8.48**