Convergence of the training as function of the execution time B = 8B = 166 B = 32B = 64Training loss 20 10 30 Execution time (hours)