A fast algorithm to simulate deep resistive networks with dissipative nonlinearities

F. Osana(1)(2), J. Grollier(1), D. Querlioz(2)

(1) Laboratoire Albert Fert, CNRS, Thales, Université Paris-Saclay, Palaiseau, France

(2) Université Paris-Saclay, CNRS, C2N, Palaiseau, France

Equilibrium Propagation has emerged as a compelling approach to train physical neural networks [1]. Deep Resistive Networks (DRN) are a natural substrate because they exhibit analog behaviour and benefit from a local learning rule [2]. Electrical circuit simulators such as SPICE, which are widely used to model analog networks, provide high precision but suffer from long simulation times, making them impractical for training.
A recent work uses a coordinate-descent (CD) solver to enable the fast simulation of DRN with ideal diodes by framing the solution as a quadratic programming (QP) problem with linear inequality constraints [3]. However, the QP formulation does not cover (more realistic) dissipative nonlinearities that contribute to the objective through convex dissipation potentials. These introduce non-quadratic terms and give rise to implicit nonlinear CD updates. Here, we show that these one-dimensional subproblems can still be solved efficiently, using exact Lambert-W updates for exponential nonlinearities and Newton–Raphson iterations when no closed-form update is available.

We implement the proposed CD solver in PyTorch and train DRNs with one to three hidden layers on Digits while varying both hidden-layer width and nonlinearity. We then sweep the validation set through each trained network using both CD and matched SPICE simulations, compare the resulting steady-state solutions, and compute the relative error. Across all configurations, the P90 relative L_1 error remains below 1.1×10^{-4}, while runtime improves by up to $4400\times$, with larger gains for wider and deeper networks. On MNIST, using the same network as in [2], our implementation reaches 3.0% test error, improving the previously reported result by 0.4% while reducing the training time per epoch by roughly $440\times$.

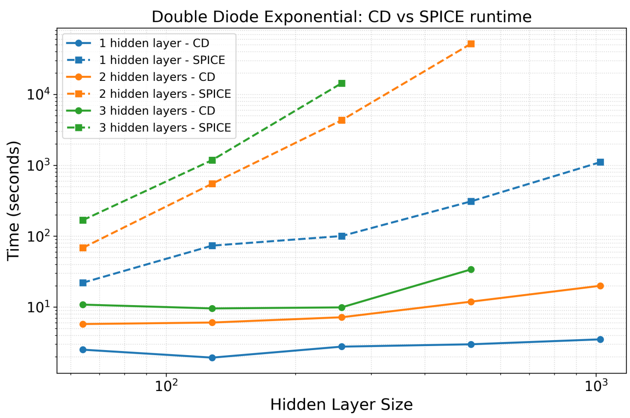

Figure 1: Comparison of runtimes for CD and SPICE simulations of DRNs with Shockley nonlinearities.

This work makes the simulation and training of DRNs with dissipative nonlinearities feasible on full benchmark datasets, and opens the way to more detailed studies of their behavior and scalability. A natural next step is to incorporate nonlinear conductances, non-ideal sources, and other circuit nonidealities.

This work benefited from France 2030 grants (ANR-22-PEEL-0010).

[1] B. Scellier and Y. Bengio, Front. Comput. Neurosci., 11, 24, 2017.

[2] J. Kendall et al., arXiv:2006.01981, 2020.

[3] B. Scellier, arXiv:2402.11674, 2024.