Synaptic Metaplasticity in Binarized Neural Networks

Axel Laborieux[1], M. Ernoult[1,2], T. Hirtzlin[1], D. Querlioz[1]
1: Center for Nanoscience and Nanotechnology, Palaiseau, France
2: Unité mixte de Physique CNRS Thales, Palaiseau, France

CVPR VIRTUAL JUNE 19-25

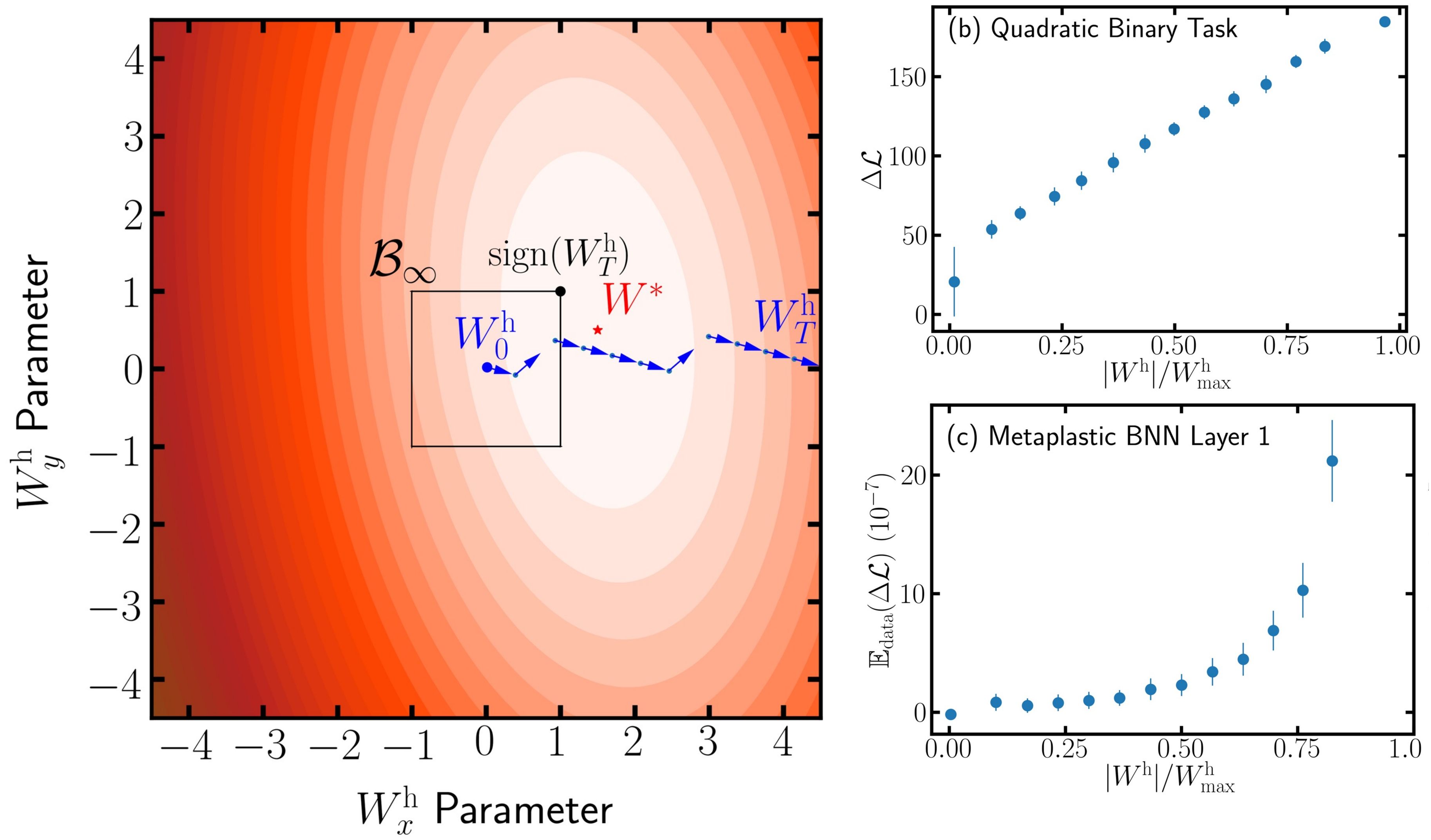
Introduction:

- **Catastrophic forgetting** is an issue common to artificial neural networks in stark contrast with the brain.
- Neuroscience suggest that **real synapses are complex** and **metaplastic**.
- Hidden weights of Binarized Neural Networks are discarded for inference, but what difference does their hidden magnitudes make?

Binary Optimization:

- We consider the quadratic binary task:

$$\mathscr{L}(\mathbf{W}) = \frac{1}{2}(\mathbf{W}-\mathbf{W}^*)^T \cdot \mathbf{H} \cdot (\mathbf{W}-\mathbf{W}^*)$$

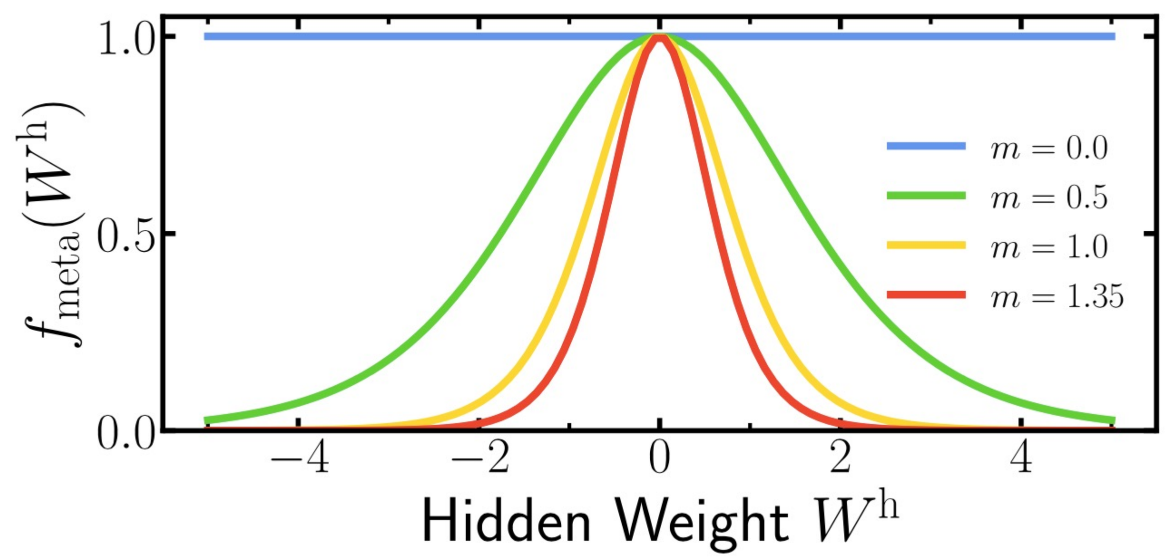


$$\Delta_i \mathscr{L}(\mathbf{W}_t) \sim 2\lambda_i + 2\frac{|\widetilde{W}_i^{\mathrm{h}}|}{\eta} \quad \text{as} \quad t \to +\infty.$$

Consolidation strategy:

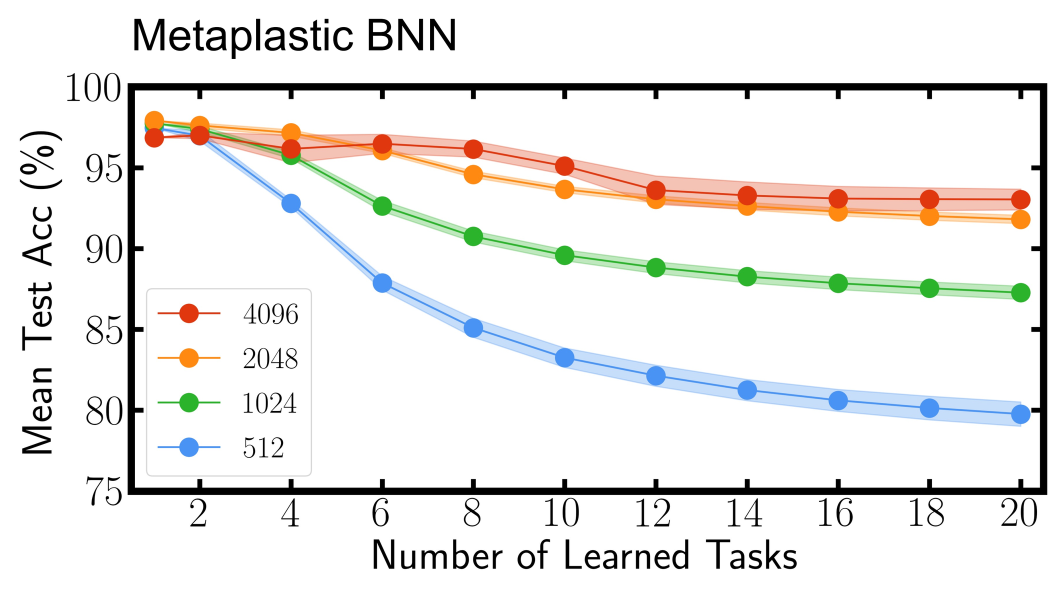
- The higher the hidden magnitude of the binarized weight, the more difficult to switch to the opposite sign.

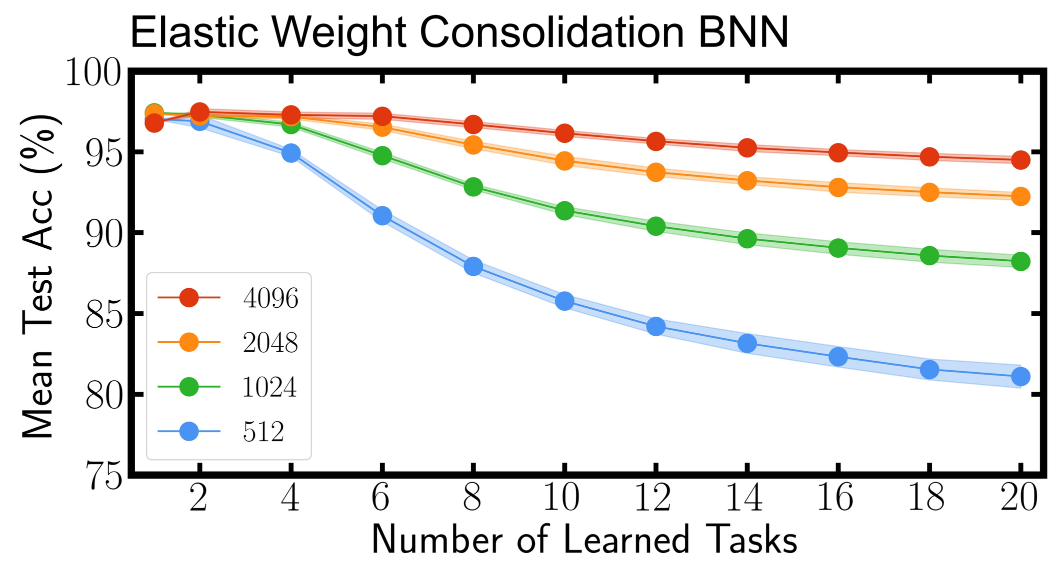
$$W^{\mathrm{h}} \leftarrow W^{\mathrm{h}} - \eta U_W \cdot f_{\mathrm{meta}}(m, W^{\mathrm{h}}) \quad \text{if} \quad U_W W^{\mathrm{h}} > 0$$
$$W^{\mathrm{h}} \leftarrow W^{\mathrm{h}} - \eta U_W \quad \text{otherwise.}$$



Continual learning:

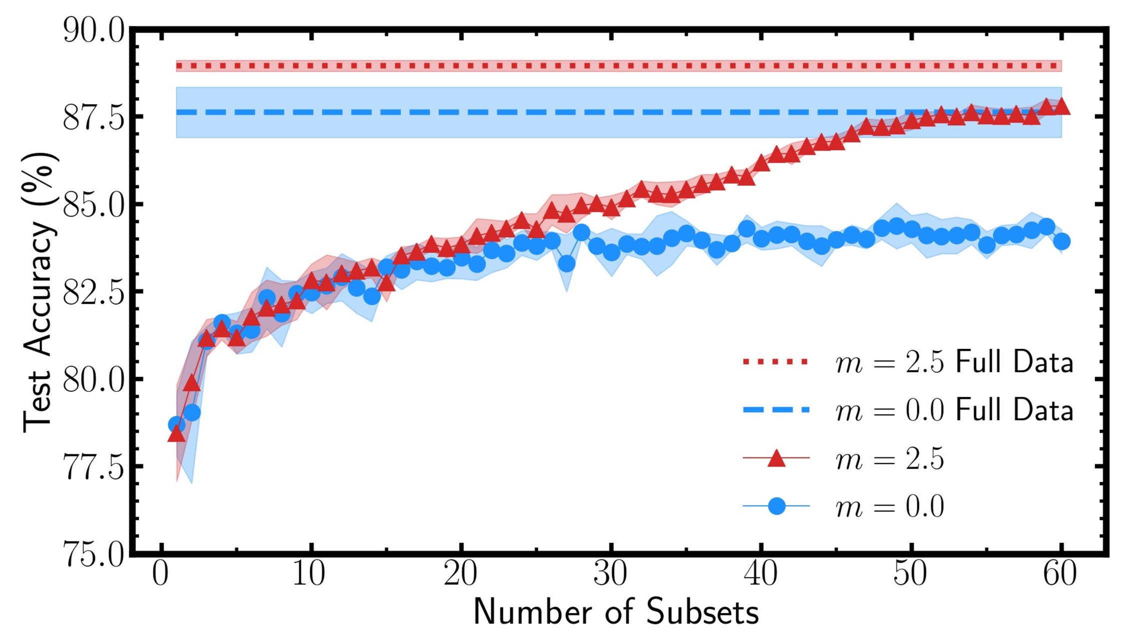
- Our approach performs similarly to Elastic Weight Consolidation on the permuted MNIST continual learning benchmark.
- Can also learn more complex sequences like MNIST -> Fashion MNIST (see paper).
- No need for tasks boundaries.

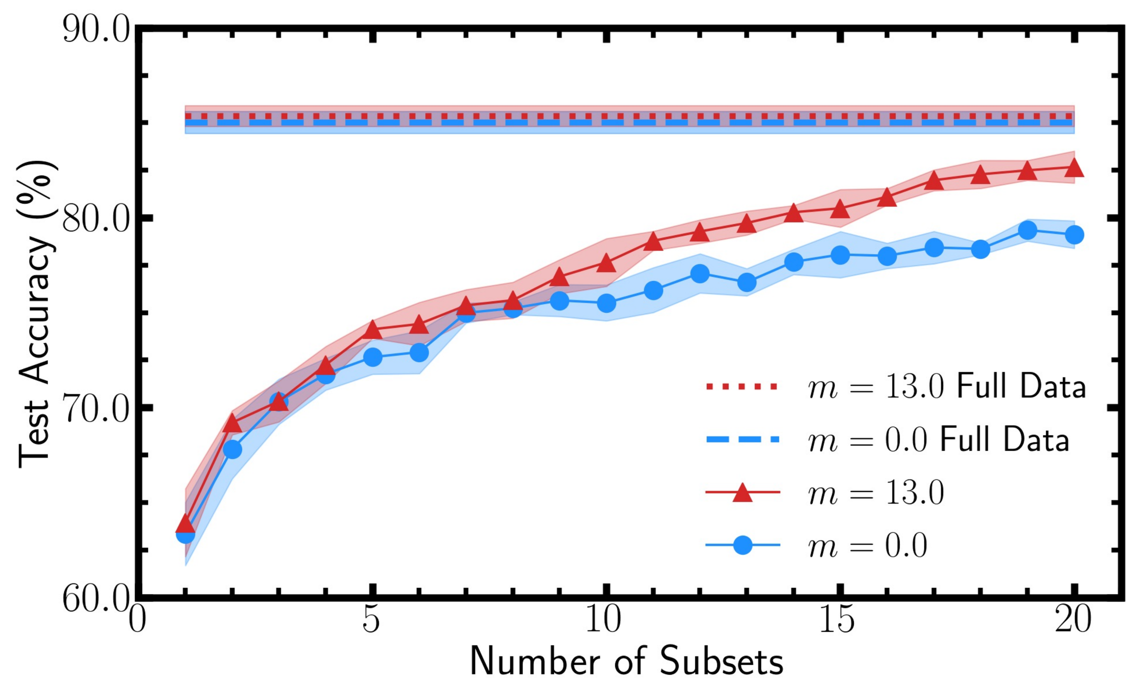




Stream Learning:

- The network is a learning a given task by **learning sequentially subsets of the whole dataset**.
- The accuracy of the Vanilla network plateaus because of it forgets previous subsets.
- Also works on CIFAR-10.





Summary/Conclusion

1. Binarized Neural Networks **hidden weights are relevant** for continual learning.
2. We show a **principled explanation** for a tractable sub problem.
3. The resulting consolidation strategy **does not need task boundaries** and can be applied to Continual learning and Stream learning.

Paper: https://www.nature.com/articles/s41467-021-22768-y
Code: https://github.com/Laborieux-Axel/SynapticMetaplasticityBNN
Contact: axel.laborieux@c2n.upsaclay.fr