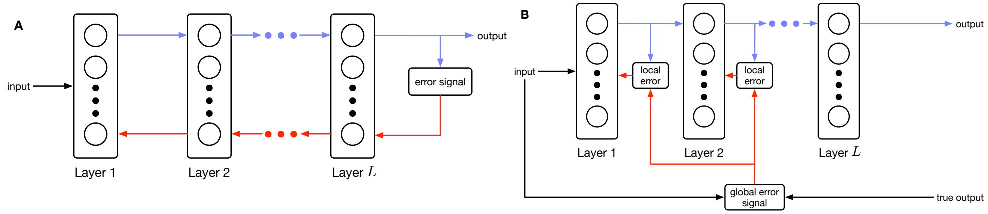
# A Biologically Plausible Learning Rule Based on the Information Bottleneck

Kyle Daruwalla and Mikko Lipasti | daruwalla@wisc.edu

## Motivation

Deep ANNs work! → what to copy?

💡 Copy the architecture not the loss!




layer output

network output

network input

$$\mathcal{L}_{\text{HSIC}} = \underbrace{\boxed{\text{HSIC}(z^\ell, x)}}_{\text{compression}} - \underbrace{\boxed{\gamma \text{HSIC}(z^\ell, y)}}_{\text{prediction}}$$

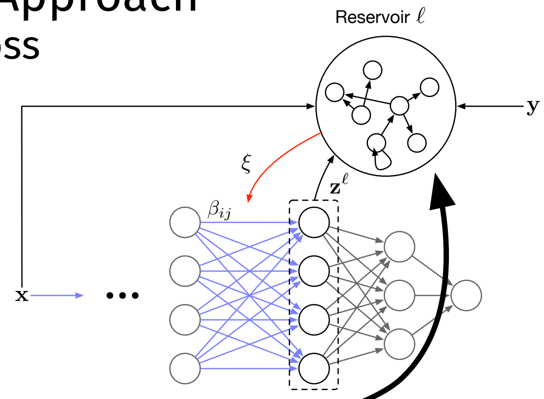
## Our Approach

Take gradient of HSIC loss

$$\Delta [W^\ell]_{ij} \propto \beta_{ij} \xi_i$$

local

$$\beta_{ij} = \partial z_0^\ell / \partial [W^\ell]_{ij} = (1 - ([z_0^\ell]_i)^2)[z_0^{\ell-1}]_j$$




depends on N samples

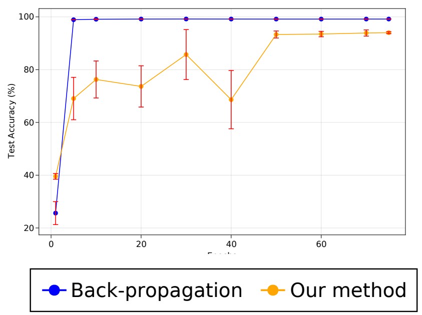
global

$$\xi_i = \sum_{p=0}^{-(N-1)} [\bar{k}(x_0, x_p) - \gamma \bar{k}(y_0, y_p)] \bar{\alpha}(z_p^\ell)$$

teach reservoir to provide global signal

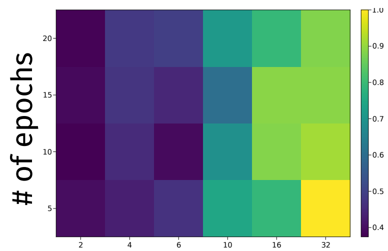
## Our Results

Final test accuracy (MNIST)




Comparable to backprop on MNIST

Normalized test accuracy (MNIST)




Capacity of reservoir matters!

## Conclusions

- Focus on layer-wise objectives
- Directly incorporate past samples with an auxiliary memory
- Better understanding of memory-modulated learning

Future: local term can be made spike-time dependent for certain neuron models

- spike is decided from probabilistic activation
- derivative of activation is the original activation