Meta-Learning Backpropagation & Improving It

Variable Shared Meta Learning (VS-ML)

@LouisKirschAI

Louis Kirsch

Jürgen Schmidhuber

NEURAL INFORMATION PROCESSING SYSTEMS

Background: General Meta Learning

- **General Meta Learning** seeks to **automate learning algorithm design**
- Learning Algorithms (LAs) should be **reusable across a wide range of environments / tasks**
- **Current Meta Learning** either **doesn't generalize very well** or **assumes many human engineered components** (e.g. backpropagation & loss functions)

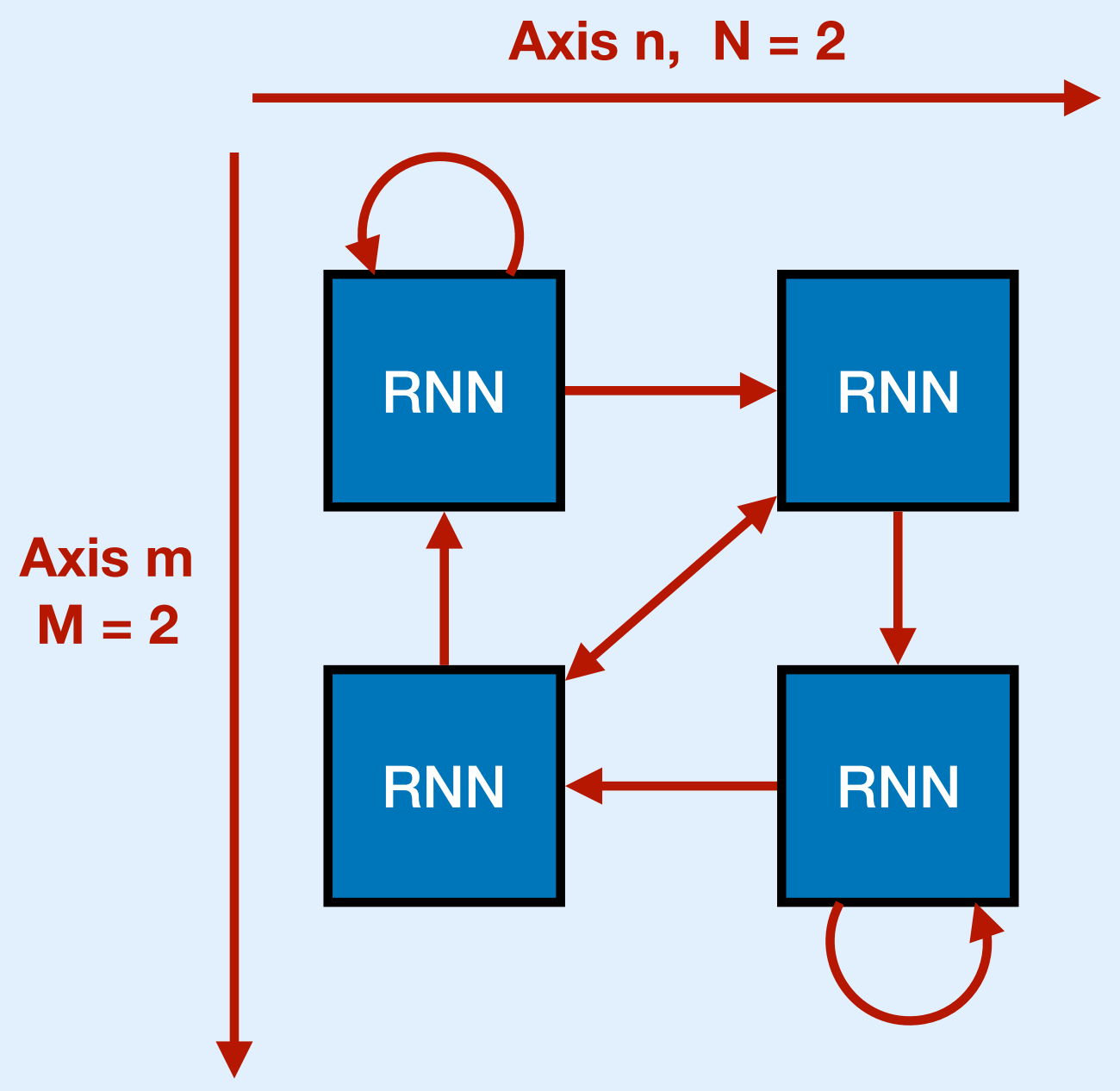
VS-ML $|V_L| \gg |V_M|$

V_M **Meta variables that encode the learning algorithm**

V_L **Learned variables that encode what is being learned**

Variable Shared Meta Learning

- Very simple: Take a 'neural network' and add **variable sharing and sparsity**
- Meta-learned **LAs generalize** due to many more $|V_L|$ (here activations) than $|V_M|$ (here parameters of the network)
- **Unifies** Learned Learning Rules / Fast Weights & Meta RNNs
- **Everything is a variable:** variables we thought of as activations can be interpreted as weights, LSTM defines time-scales

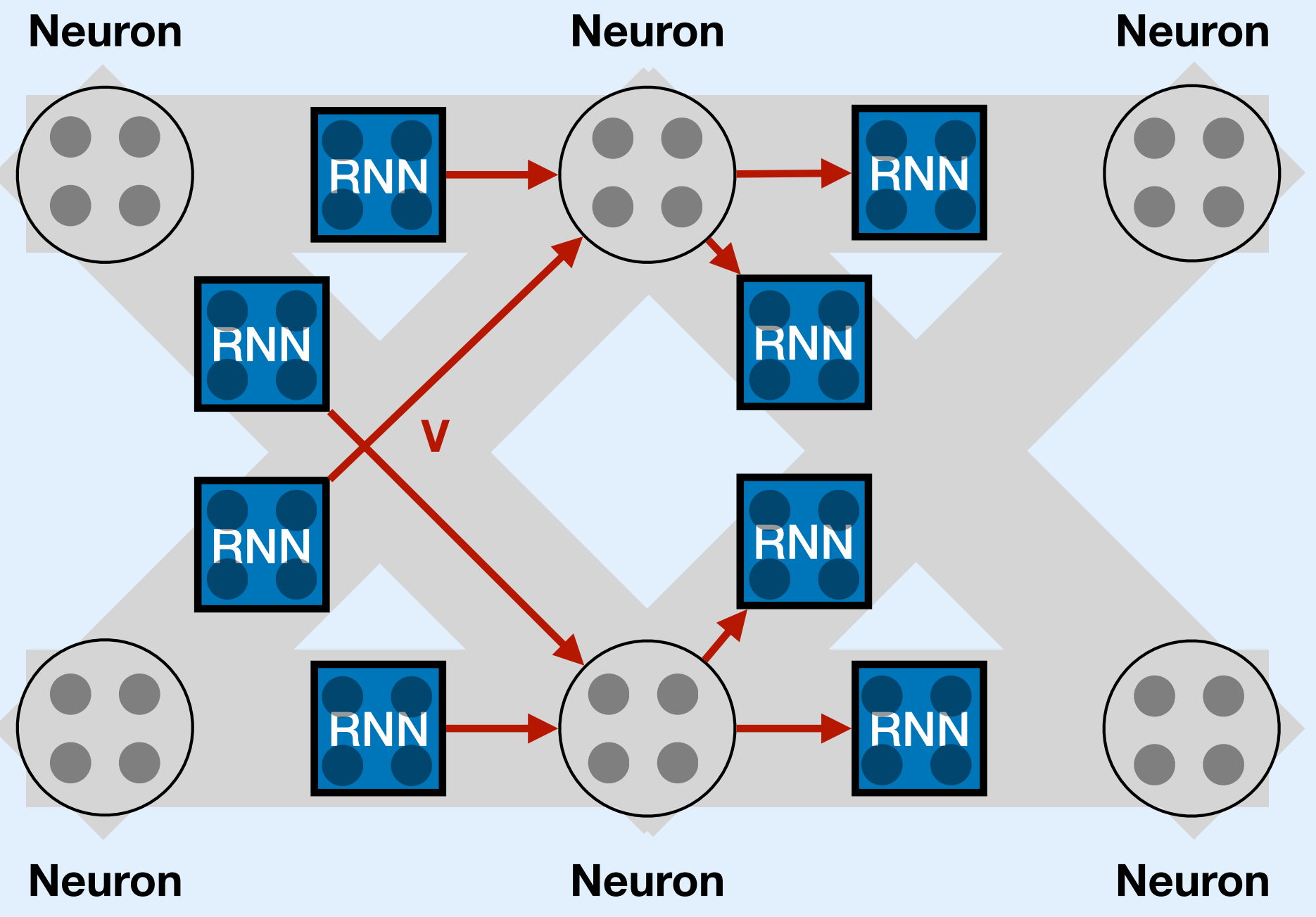


Variable Shared Meta RNN

- Multiple RNNs with shared parameters V_M
- RNNs are connected (e.g. like weights in a neural network)
- Can implement backpropagation and other learning algorithms

Neural interpretation

- Can be interpreted as more complex weights and neurons



Join my invited talk at the Meta Learn Workshop
11 Dec 16:00 UTC

VS-ML generalizes Learned Learning Rules & Fast weights

Learning in Weights

- **Versatile** approach
- **Learning rule** (weight update) 'augments' the standard neural network

$$\Delta w_{ab}(t+1) = f_\theta(x_a(t), x_b(t+1), w_{ab}(t), r_{ab}(t))$$

V_M ↑ f_θ ; V_L ↑ $w_{ab}(t)$

- But **introduces a lot of complexity & design decisions**

Malsburg [1981], Bengio et al [1992], Schmidhuber [1992, 1993], Schlag et al [2017], Miconi et al [2019], Gregor, Najarro, Randazzo [2020]

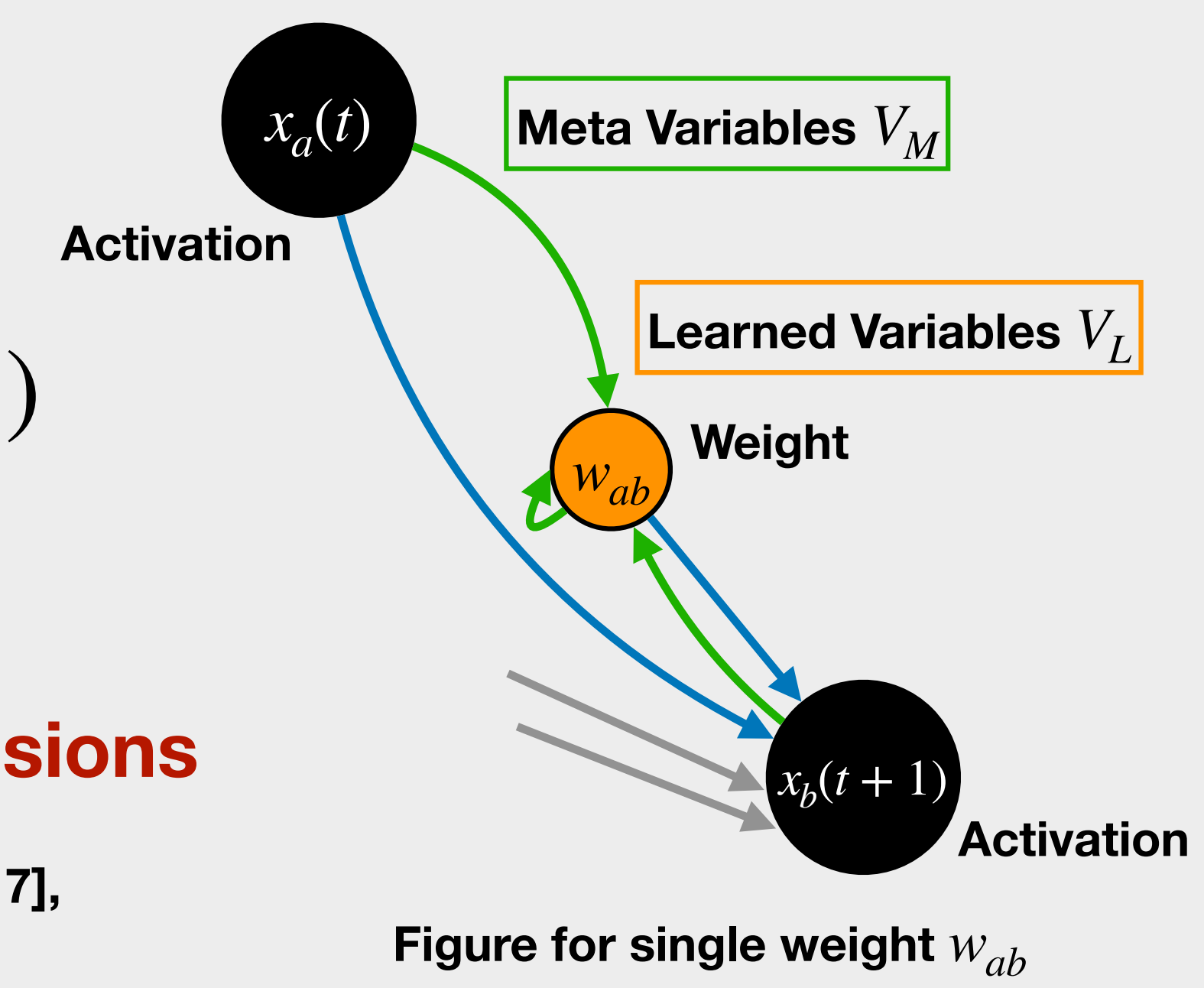
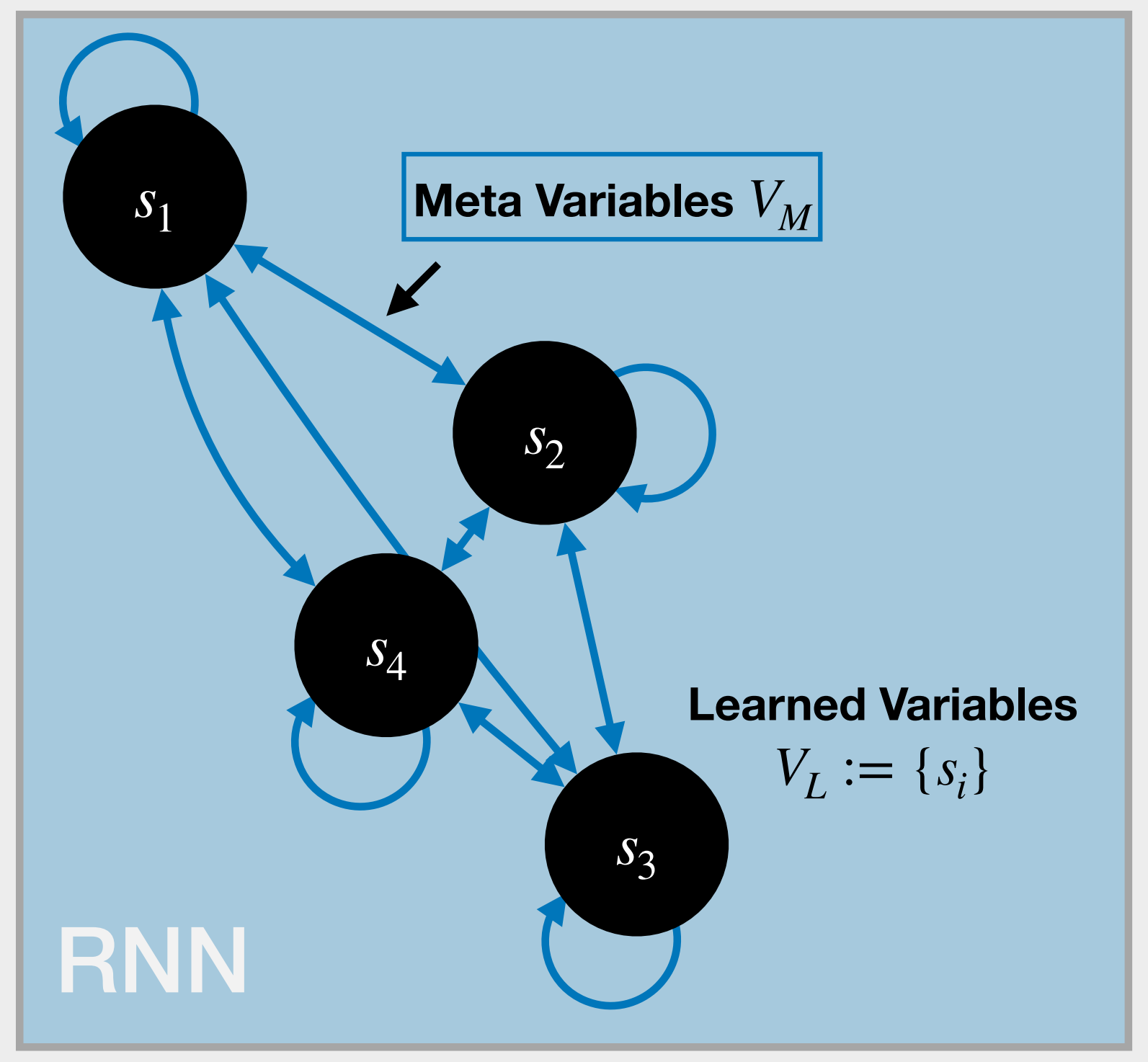


Figure for single weight w_{ab}



Generalize update rules

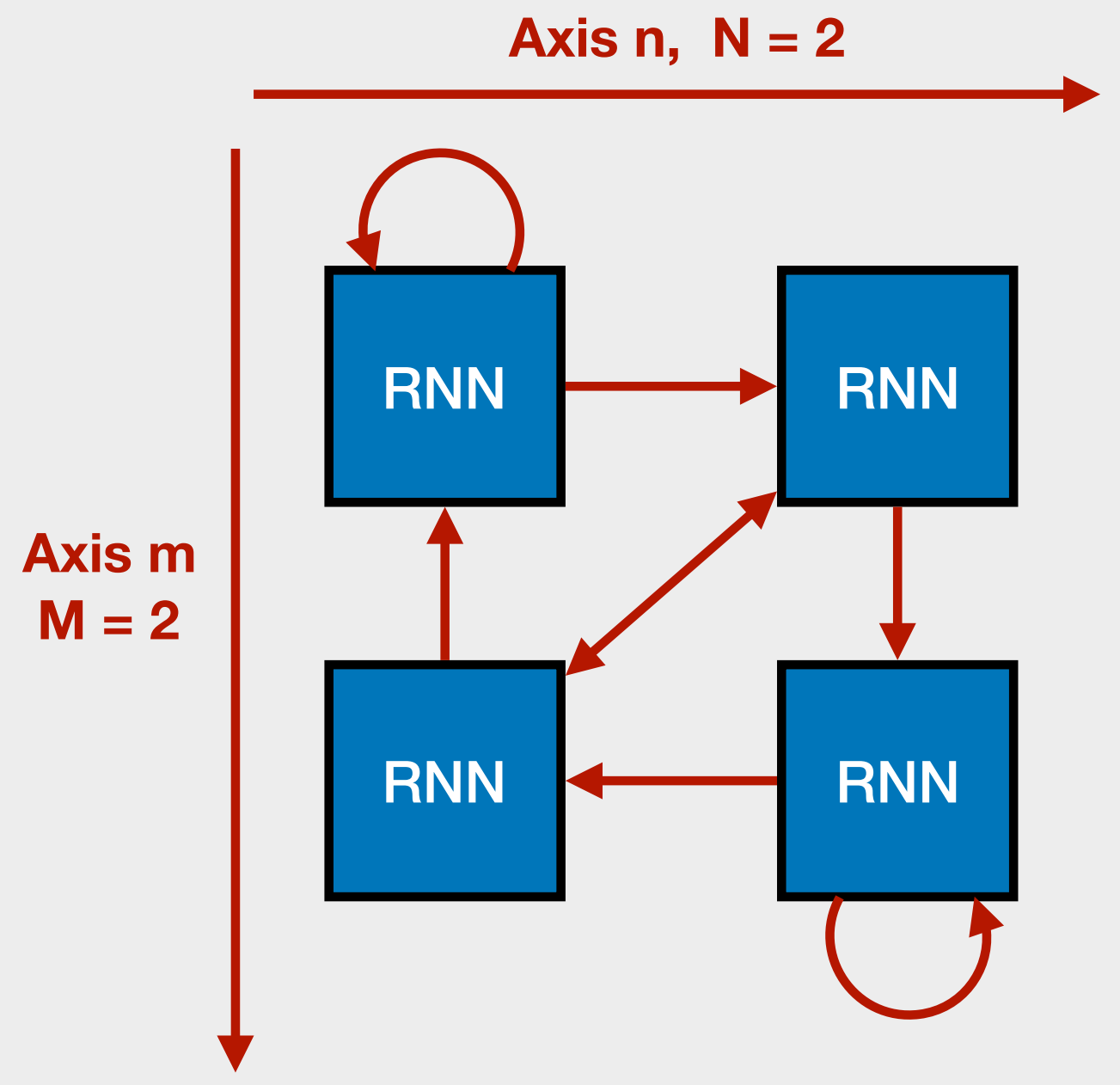
- Interpreting the **weight as just another activation, everything is a variable**
- All activations are learned variables
- Becomes a **generic RNN** (LSTM)
- Now the forward function is also learned

$$s_j \leftarrow \sigma(b_j + \sum_i s_i W_{ij})$$

1 RNN ≙ 1 Weight + Activations

Duplicate RNNs & Choose interactions

here like weights in a neural network



$$s_{mnj} \leftarrow \sigma(b_j + \sum_i s_{mni} W_{ij} + \underbrace{\sum_{c,i} s_{cmi} V_{ij}}_{\text{interactions}} \quad)$$

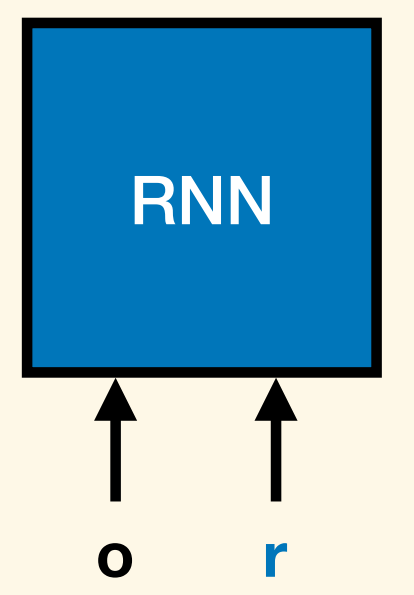
Shared meta variables V_M ↑ b_j, ↑ W_{ij}

N * M RNNs implement an entire neural network including its learning algorithm

Interpretation of VS-ML as a Meta RNN

Learning in Activations

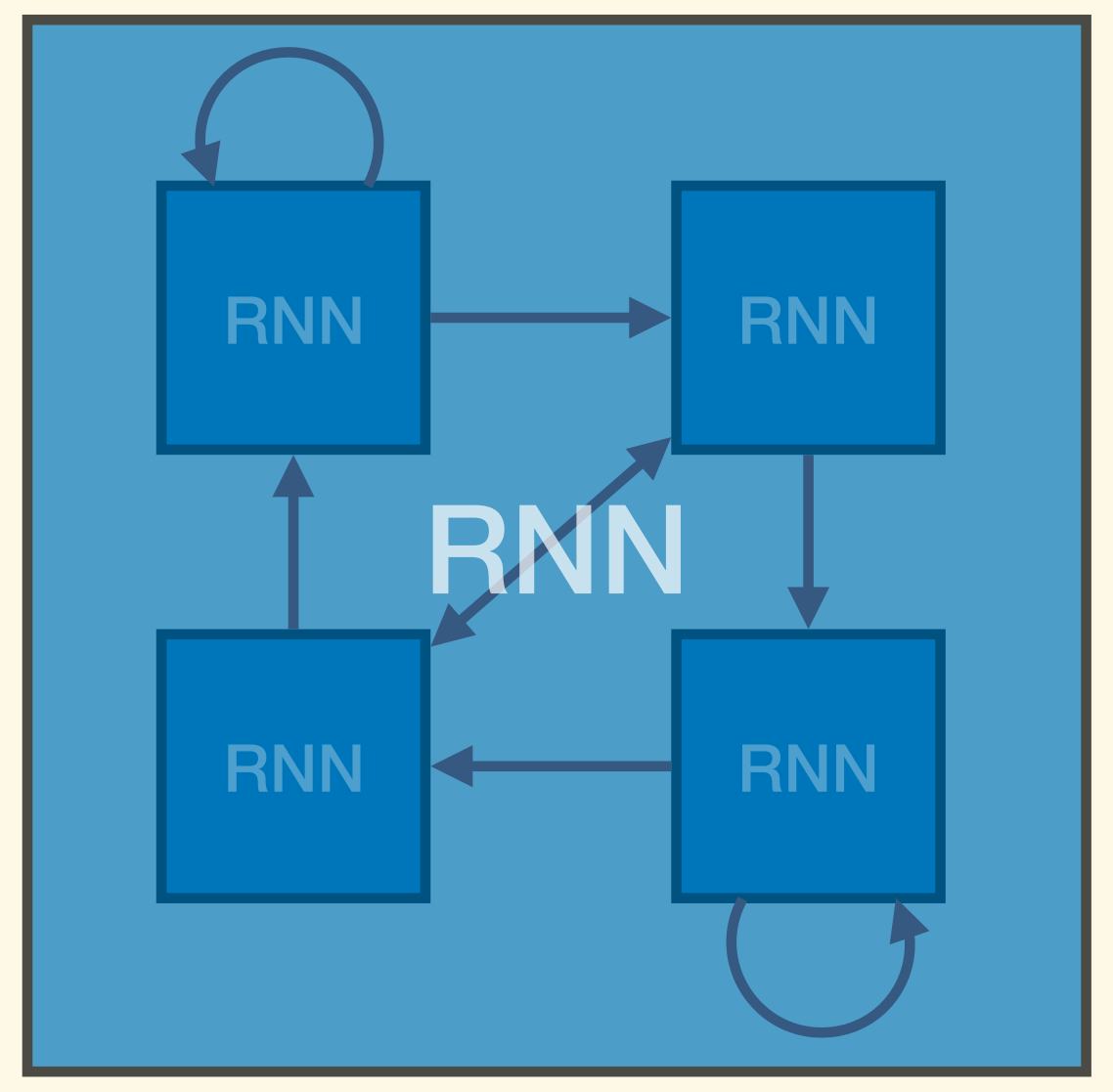
- **Simple model** - just a vanilla RNN / LSTM
- **Meta Variables** V_M : Parameters of a RNN
- **Learned Variables** V_L : State / Activations of an RNN
- **Crucial ingredient: Feedback signal** (e.g. reward) needs to be fed as an input [Hochreiter et al 2001]



- LA is overparameterized, prone to **overfitting** [Kirsch 2020] $|V_L| \in O(N)$ and $|V_M| \in O(N^2)$

$$|V_L| \ll |V_M|$$

Hochreiter et al [2001], Duan et al [2016], Wang et al [2016]



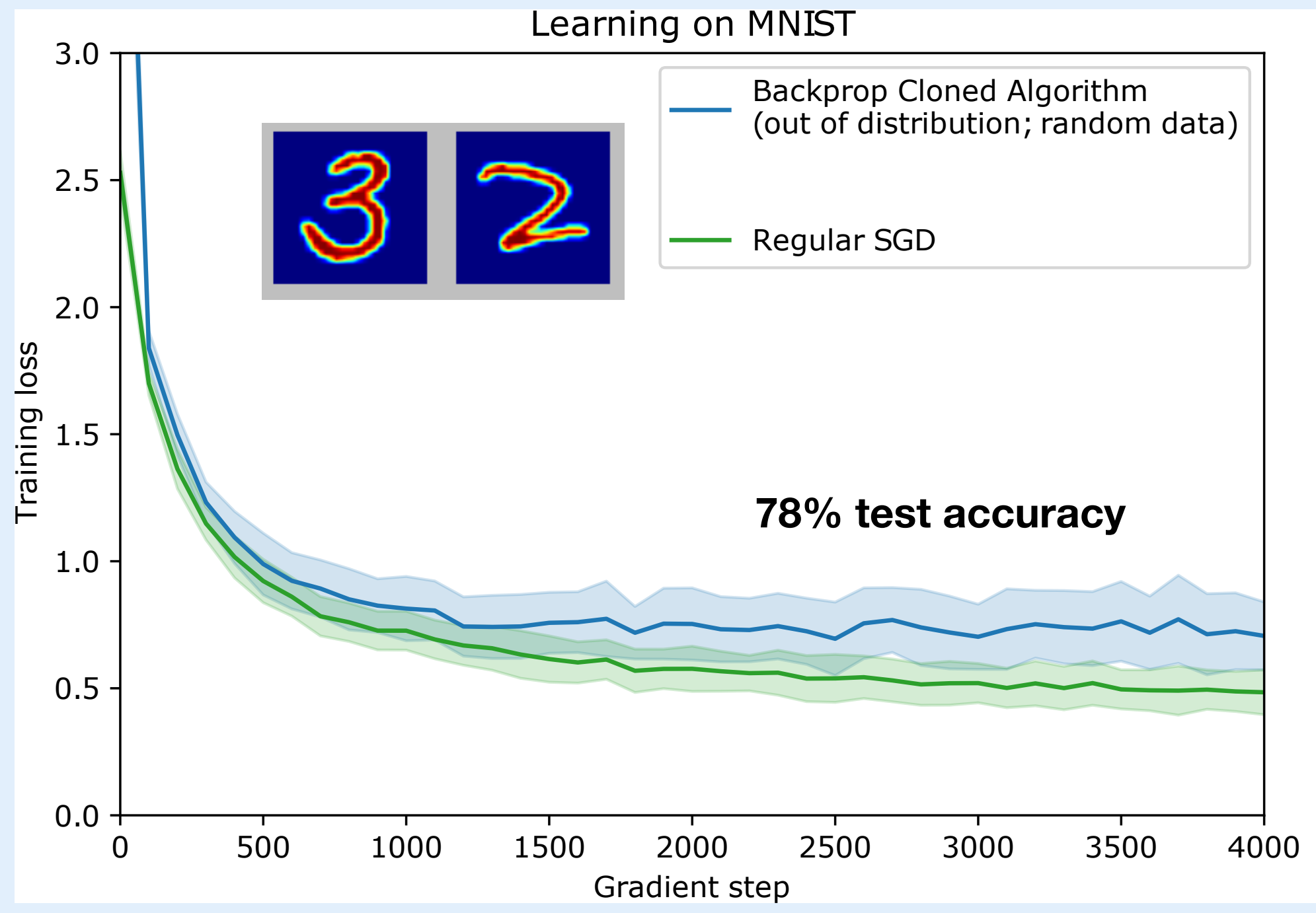
Share RNN weight matrix entries

1. Our method reduces to a **Meta RNN** for $M = 1$ and $N = 1$
2. Or for any M and N to a **Meta RNN** with **shared and zero entries** in its 'weight matrix'

Experiments

Meta Learning Backpropagation

- Can an LSTM implement backpropagation purely in its recurrent dynamics? Yes!
- A sort of 'Learning Algorithm Cloning'

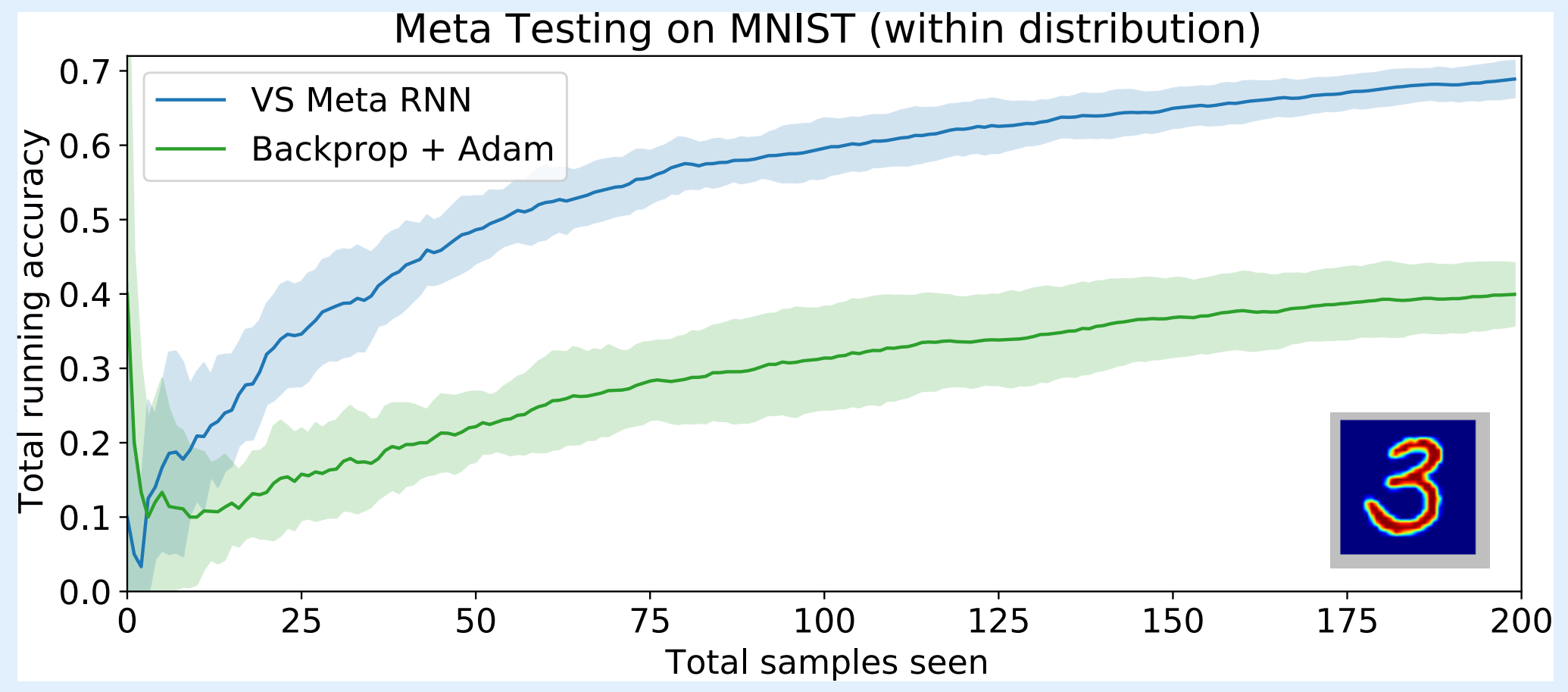


Meta Learning Supervised Learning from Scratch

- Learning is completely **online**, **no batching**, **no pre-training**

Learns on MNIST →

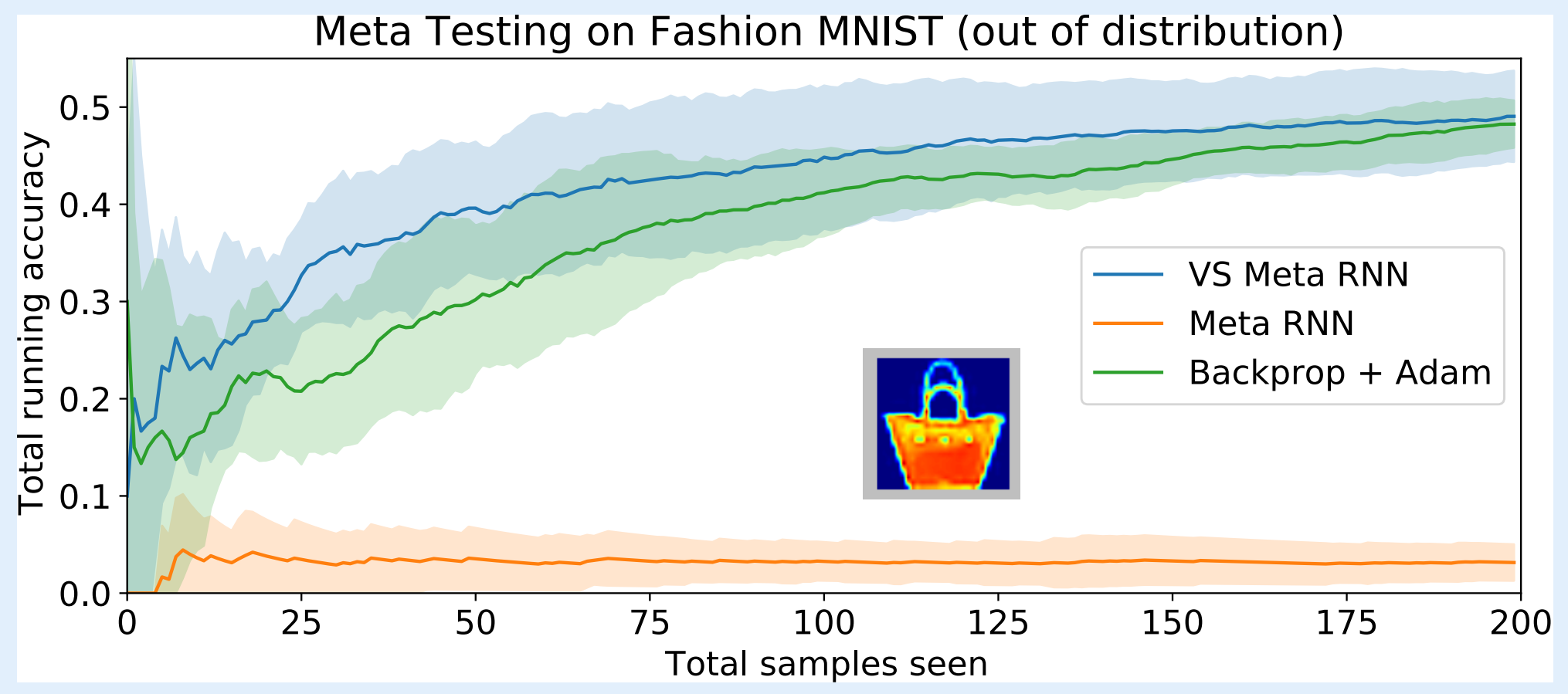
✓ **Sample efficient**



- Use LSTM and run on **Fashion MNIST** (**out of distribution**)

✓ **Strong generalization**

✓ **No gradients required during meta-testing**


Meta Testing on Fashion MNIST (out of distribution)
VS Meta RNN
Meta RNN
Backprop + Adam
Total running accuracy
Total samples seen