Neural Turing Machines

Presented by Nish Khandwala

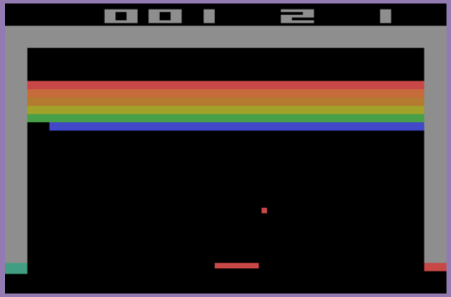
Papers covered:

- "Neural turing machines."
  Graves, Alex, Greg Wayne, and Ivo Danihelka. arXiv:1410.5401 (2014).
- "Hybrid computing using a neural network with dynamic external memory."
  Graves, Alex, et al. Nature 538.7626 (2016): 471-476.

The Problem

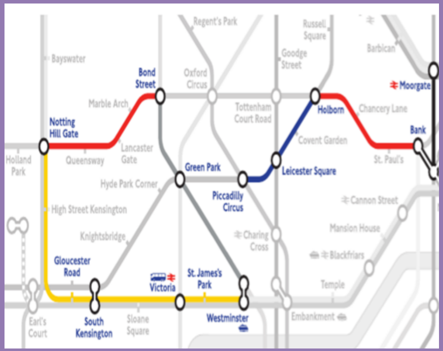
The neural networks (NN) we have seen in class so far excel at pattern recognition and reactive decision making but are unable to *deliberate or reason using knowledge*.

Easy for NN



Playing Breakout

Hard for NN



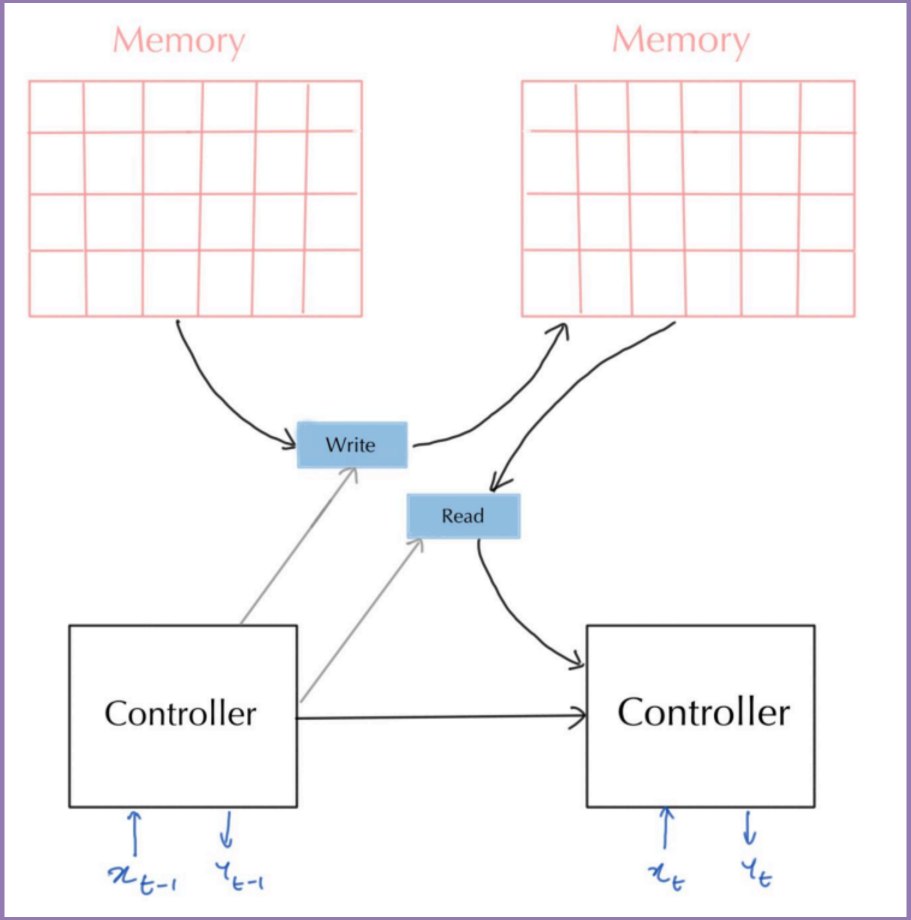
Finding Shortest Path

The Solution? Memory...

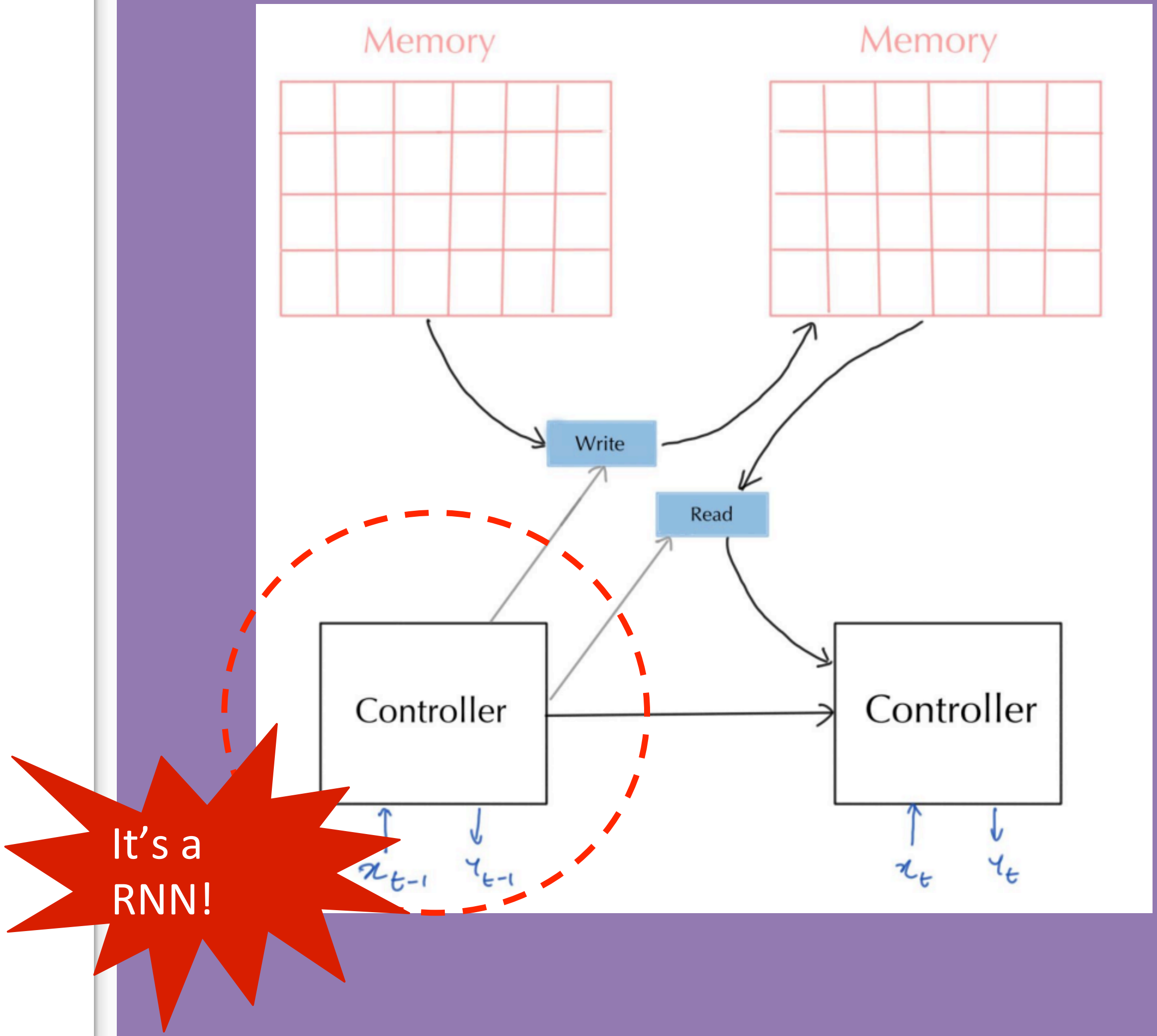
Wait, I thought LSTMs had memory cells...?

Yes, but we need an *external* memory bank. If a LSTM's memory cell is a cache, what we need is RAM.

Neural Turing Machines

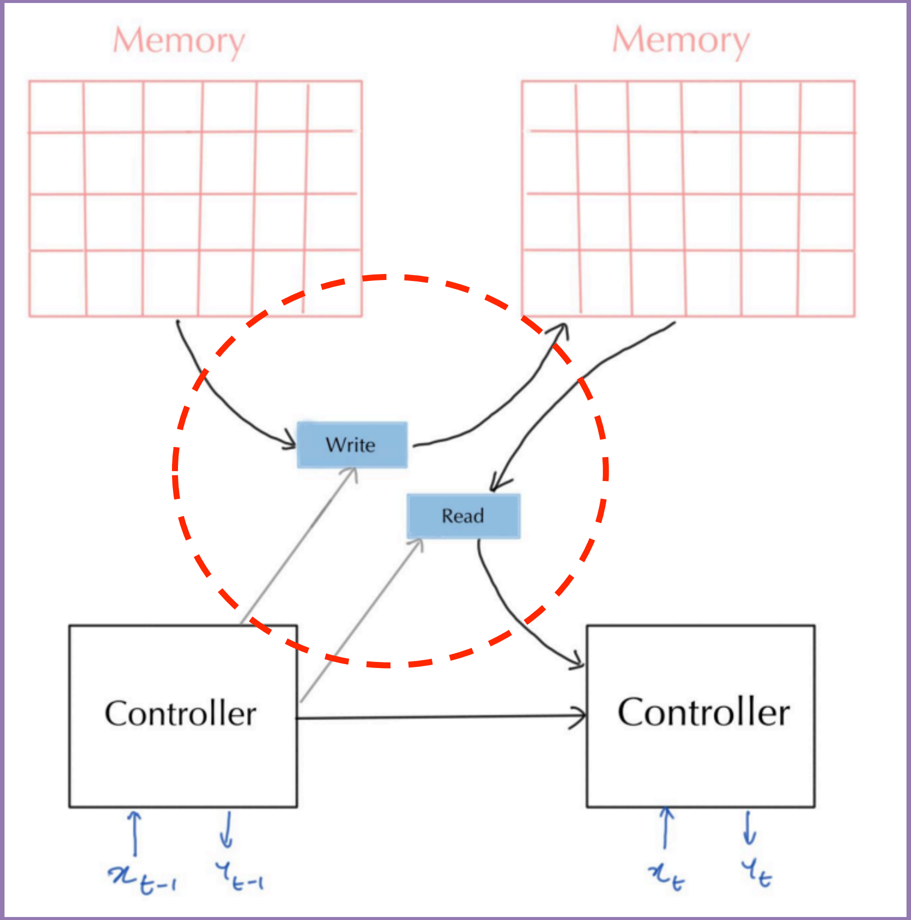


Neural Turing Machines



It's a RNN!

Neural Turing Machines



How does reading and writing work?

Reading from and writing to memory should be differentiable with respect to the location we need to operate on.

But this is tricky! Why?

How does reading and writing work?

Reading from and writing to memory should be differentiable with respect to the location we need to operate on.

But this is tricky! Why? Memory is fundamentally discrete. :(

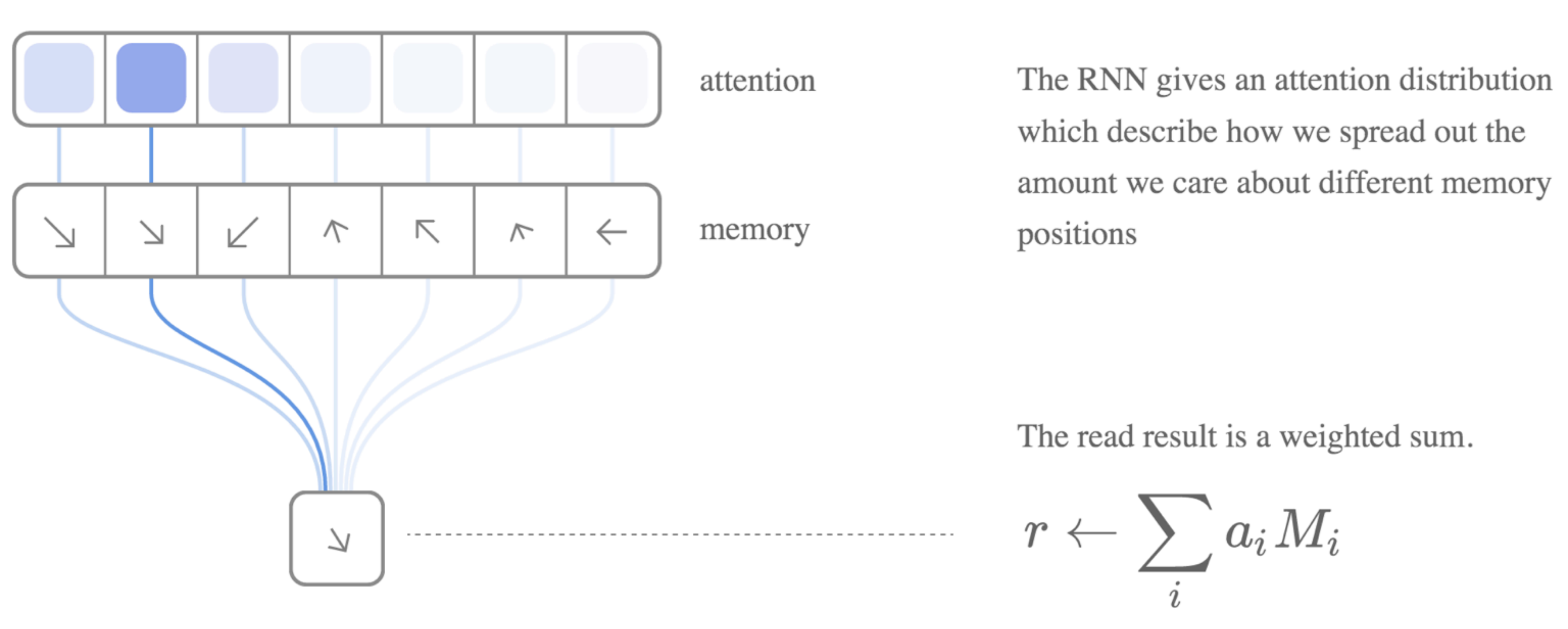
How does reading and writing work?

Reading from and writing to memory should be differentiable with respect to the location we need to operate on.

But this is tricky! Why? Memory is fundamentally discrete. :(

Solution? Read and write everywhere, just to different extents.. (think attention)

Reading from Memory

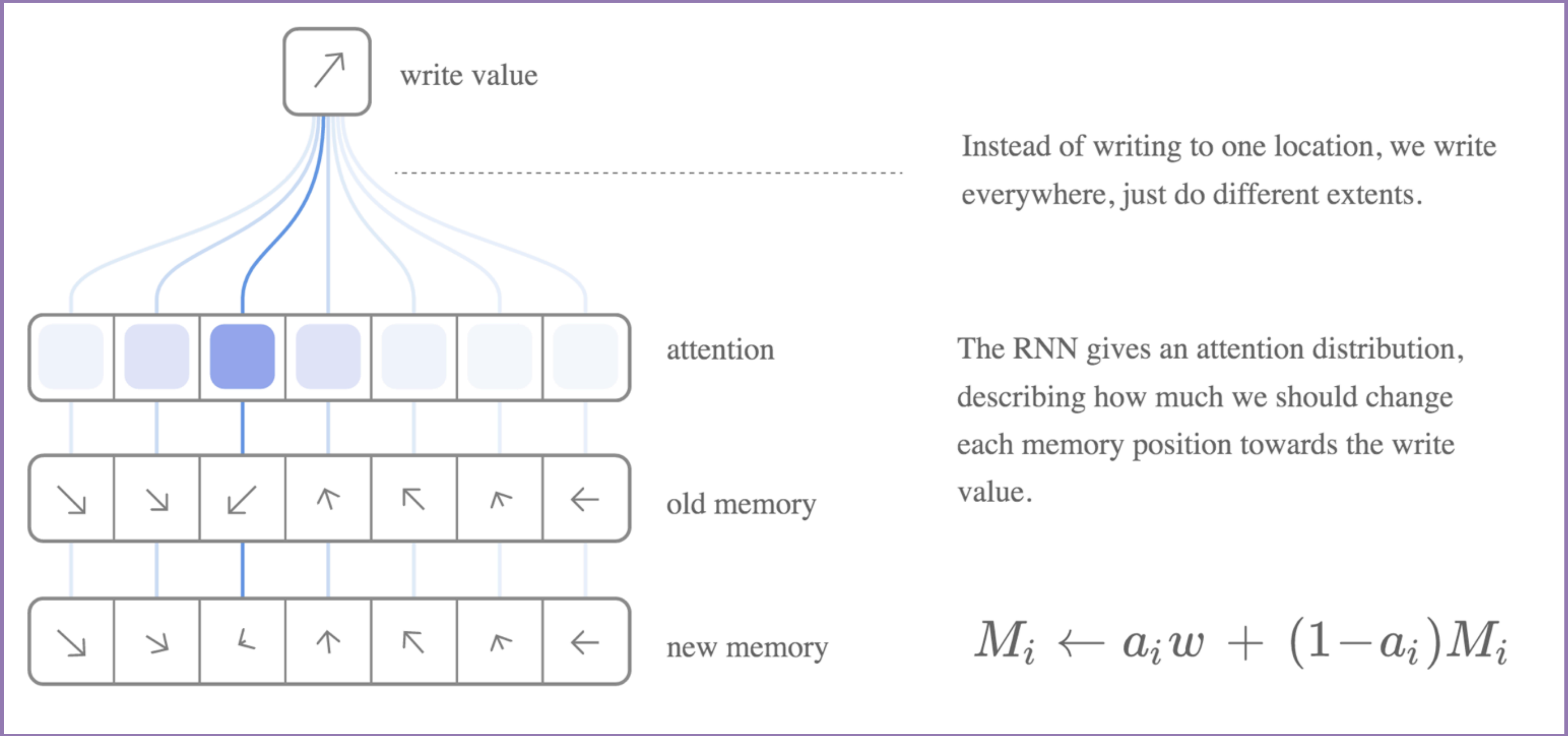


The RNN gives an attention distribution which describe how we spread out the amount we care about different memory positions

The read result is a weighted sum.

$$r \leftarrow \sum_i a_i M_i$$

Writing to Memory

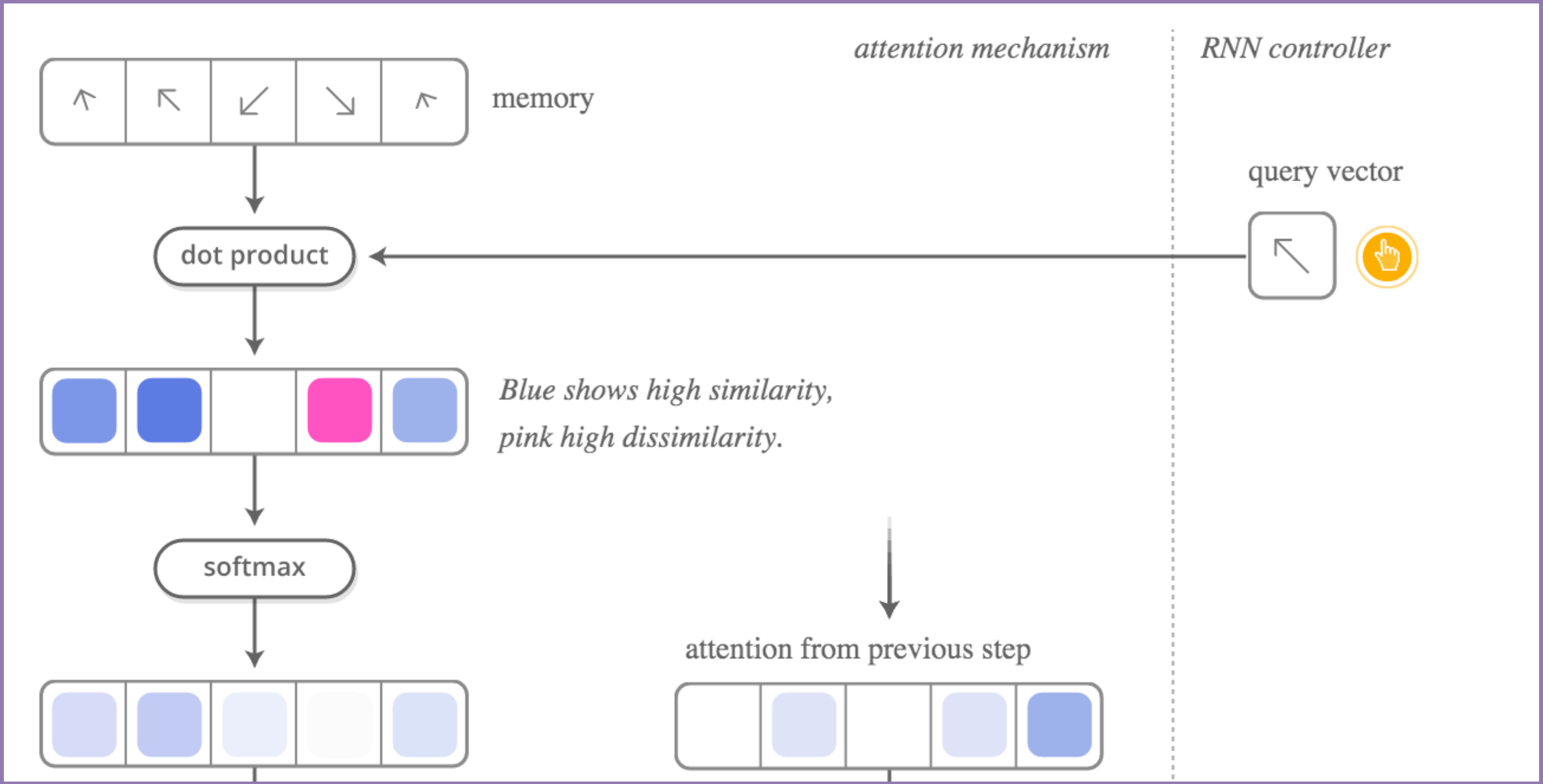


Instead of writing to one location, we write everywhere, just do different extents.

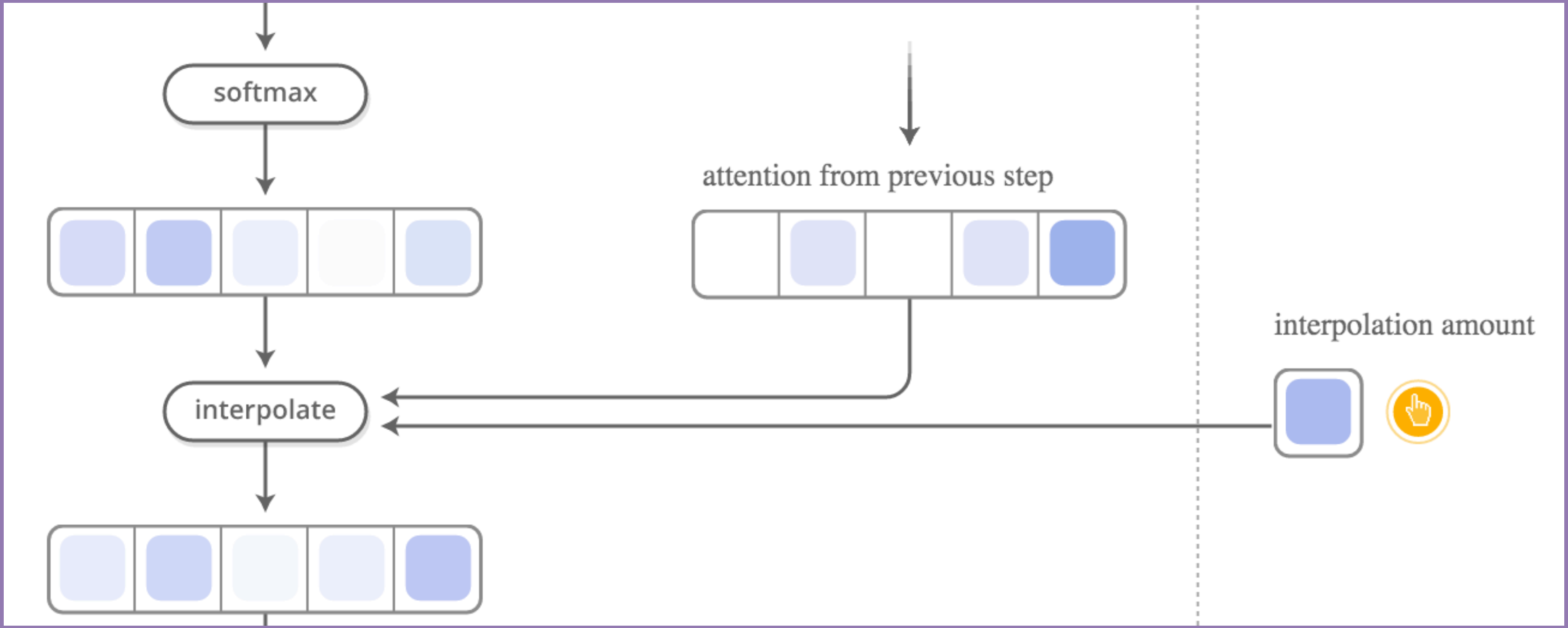
The RNN gives an attention distribution, describing how much we should change each memory position towards the write value.

$$M_i \leftarrow a_i w + (1 - a_i) M_i$$

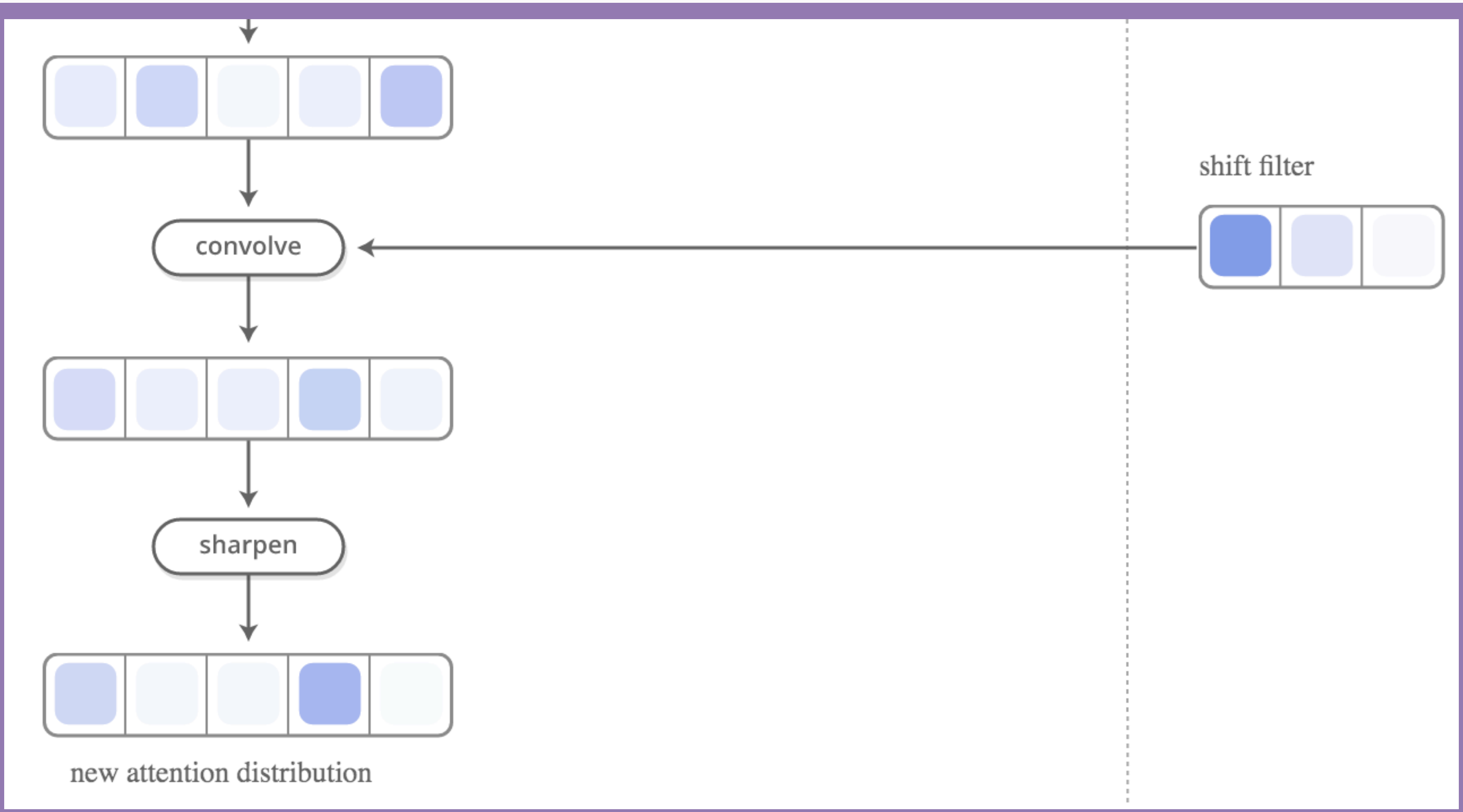
Updating Attention - 1/3



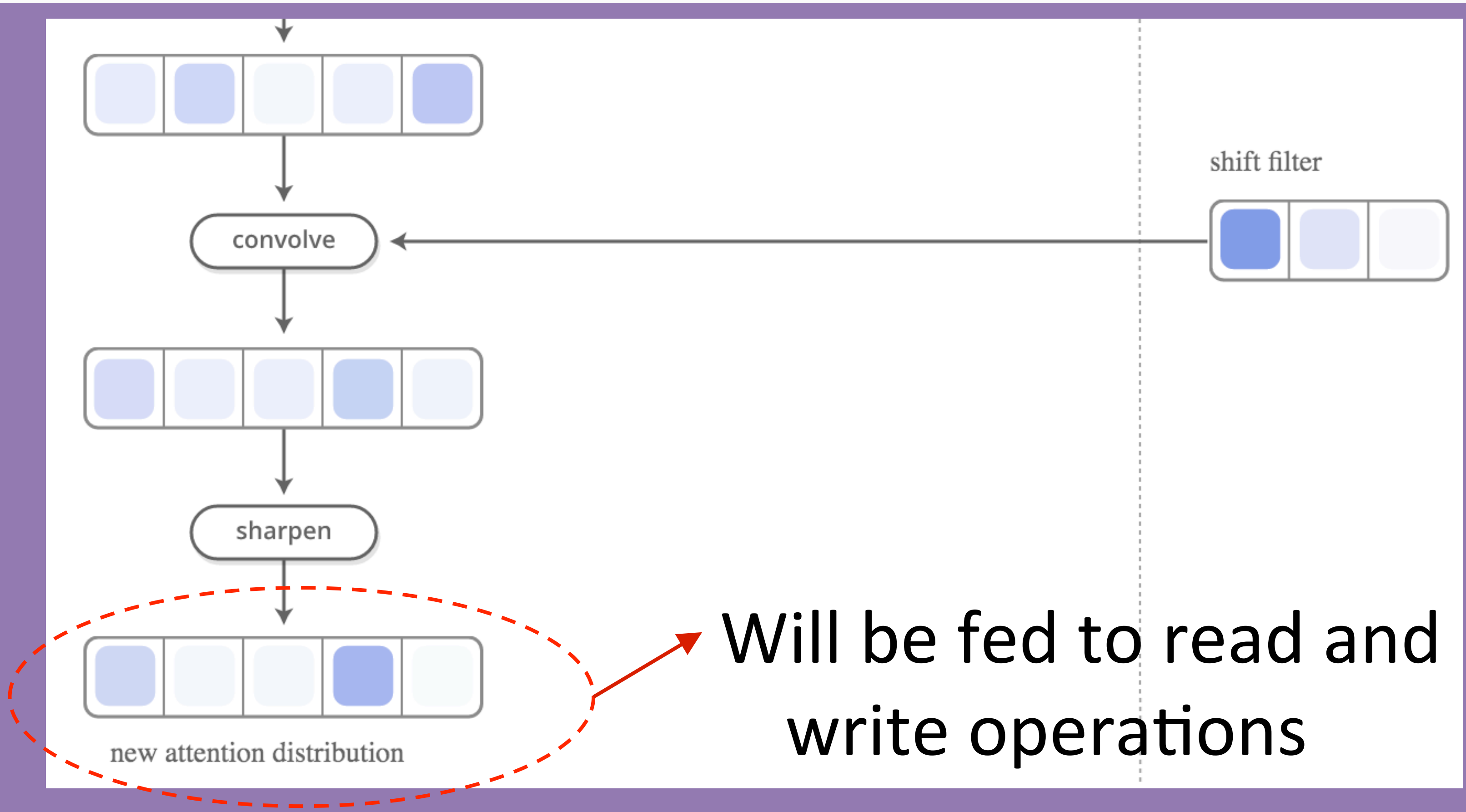
Updating Attention - 2/3



Updating Attention - 3/3



Updating Attention - 3/3



Will be fed to read and write operations

Results*

* using a differentiable neural computer (follow up architecture to NTMs)

Acknowledgements

- Olah & Carter, "Attention and Augmented Recurrent Neural Networks", Distill, 2016.

- https://deepmind.com/blog/differentiable-neural-computers/