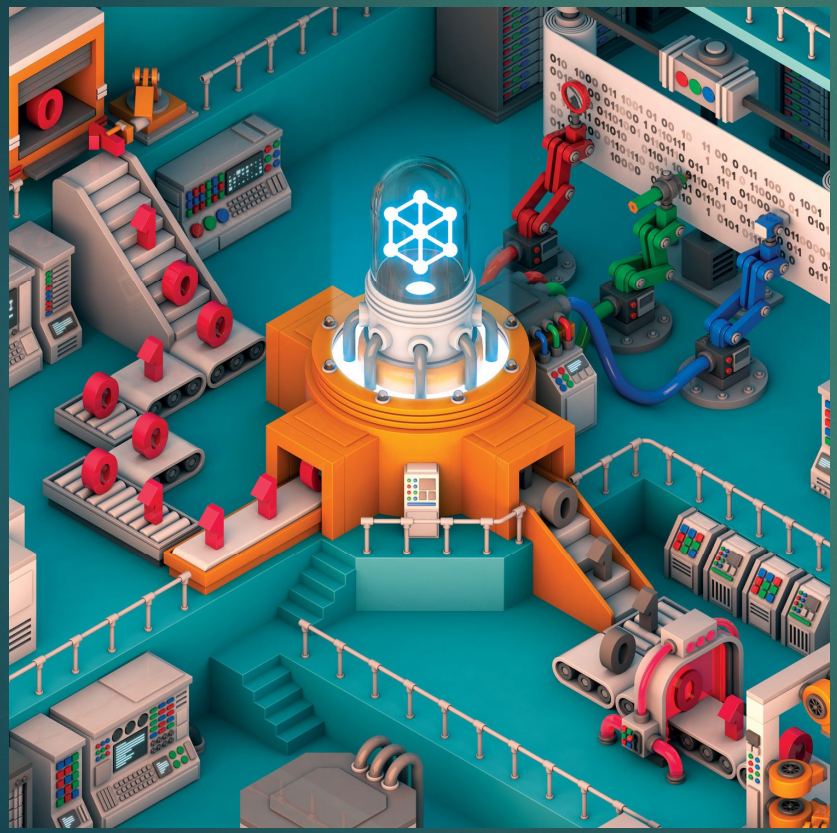
Ms.Comsci KMITL , Subject : Seminar I 1st

Hybrid computing using a neural network with dynamic external memory

from Alex Graves et al.



[1]

DNC

presented by

Lada Phonrungwong

[1] https://deepmind.com/blog/differentiable-neural-computers/

Hybrid computing using a neural network with dynamic external memory

DNC & NTM

Graves, Alex; Wayne, Greg; Reynolds, Malcolm; Harley, Tim; Danihelka, Ivo; Grabska-Barwińska, Agnieszka; Colmenarejo, Sergio Gómez; Grefenstette, Edward; Ramalho, Tiago (2016-Oct). Nature.

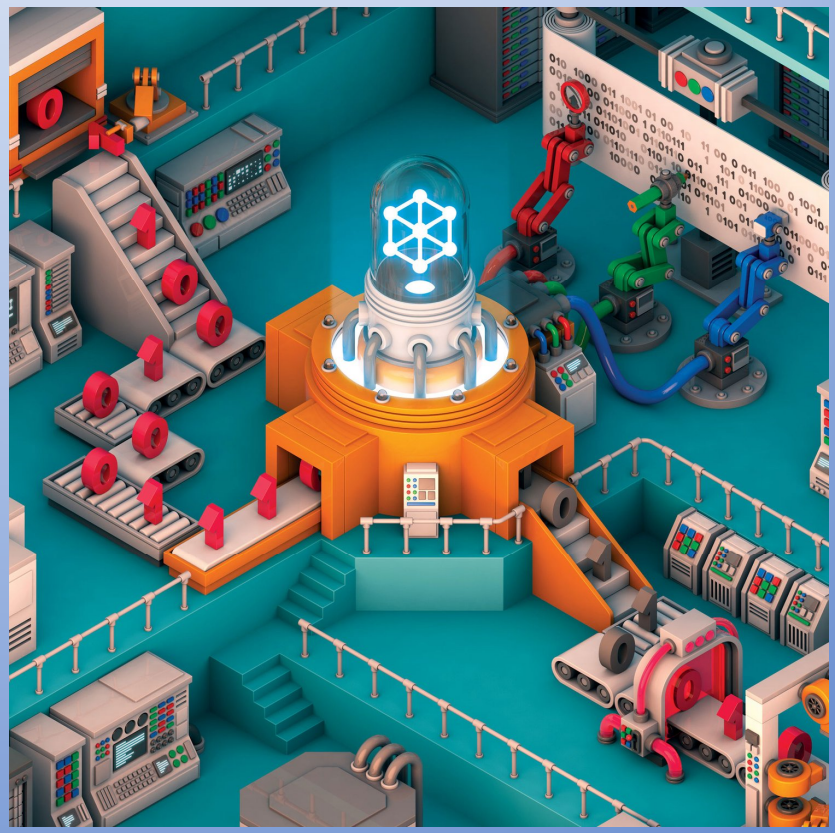
Outline

What is DNC ?

Concept

Experiment

Summary

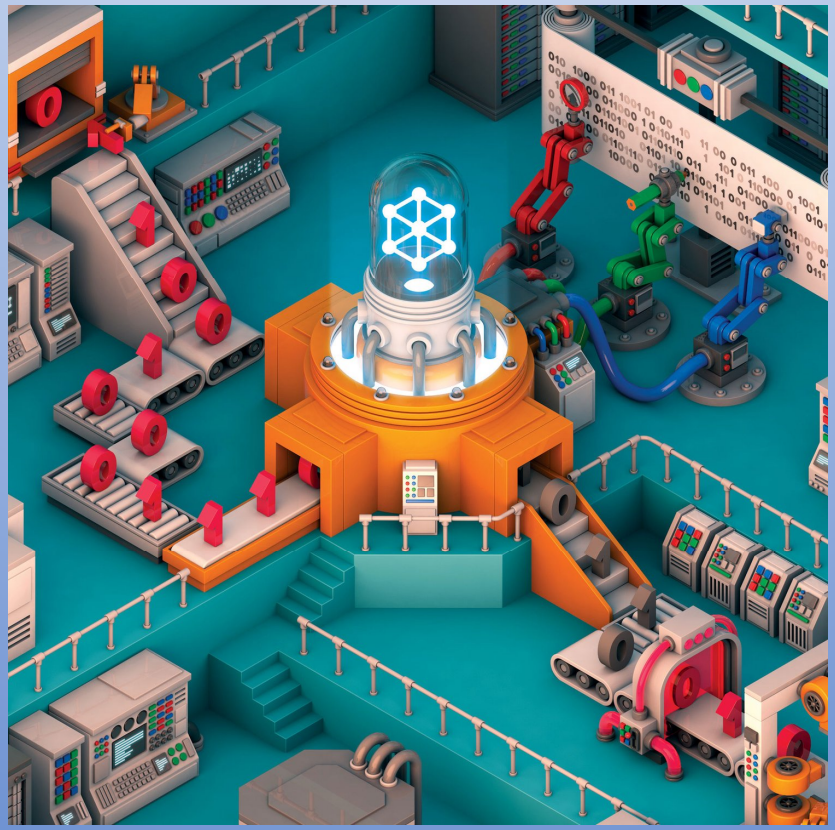


[1]

[1] https://deepmind.com/blog/differentiable-neural-computers/

What is DNC

Differentiable **N**eural **C**omputer

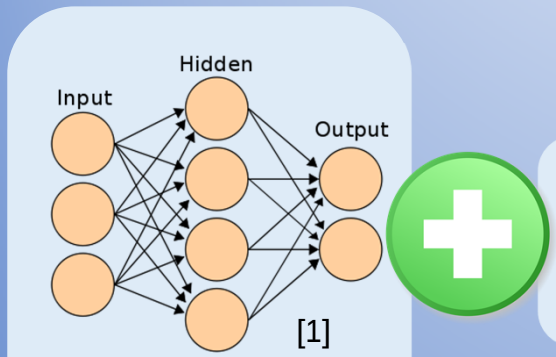


[1]

[1] https://deepmind.com/blog/differentiable-neural-computers/

What is DNC

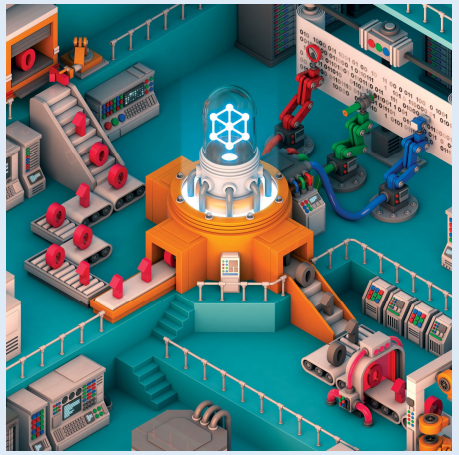
Differentiable **N**eural **C**omputer



NN

external memory

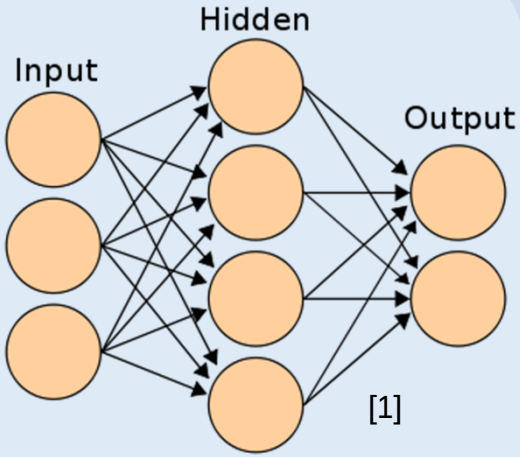
differentiable attention mechanisms



DNC [2]

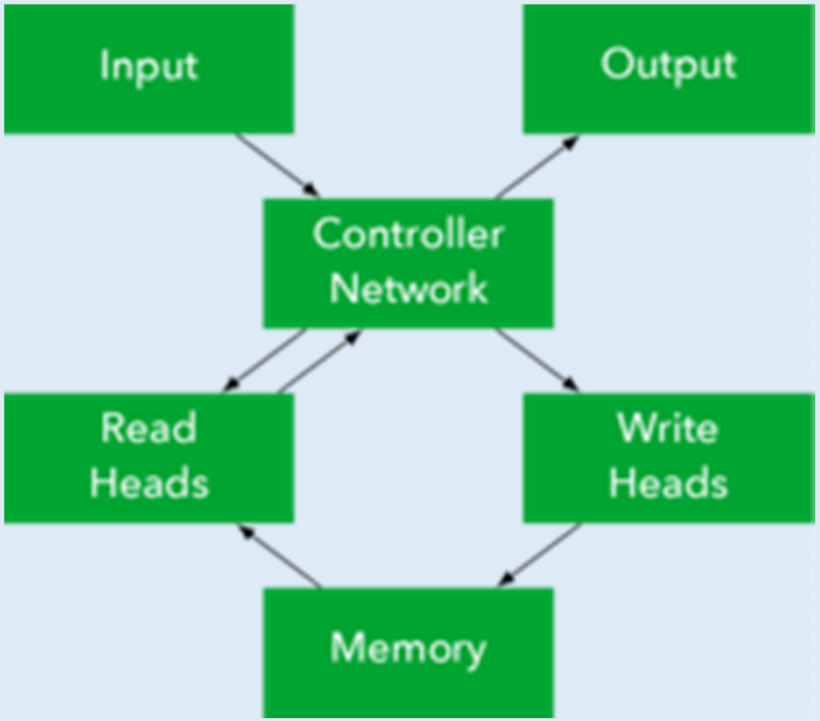
[1] https://www.analyticsvidhya.com/blog/2014/10/ann-work-simplified/
[2] https://deepmind.com/blog/differentiable-neural-computers/

Differentiable Neural Computer



[1]

NN

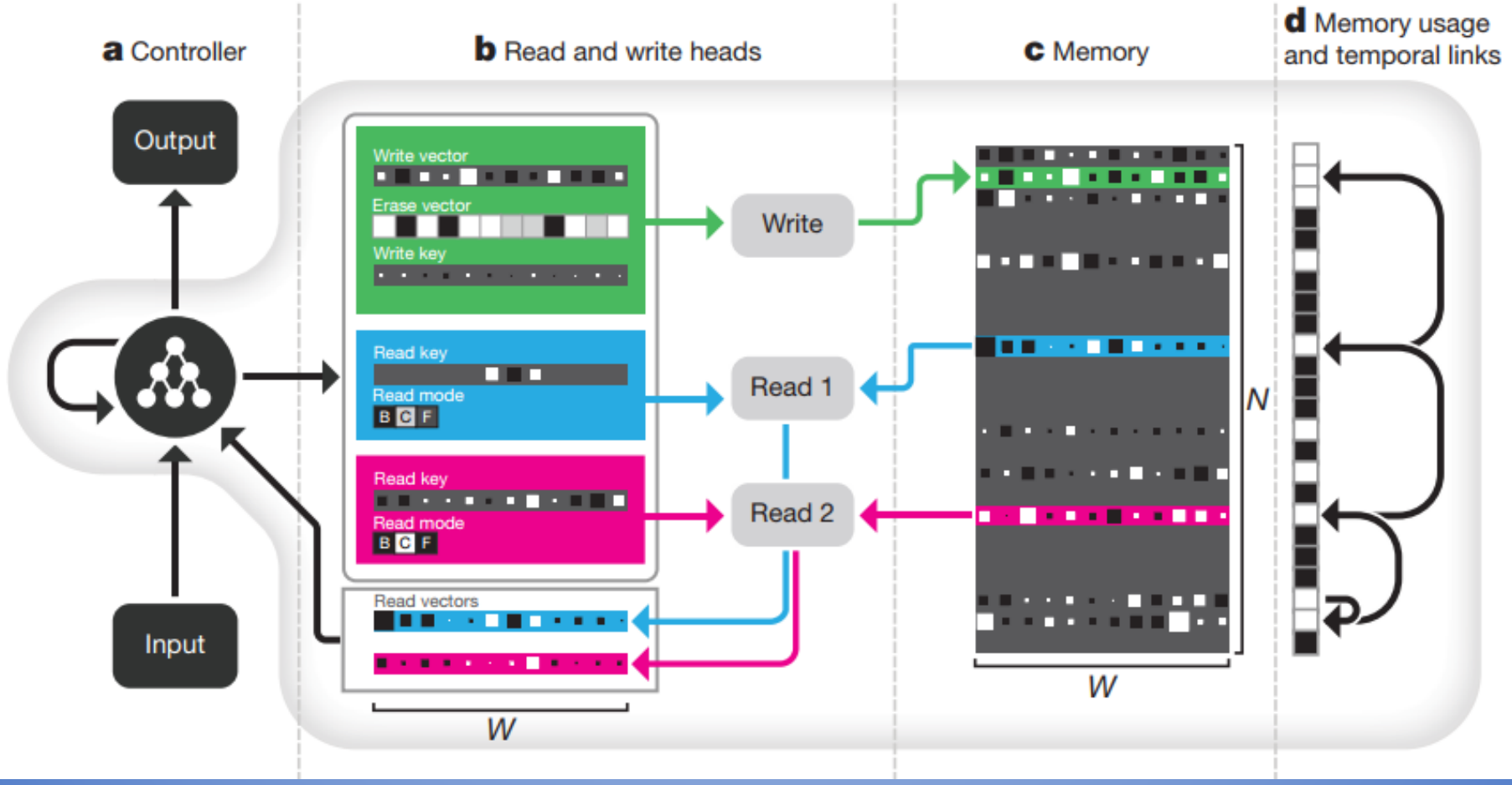


[2]

DNC

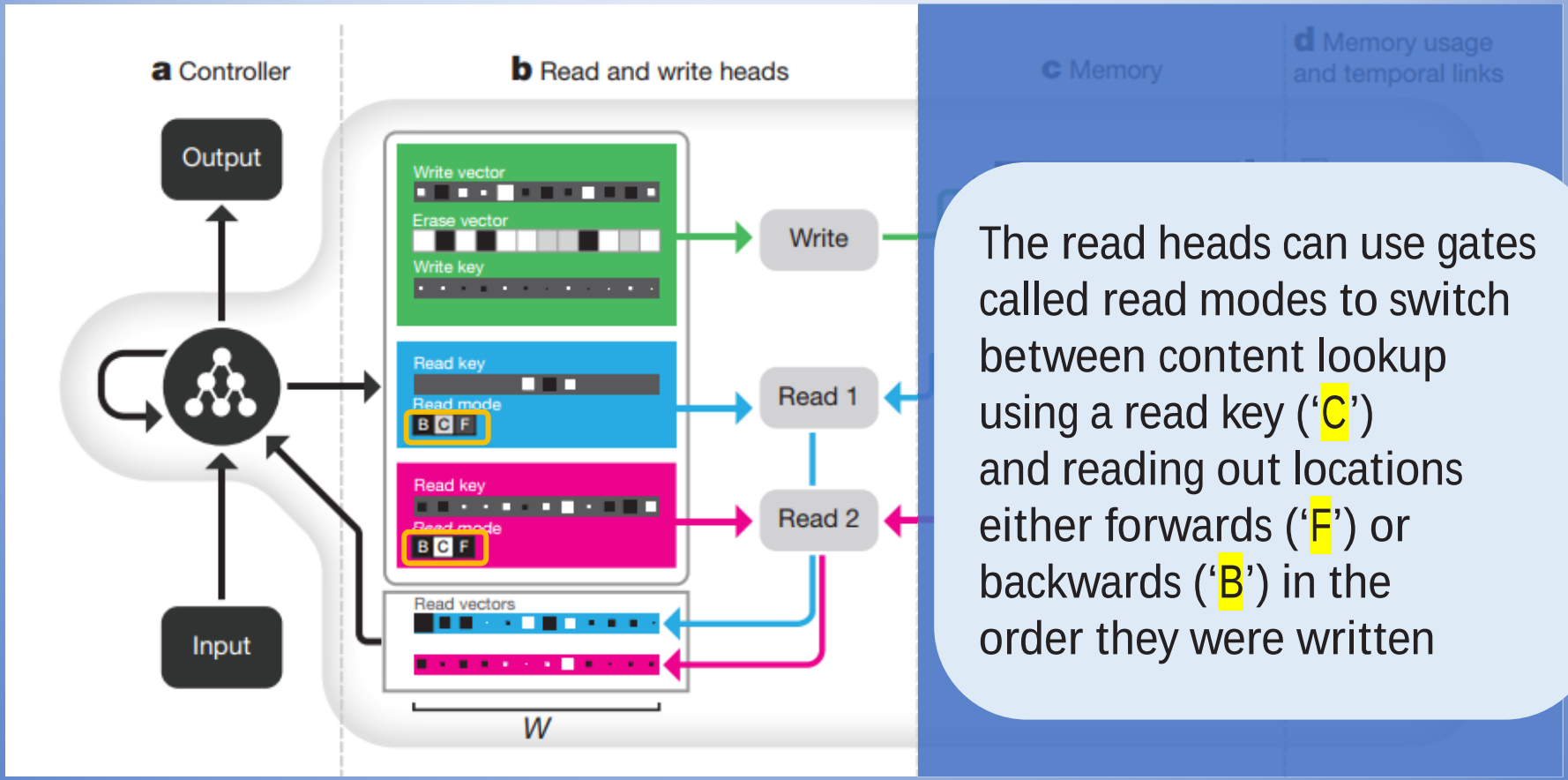
[1] https://www.analyticsvidhya.com/blog/2014/10/ann-work-simplified/
[2] https://en.wikipedia.org/wiki/Differentiable_neural_computer#cite_note-DNCnature2016-1

DNC architecture



Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

DNC architecture

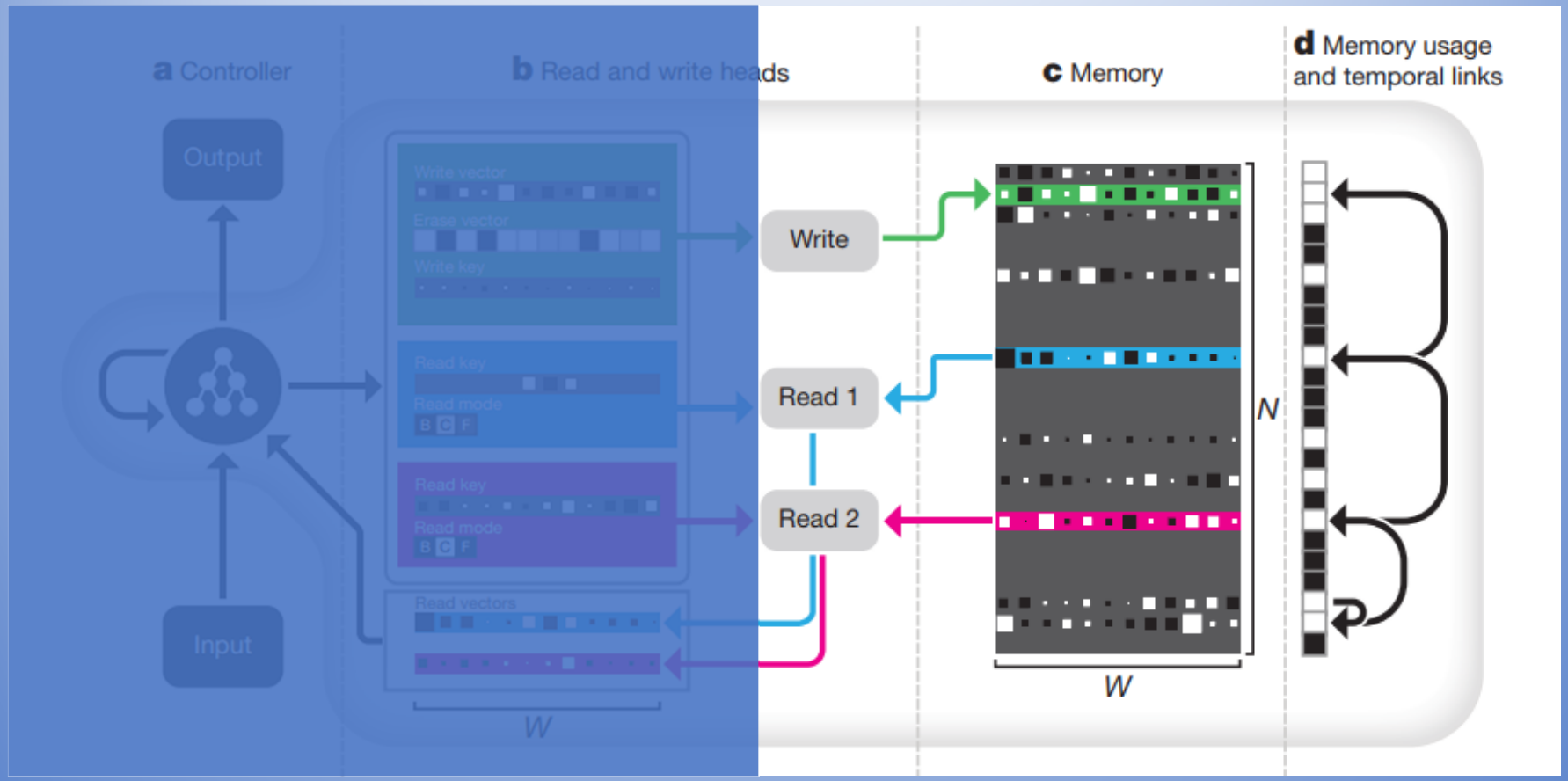


The read heads can use gates called read modes to switch between content lookup using a read key (‘C’) and reading out locations either forwards (‘F’) or backwards (‘B’) in the order they were written

The controller also outputs vectors that parameterize one write head and multiple read heads (two in this case, blue and pink)

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

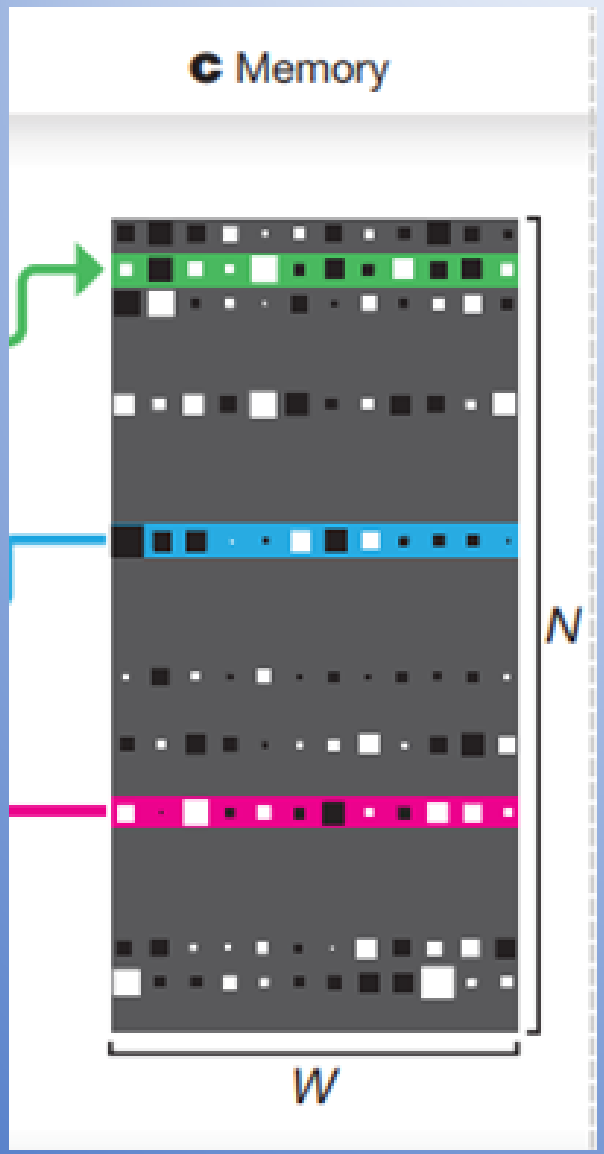
DNC architecture



The write head defines a write and an erase vector that are used to edit the $N \times W$ memory matrix

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

DNC architecture



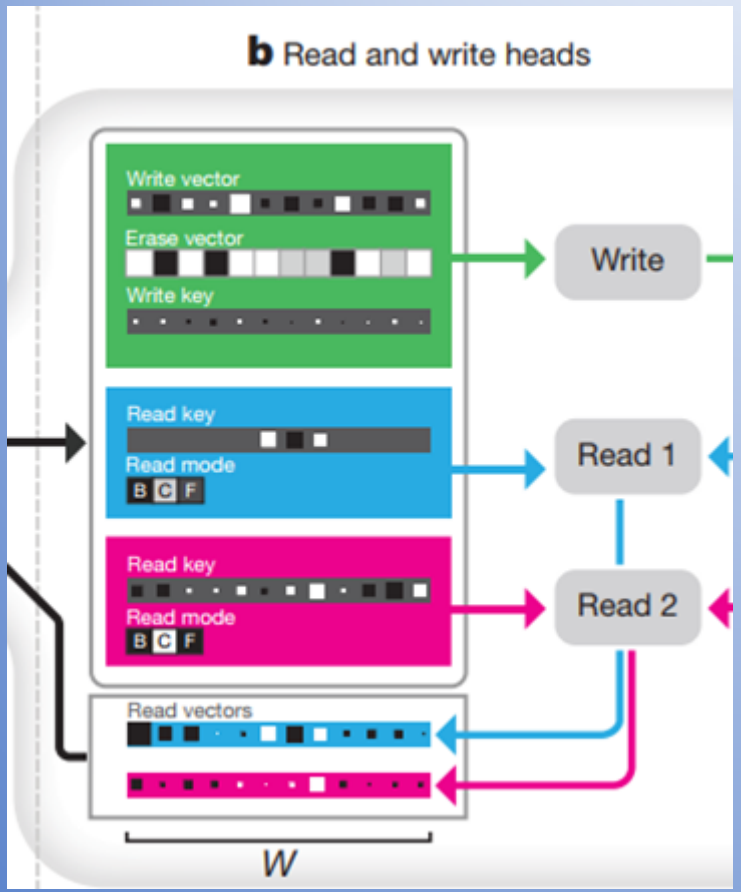
Whereas conventional computers use unique addresses
to access memory contents,

a DNC uses
differentiable attention mechanisms to define distributions over
the *N* rows, or ‘locations’,
in the *N*×*W* memory matrix *M*

These distributions, which we call weightings, represent the degree to which each location is involved in a read or write operation.

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

DNC architecture



The read vector ***r*** returned by a read weighting $\boldsymbol{w}^{\mathrm{r}}$ over memory *M*

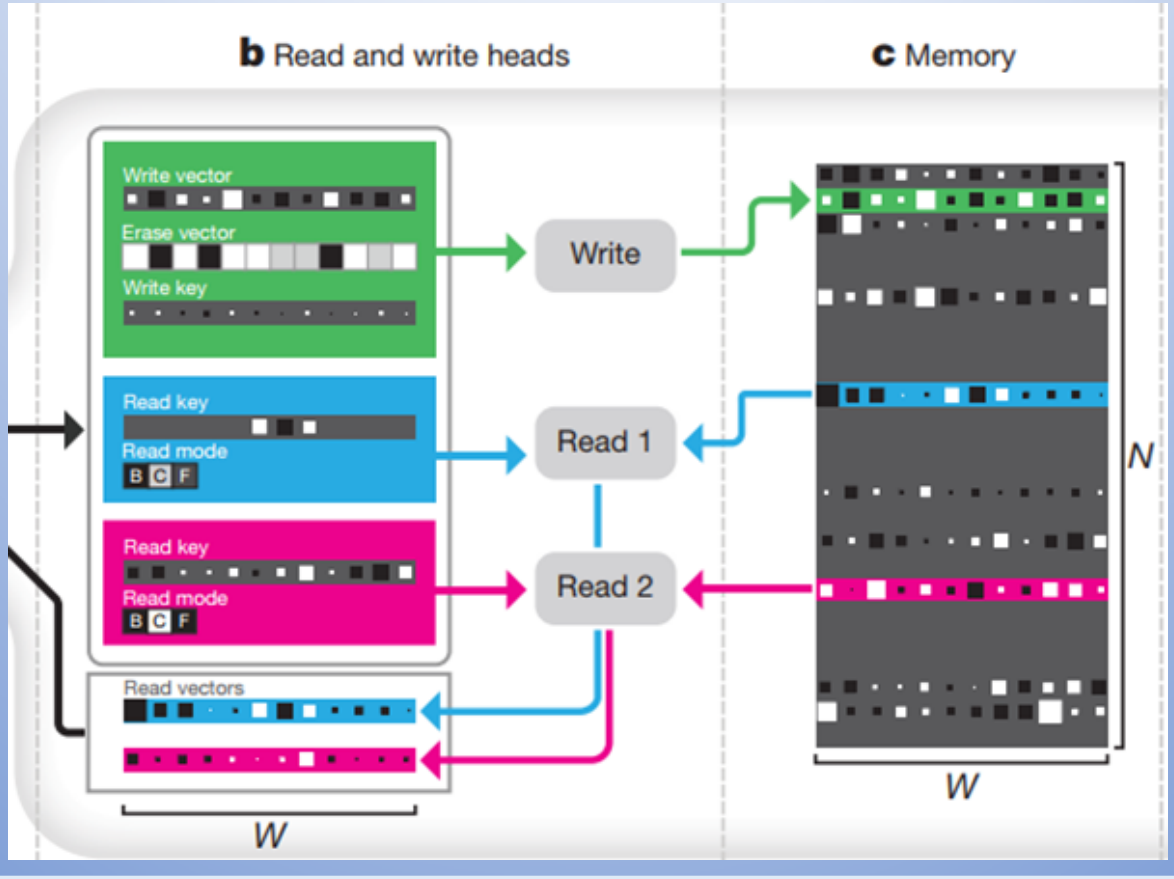
is a weighted sum over the memory locations:

$$\boldsymbol{r} = \sum_{i=1}^{N} M[i, \cdot]\boldsymbol{w}^{\mathrm{r}}[i]$$

where the '·' denotes all *j*=1, ..., *W*

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

DNC architecture

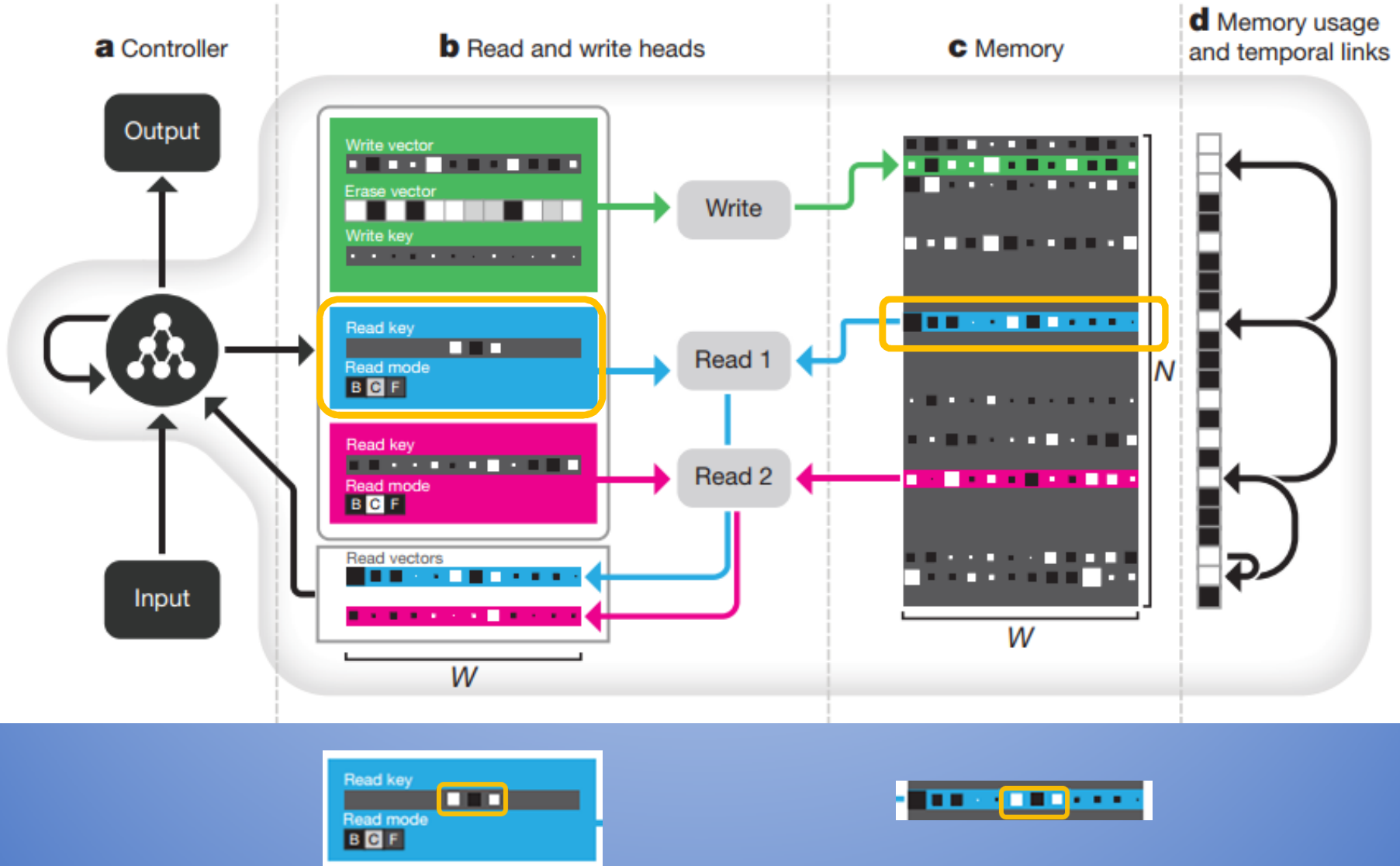


Similarly, the write operation uses a write weighting $\boldsymbol{w}^{\mathrm{w}}$ to first erase with an erase vector $\boldsymbol{e}$, then add a write vector

$$\boldsymbol{v}:\ M[i,j] \leftarrow M[i,j]\,(1 - \boldsymbol{w}^{\mathrm{w}}[i]\boldsymbol{e}[j]) + \boldsymbol{w}^{\mathrm{w}}[i]\boldsymbol{v}[j]$$

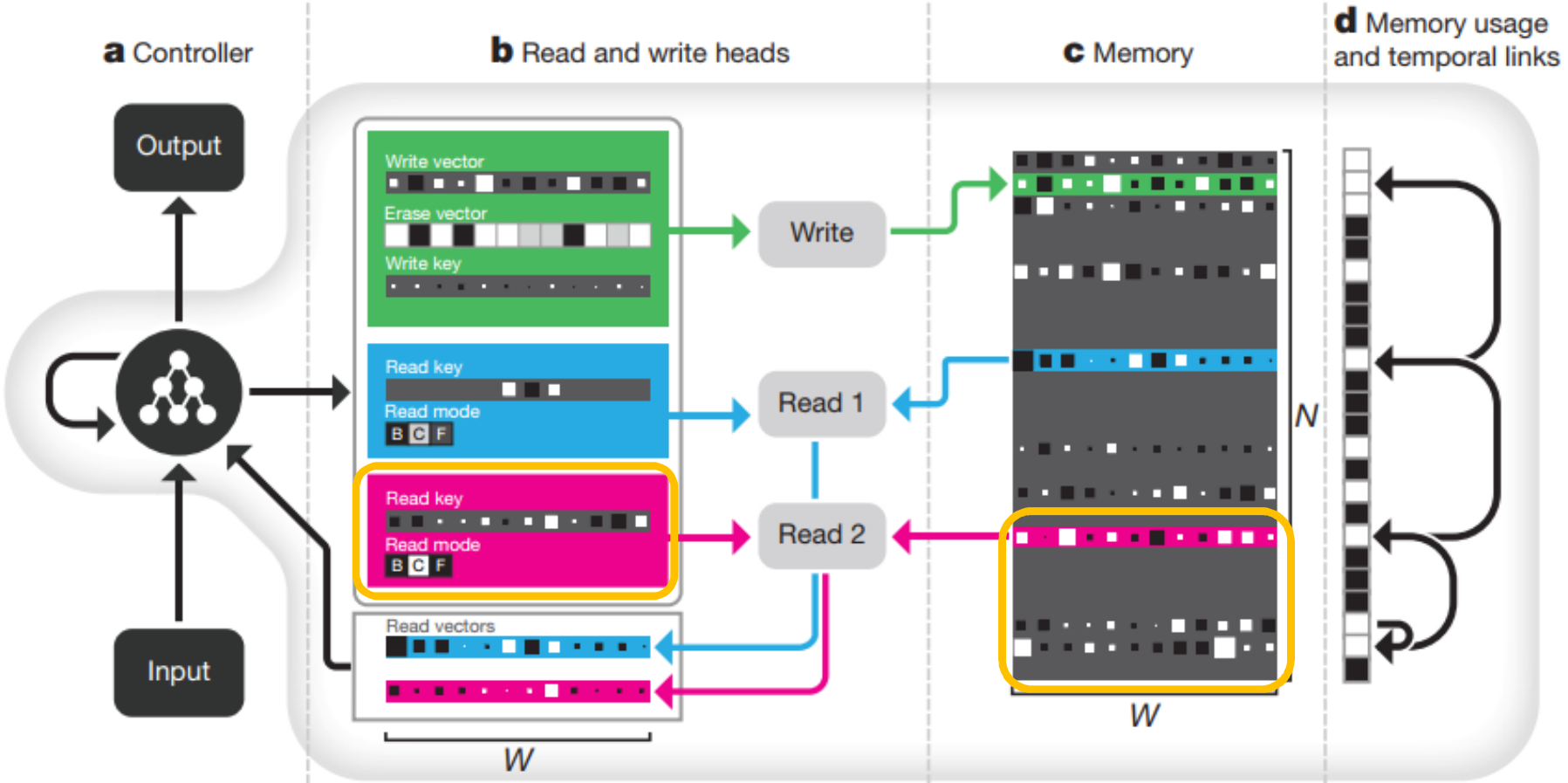
Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

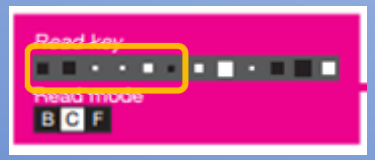
DNC architecture

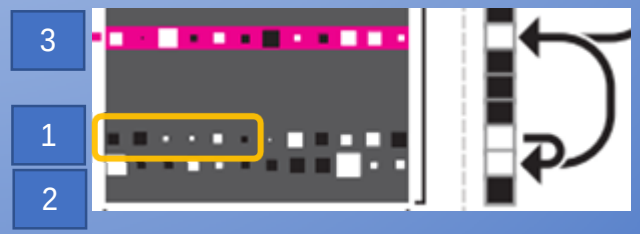


Hybri computing using a neural network with dynamic external memory by Alex Graves et al.

DNC architecture







Hybri computing using a neural network with dynamic external memory by Alex Graves et al.

Memory Addressing

The system uses a combination of content-based addressing and dynamic memory allocation to determine where to write in memory,

and a combination of content-based addressing and temporal memory linkage to determine where to read.

Content-based addressing. All content lookup operations on the memory $M \in \mathbb{R}^{N \times W}$ use the following function :

$$\mathcal{C}(M, \boldsymbol{k}, \beta)[i] = \frac{\exp\{\mathcal{D}(\boldsymbol{k}, M[i, \cdot\,])\beta\}}{\sum_j \exp\{\mathcal{D}(\boldsymbol{k}, M[j, \cdot\,])\beta\}}$$

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Memory Addressing

Content-based addressing:

$$\mathcal{C}(M, \boldsymbol{k}, \beta)[i] = \frac{\exp\{\mathcal{D}(\boldsymbol{k}, M[i, \cdot\,])\beta\}}{\sum_j \exp\{\mathcal{D}(\boldsymbol{k}, M[j, \cdot\,])\beta\}}$$

$\boldsymbol{k} \in \mathbb{R}^W$ is a lookup key,

$\beta \in [1, \infty)$ is a scalar representing key strength

$\mathcal{D}$ is the cosine similarity:

$$\mathcal{D}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}$$

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Differentiable Attention Mechanisms<1>

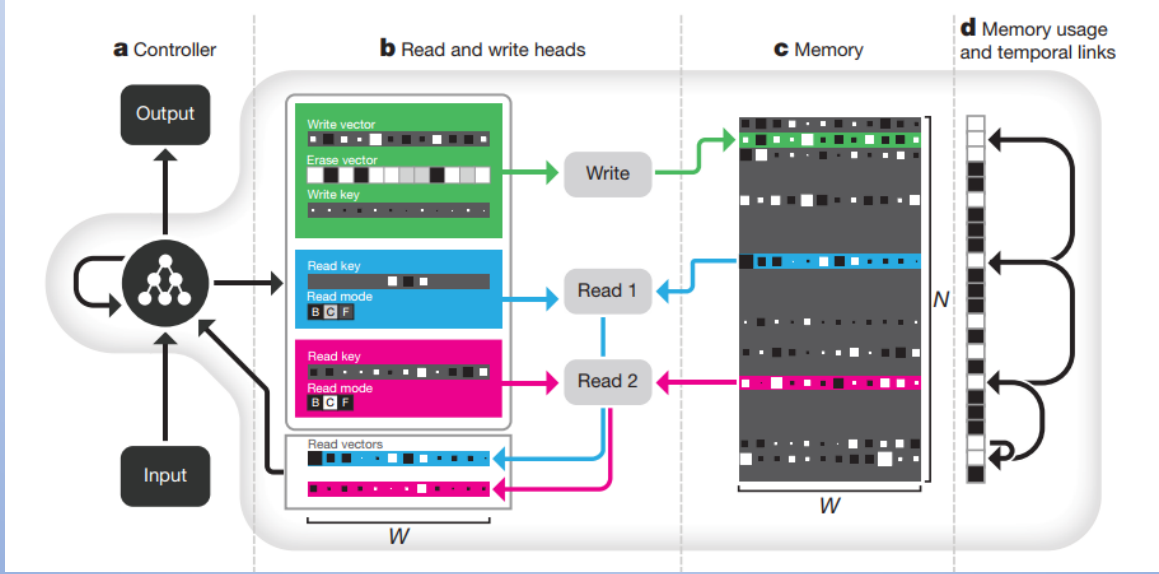
The first is content lookup in which a key vector emitted by the controller is compared to the content of each location in memory according to a similarity measure (here, cosine similarity).

The similarity scores determine a weighting that can be used by the read heads for associative recall or by the write head to modify an existing vector in memory.

Importantly, a key that only partially matches the content of a memory location can still be used to attend strongly to that location

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Differentiable Attention Mechanisms<2>



A second attention mechanism records transitions between consecutively written locations in an $N \times N$ temporal link matrix L. $L[i, j]$ is close to 1 if i was the next location written after j, and is close to 0 otherwise. For any weighting $\boldsymbol{w}$, the operation $L\boldsymbol{w}$ smoothly shifts the focus forwards to the locations written after those emphasized in $\boldsymbol{w}$, whereas $L^{\top}\boldsymbol{w}$ shifts the focus backwards. This gives a DNC the native ability to recover sequences in the order in which it wrote them, even when consecutive writes did not occur in adjacent time-steps.

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Differentiable Attention Mechanisms<3>

The third form of attention allocates memory for writing.

The ‘usage’ of each location is represented as a number between 0 and 1,

and a weighting that picks out unused locations is delivered to the write head.

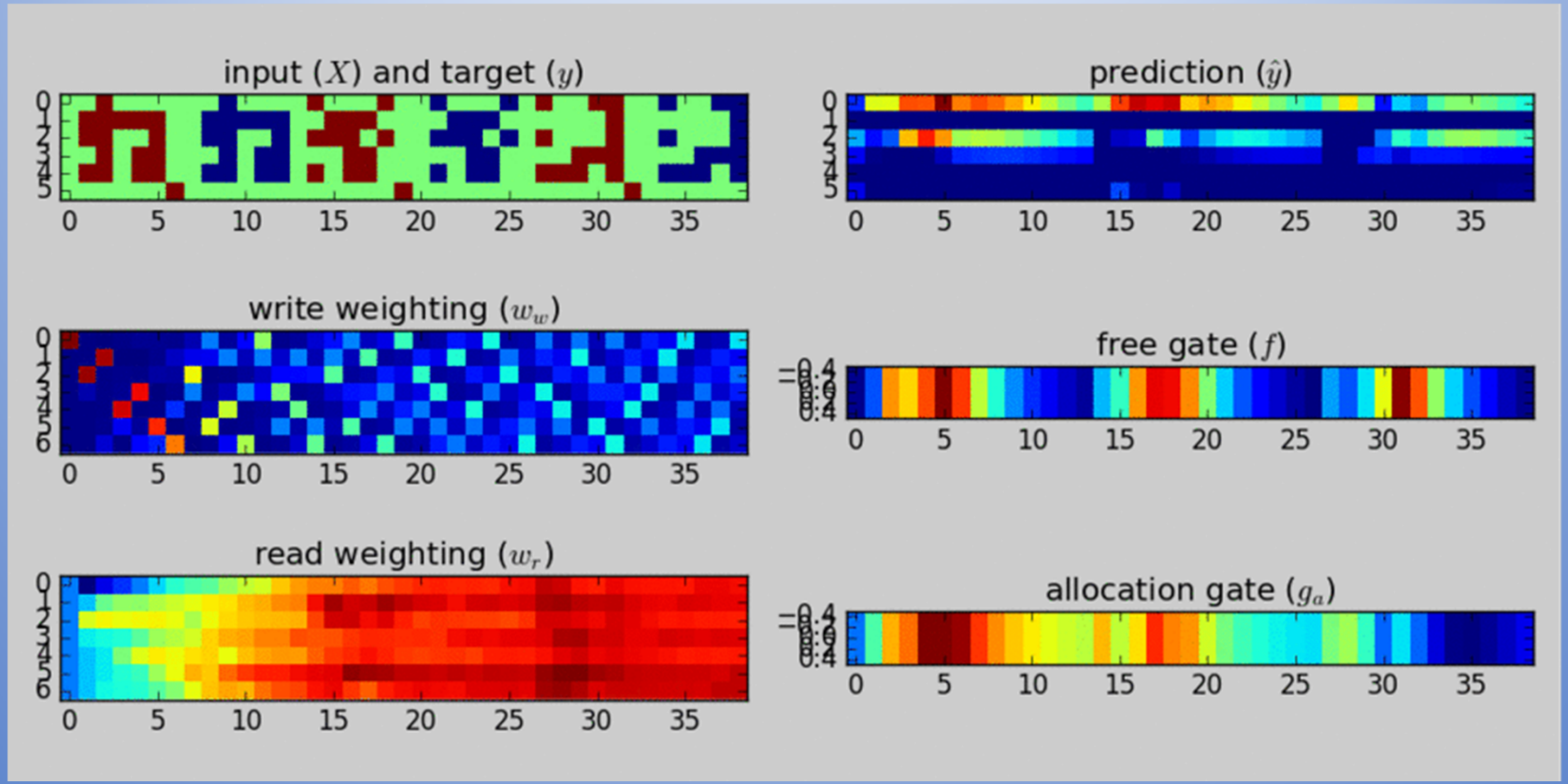
As well as automatically increasing with each write to a location, usage can be decreased after each read.

This allows the controller to reallocate memory that is no longer required The allocation mechanism is independent of the size and contents of the memory,

meaning that DNCs can be trained to solve a task using one size of memory and later upgraded to a larger memory without retraining

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Experiment



[1]

[1] https://en.wikipedia.org/wiki/Differentiable_neural_computer

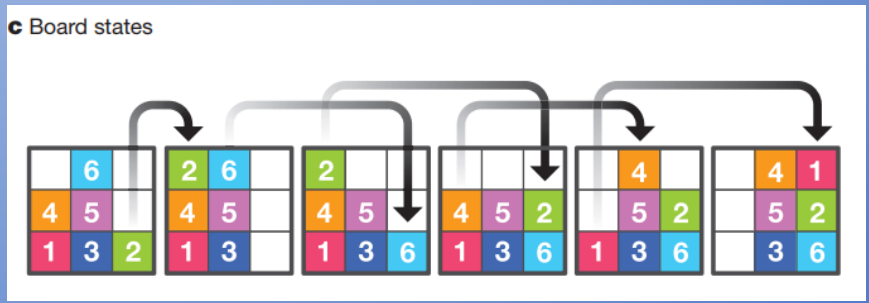
Experiment

1. Question answering :
   bAbI

```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?        bathroom        1
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?      hallway 4
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?      hallway 4
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?     office  11
13 John went back to the garden.
14 John moved to the bedroom.
15 Where is Sandra?     bathroom        8
```

2. Graph :
   London Underground
   The family tree

3. Block puzzle :
   Mini-SHRDLU



Question answering experiments

bAbI dataset : example

```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?        bathroom        1
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?      hallway 4
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?      hallway 4
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?     office  11
13 John went back to the garden.
14 John moved to the bedroom.
15 Where is Sandra?     bathroom        8
```

[1]

To compare DNCs to other NN architectures,

considered the bAbI dataset,

generated questions designed to mimic aspects of textual reasoning.

[1] https://jamesmccaffrey.wordpress.com/2018/07/06/the-babi-dataset/

Question answering experiments

We found that a single DNC was able to achieve a mean test error rate of 3.8%

compared to 7.5% mean error for the best previous jointly trained result.

DNCs performed much better than both LSTM and the neural Turing machine

Unlike previous results on this dataset, the inputs to our model were single word tokens without any preprocessing or sentence-level features

Graph experiments : London Underground

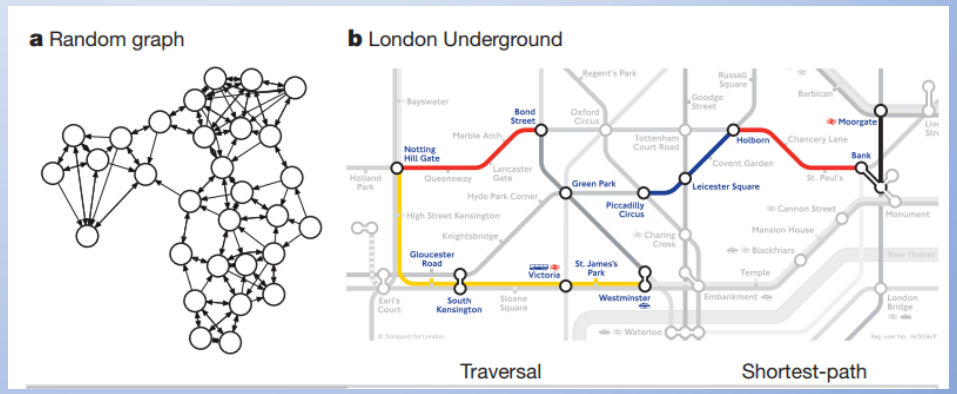
a Random graph

b London Underground

	Traversal	Shortest-path
Underground input: (OxfordCircus, TottenhamCtRd, Central) (TottenhamCtRd, OxfordCircus, Central) (BakerSt, Marylebone, Circle) (BakerSt, Marylebone, Bakerloo) (BakerSt, OxfordCircus, Bakerloo) :	Traversal question: (BondSt, _, Central), (_, _, Circle), (_, _, Circle), (_, _, Circle), (_, _, Circle), (_, _, Jubilee), (_, _, Jubilee),	Shortest-path question: (Moorgate, PiccadillyCircus, _)

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Graph experiments :
London Underground



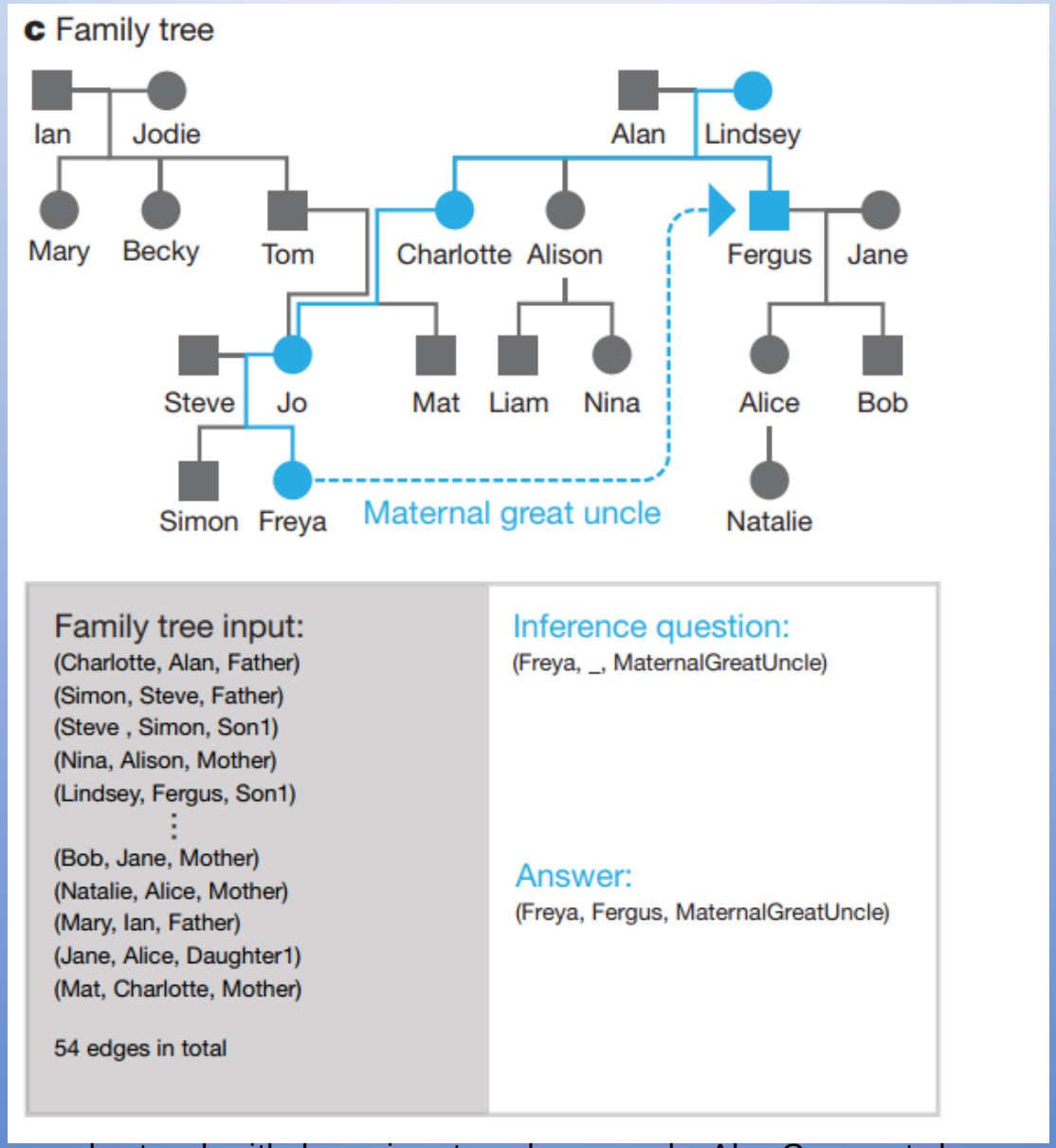
As a benchmark we again compared DNCs with LSTM.

the best LSTM network we found reaching an average of only 37% accuracy after almost two million training examples;

DNCs reached an average of 98.8% accuracy of the same curriculum after around one million training examples.

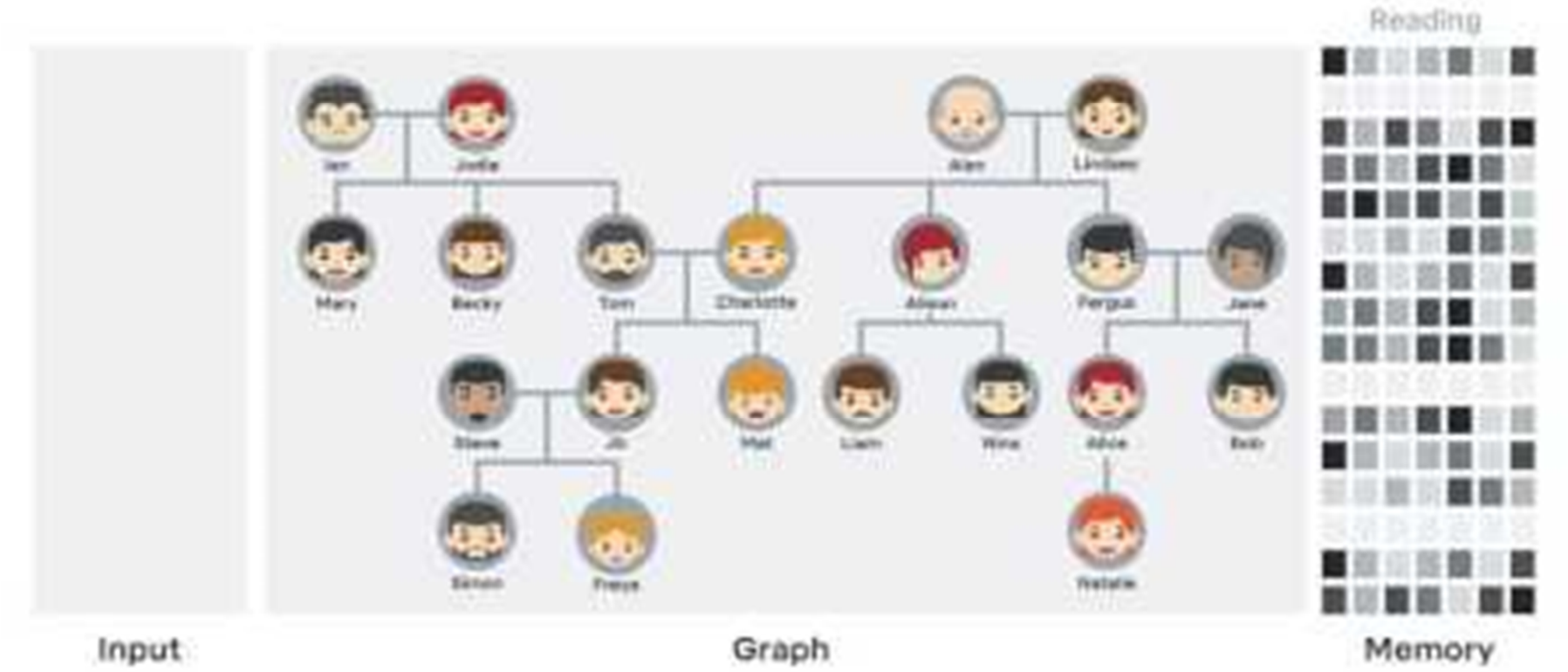
Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Graph experiments : The family tree



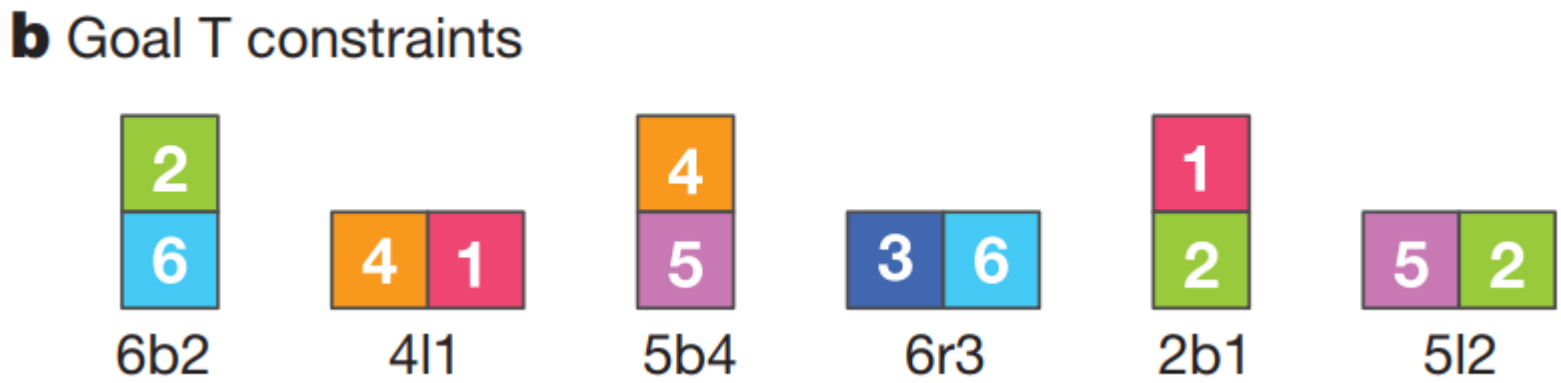
Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Graph experiments : The family tree



https://www.youtube.com/watch?v=B9U8sI7TcMY

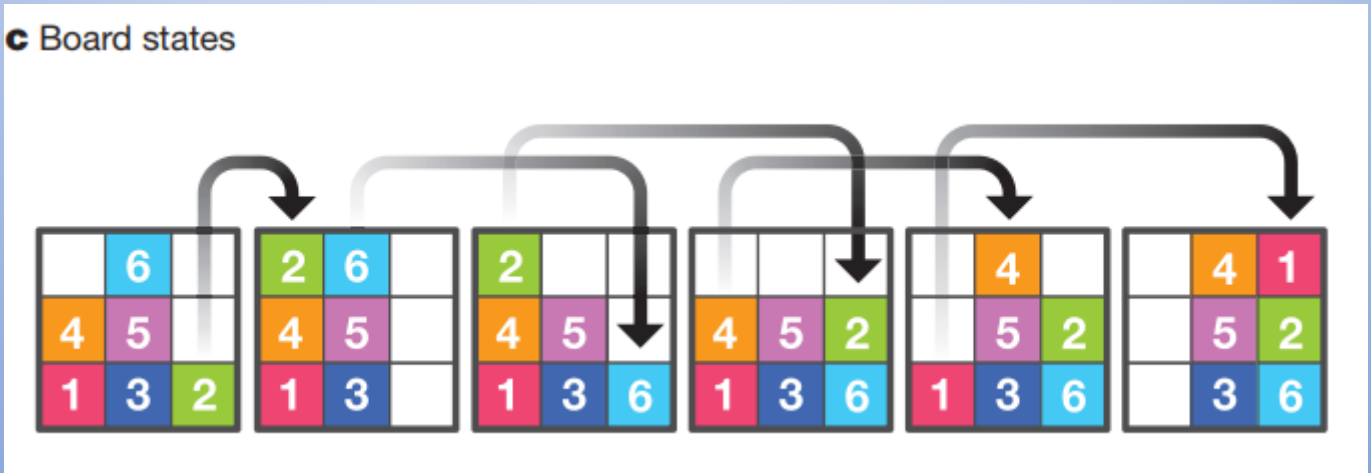
Block puzzle experiments : Mini-SHRDLU



‘b’, ‘a’, ‘l’ and ‘r’ denote
‘below’, ‘above’, ‘left of ’ and ‘right of ’, respectively

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Block puzzle experiments : Mini-SHRDLU



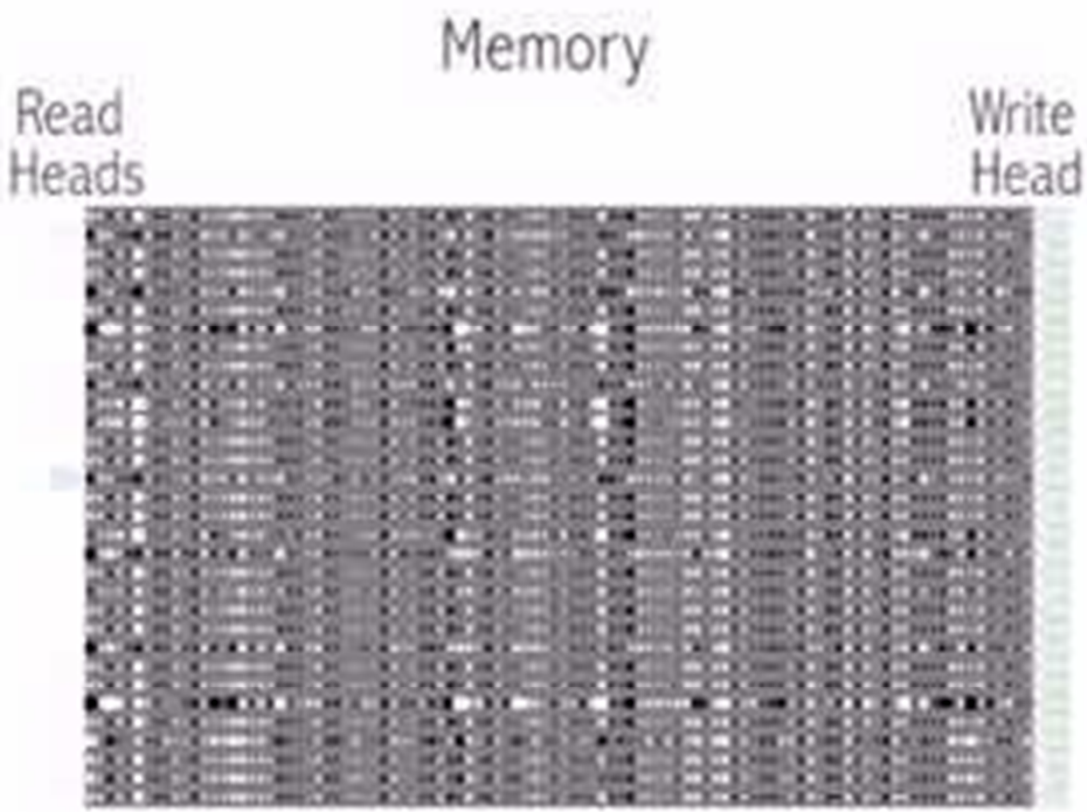
‘b’, ‘a’, ‘l’ and ‘r’ denote
‘below’, ‘above’, ‘left of ’ and ‘right of ’, respectively

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Block puzzle experiments : Mini-SHRDLU

Q?

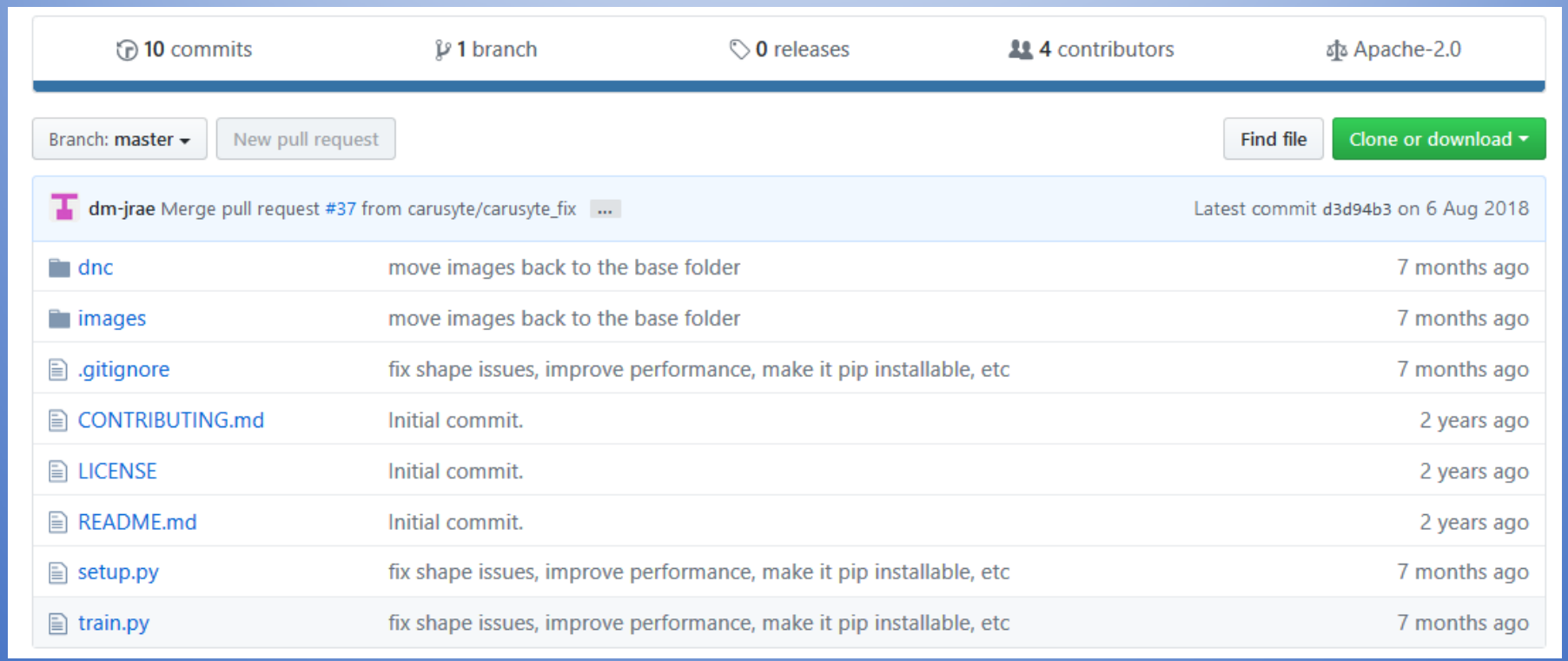
(Q: 6b4 2b6 1b5 3b1)



https://www.youtube.com/watch?v=BTSJ0-saaUc

Code and Implementation from author

https://github.com/deepmind/dnc



10 commits | 1 branch | 0 releases | 4 contributors | Apache-2.0

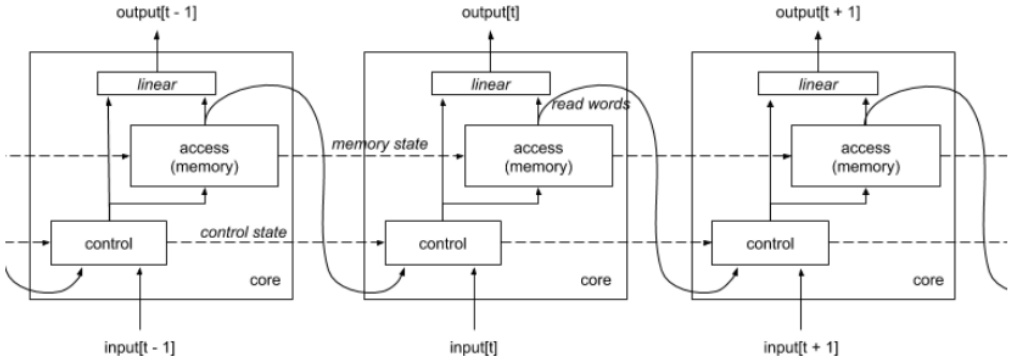
Branch: master | New pull request | Find file | Clone or download

dm-jrae Merge pull request #37 from carusyte/carusyte_fix ... Latest commit d3d94b3 on 6 Aug 2018

dnc	move images back to the base folder	7 months ago
images	move images back to the base folder	7 months ago
.gitignore	fix shape issues, improve performance, make it pip installable, etc	7 months ago
CONTRIBUTING.md	Initial commit.	2 years ago
LICENSE	Initial commit.	2 years ago
README.md	Initial commit.	2 years ago
setup.py	fix shape issues, improve performance, make it pip installable, etc	7 months ago
train.py	fix shape issues, improve performance, make it pip installable, etc	7 months ago

Code and Implementation from author

https://github.com/deepmind/dnc



Train

The `DNC` requires an installation of TensorFlow and Sonnet. An example training script is provided for the algorithmic task of repeatedly copying a given input string. This can be executed from a python interpreter:

```
$ ipython train.py
```

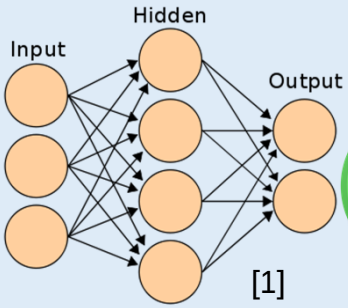
You can specify training options, including parameters to the model and optimizer, via flags:

```
$ python train.py --memory_size=64 --num_bits=8 --max_length=3

# Or with ipython:
$ ipython train.py -- --memory_size=64 --num_bits=8 --max_length=3
```

What is NTM

Neural Turing Machine

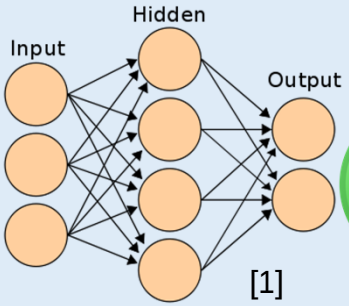


[1]

NN

external memory

NTM

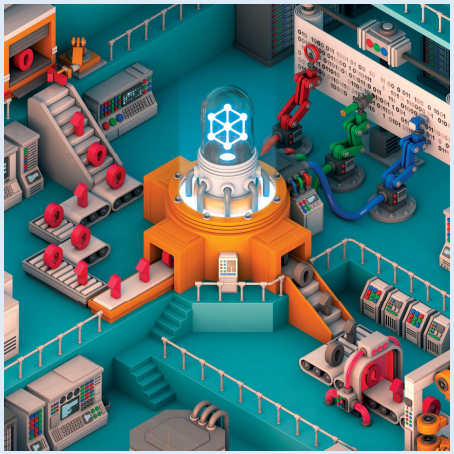


[1]

NN

external memory

differentiable attention mechanisms



[2]

DNC

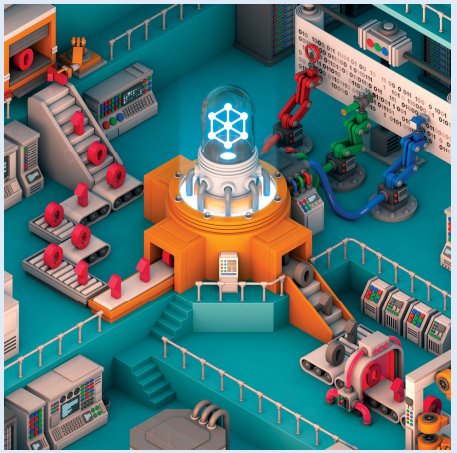
[1] https://www.analyticsvidhya.com/blog/2014/10/ann-work-simplified/
[2] https://deepmind.com/blog/differentiable-neural-computers/

NTM and DNC

NTM

2014

differentiable attention mechanisms



DNC [1]

2016

from Alex Graves et al.

DeepMind

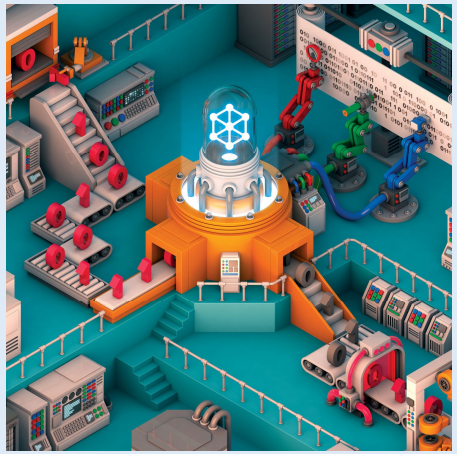
[1] https://deepmind.com/blog/differentiable-neural-computers/

DNC VS. NTM

NTM

external memory

store and retrieve temporal sequences in contiguous blocks of memory



DNC [1]

external memory

differentiable attention mechanisms

[1] https://deepmind.com/blog/differentiable-neural-computers/

DNC VS. NTM

1.the NTM has no mechanism to ensure that blocks of allocated memory do not overlap and interfere

2.the NTM has no way of freeing locations that have already been written to and, hence, no way of reusing memory when processing long sequences.

3. sequential information is preserved only as long as the NTM continues to iterate through consecutive locations;

The temporal link matrix used by DNCs does not suffer from this problem because it tracks the order in which writes were made

Hybrid computing using a neural network with dynamic external memory by Alex Graves et al.

Thank you

Q & A