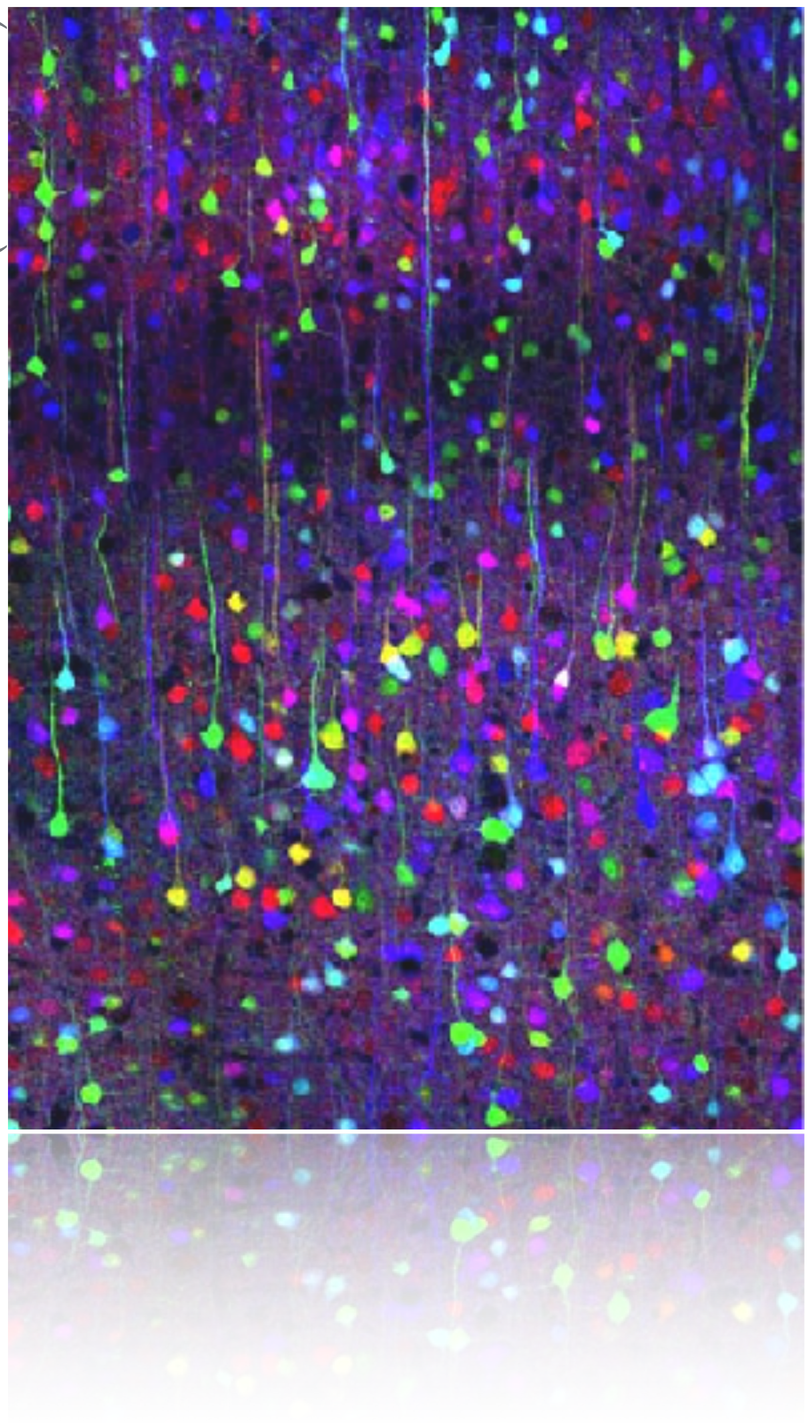
Neural Information Processing 2018/2019

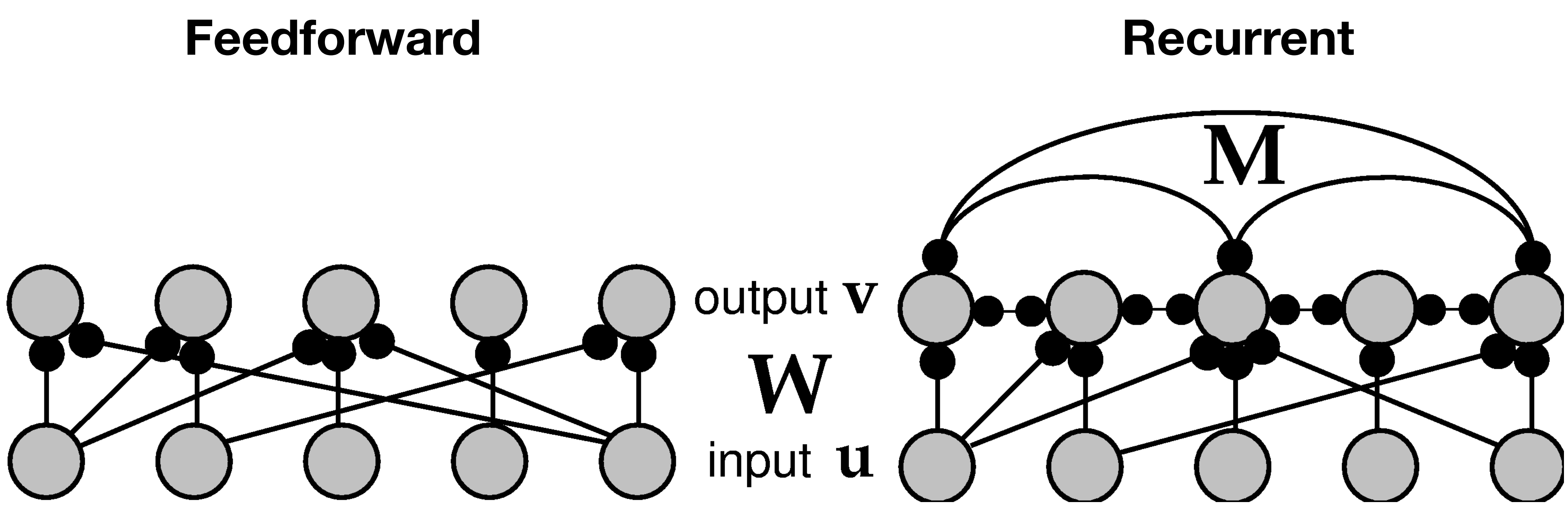
University of BRISTOL

Brainbow (Litchman Lab)



Lecture 16
Neural circuits and learning:
Microcircuits and
gated recurrent neural networks

Previously on *Neural Information Processing*...



Dayan and Abbott book (2001)

But, recurrent neural networks may benefit from additional structure…

How can you solve the delayed addition task with a RNN?

time →

number	5	**7**	3	2	4	7	**5**	9
mask	A	**B**	A	A	A	A	**B**	A

12

output
(input 7 + 5)

But, recurrent neural networks may benefit from additional structure…

Delayed addition task:

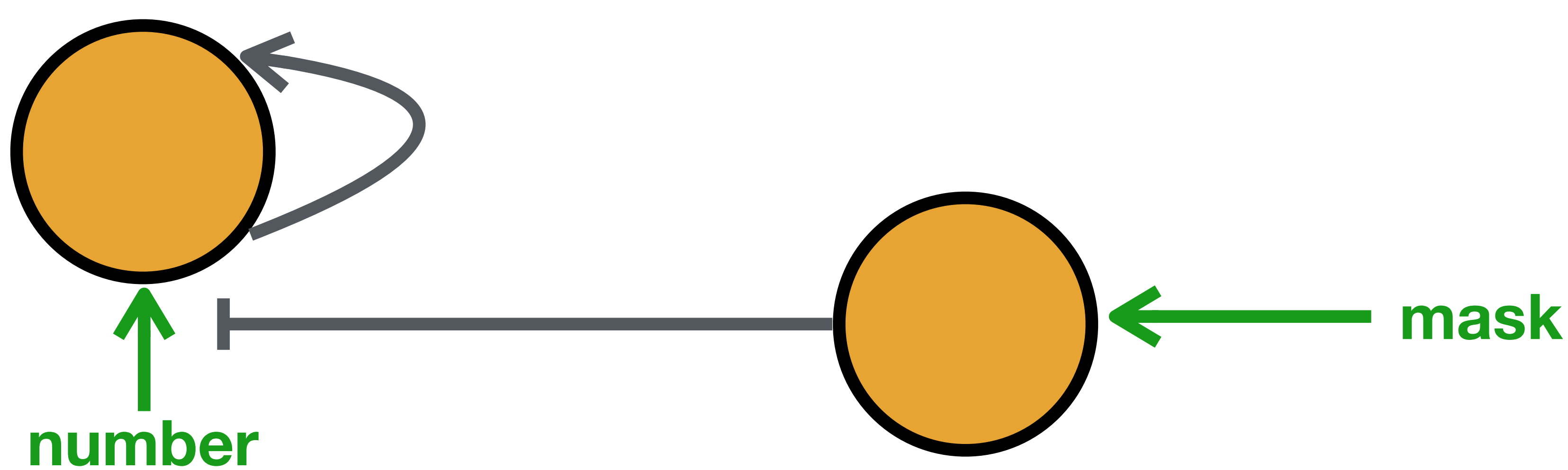
time →

number	5	**7**	3	2	4	7	**5**	9
mask	A	**B**	A	A	A	A	**B**	A

12

output
(input 7 + 5)

You would need an **integrator**: and **mask/gating** neuron:

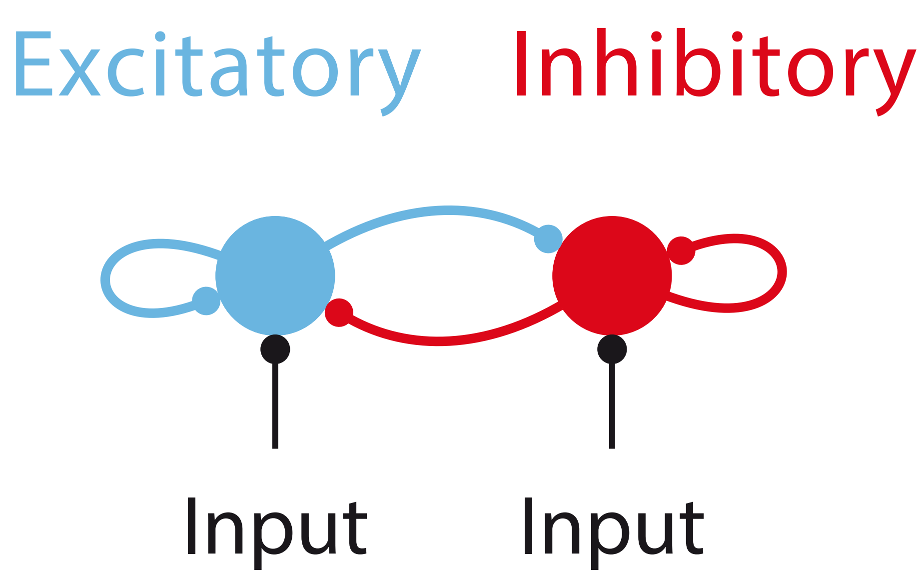


Outline

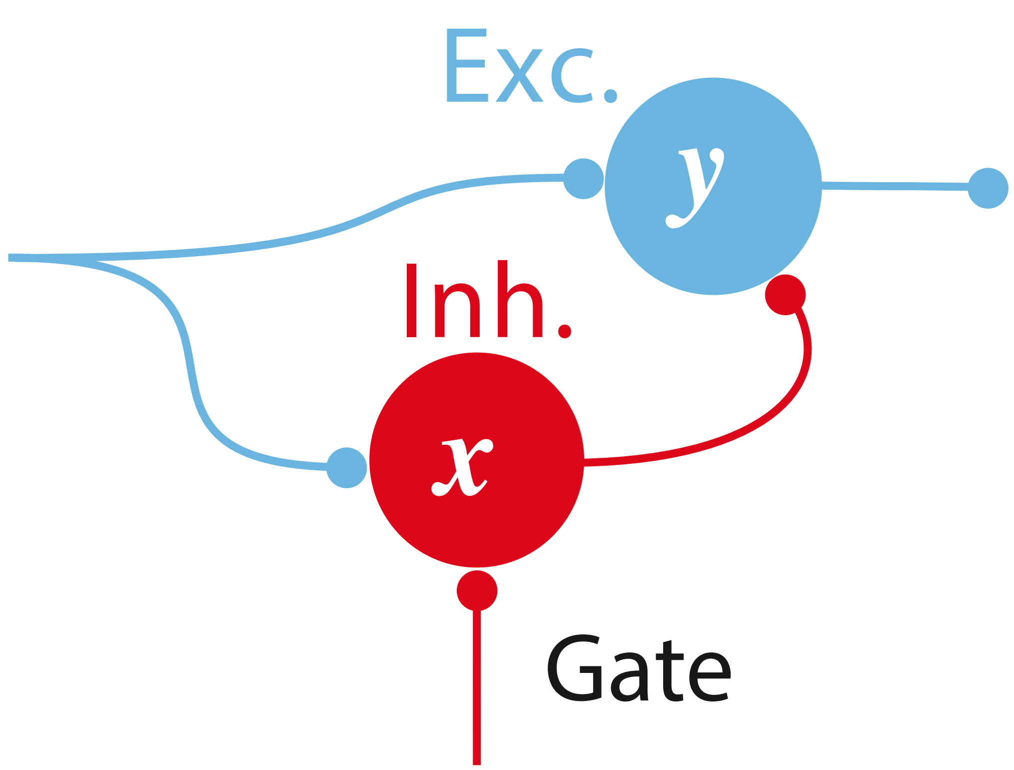
1. **Excitatory and inhibitory cell types, and their dynamics**
2. **Cortical excitatory and inhibitory microcircuits**
3. **Gated RNNs:** long short-term memory networks
4. **A biological plausible version:** Subtractive gated-RNNs

The excitatory and inhibitory dance

The brain contains two main types of neurons: **excitatory** (i.e. make synapses onto other neurons with positive synaptic weights) and **inhibitory** (i.e. make synapses onto other neurons with negative weights).



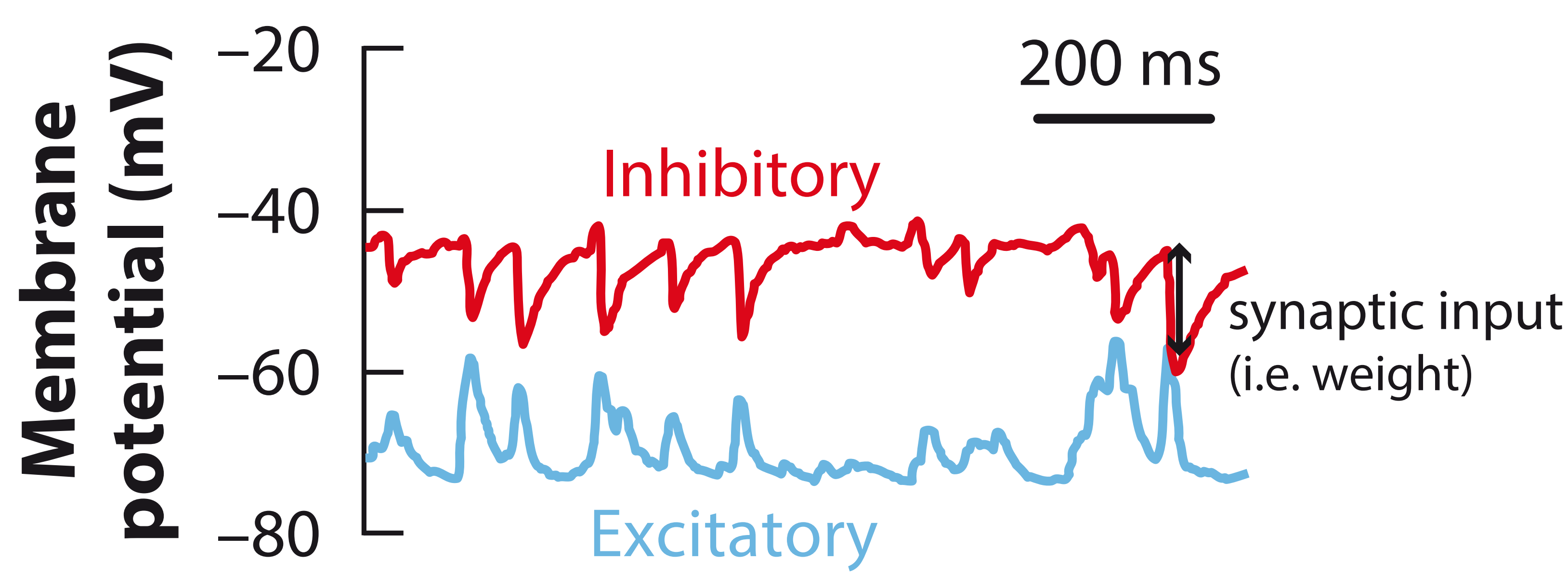
Inhibitory neurons act as gates:



Hennequin et al. review (2017)

The excitatory and inhibitory dance

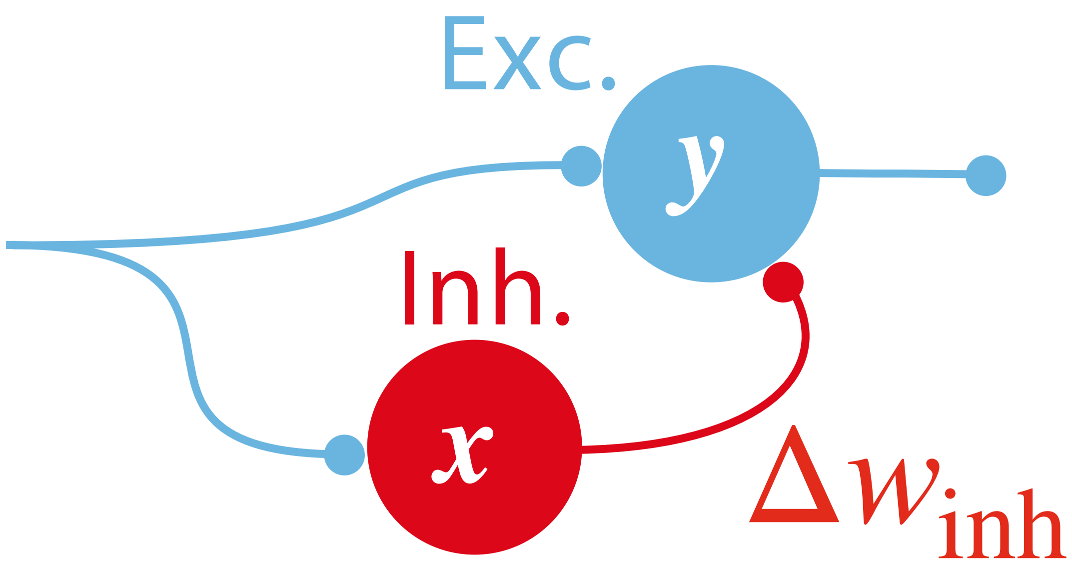
In vivo excitation-inhibition (detailed) balance:
(i.e. excitation and inhibition have similar weights)



Okun and Lampl Nat. Neurosci. (2008)

Hennequin et al. review (2017)

Learning to balance excitation and inhibition



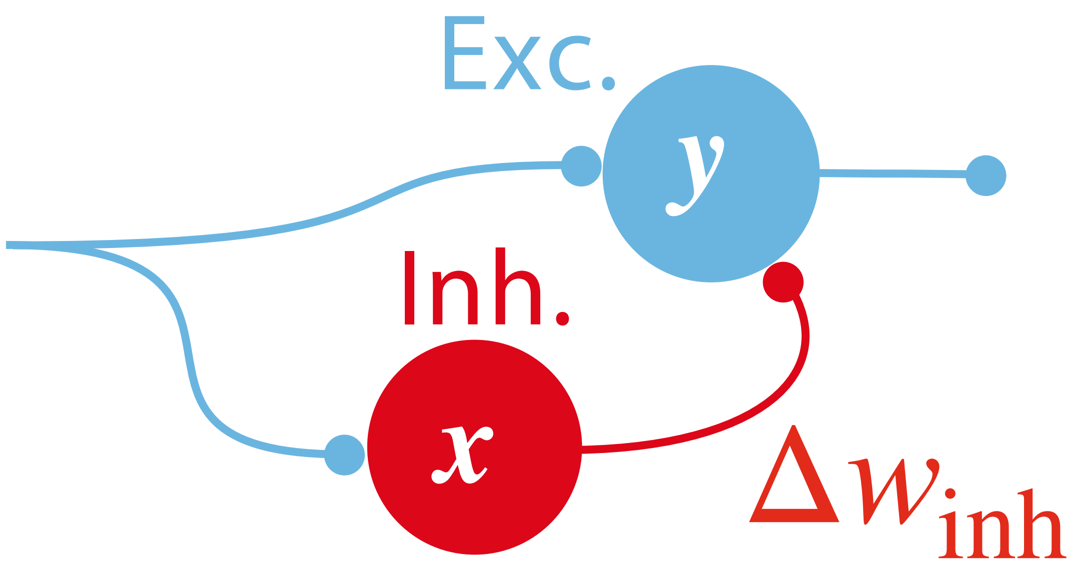
Inhibitory learning rule:

$$\Delta w_{\mathrm{inh}} = \eta x(y - r_0)$$

target rate

Vogels et al. Science (2011)

Learning to balance excitation and inhibition



Inhibitory learning rule:

$$\Delta w_{\mathrm{inh}} = \eta x(y - r_0)$$

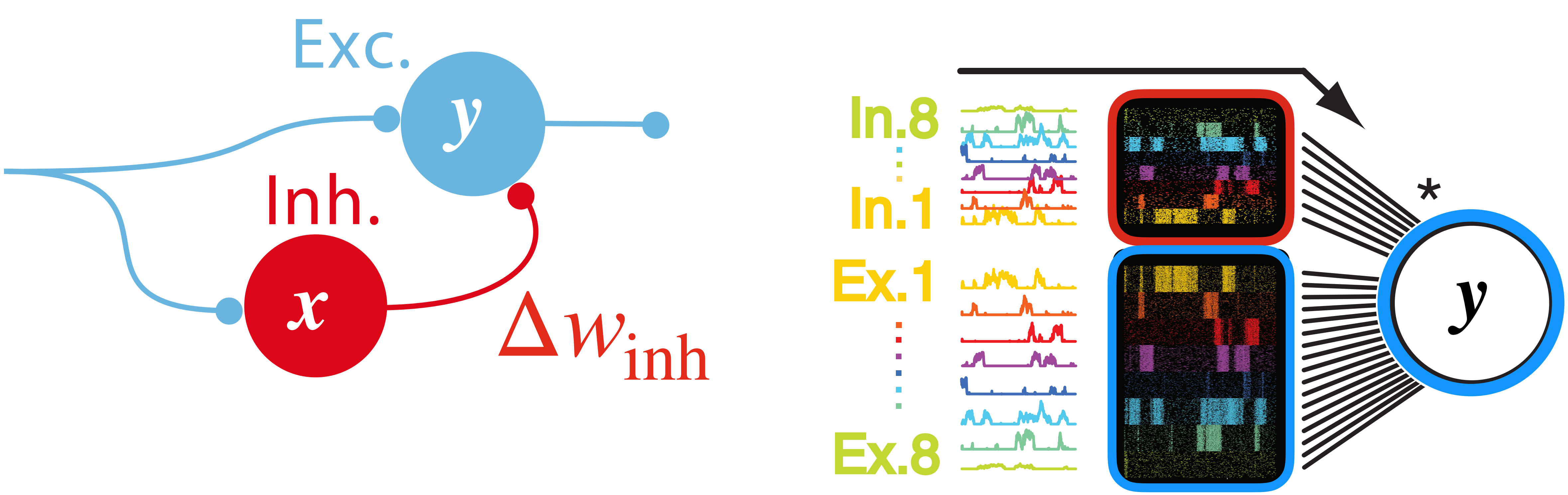
$$0 = \eta x(y - r_0)$$

$$y = r_0$$

postsynaptic neuron, *y* = target rate (*r_0*)

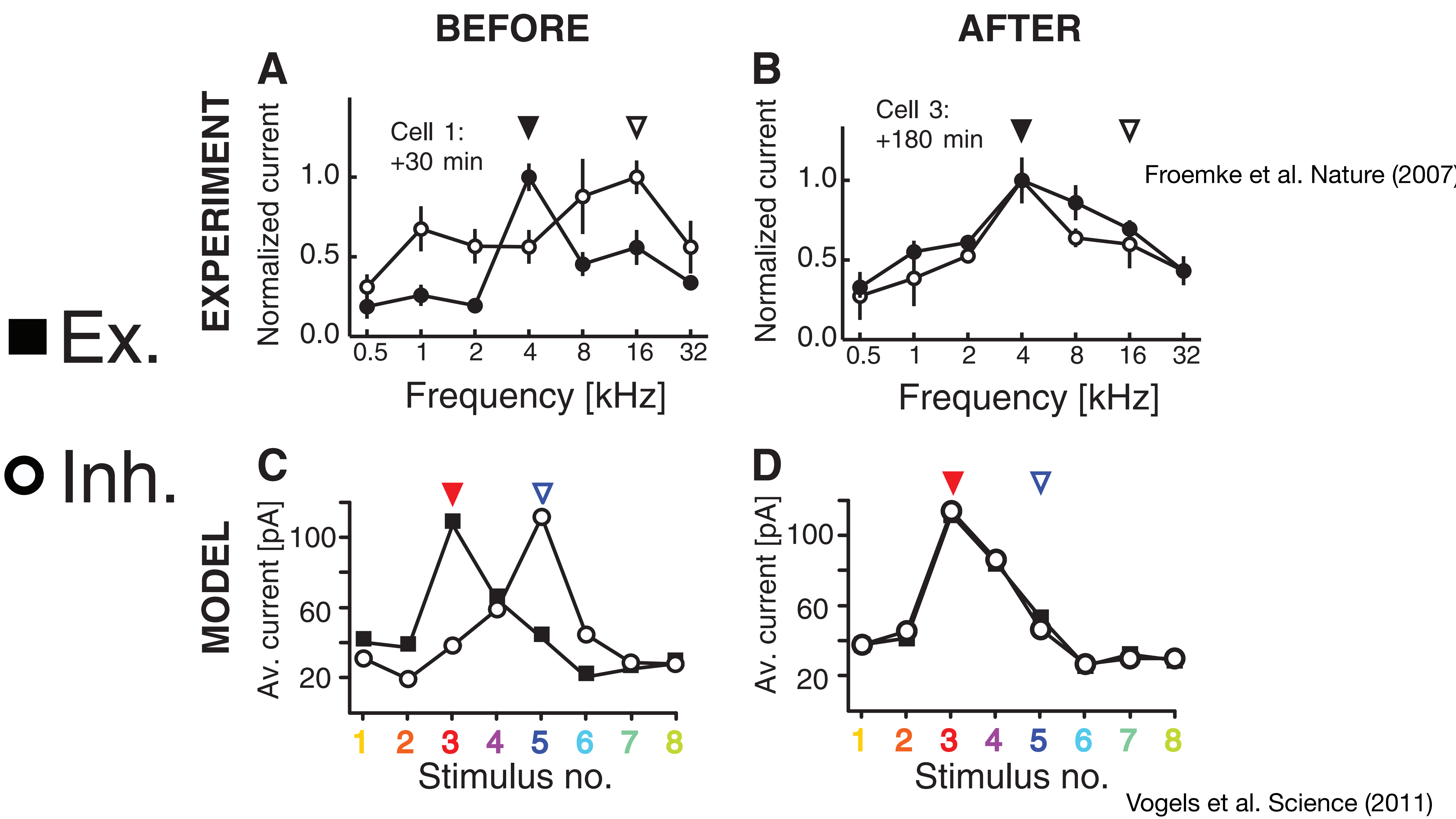
Vogels et al. Science (2011)

Learning to balance excitation and inhibition

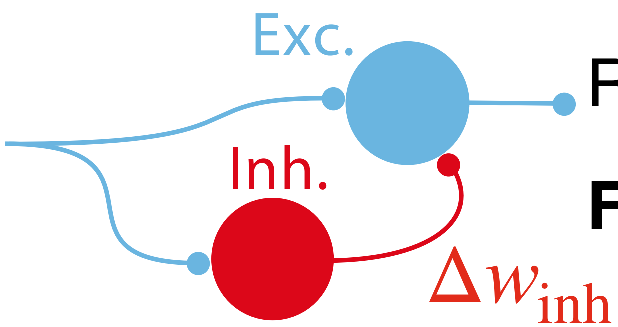


Vogels et al. Science (2011)

Inhibitory plasticity balances receptive fields



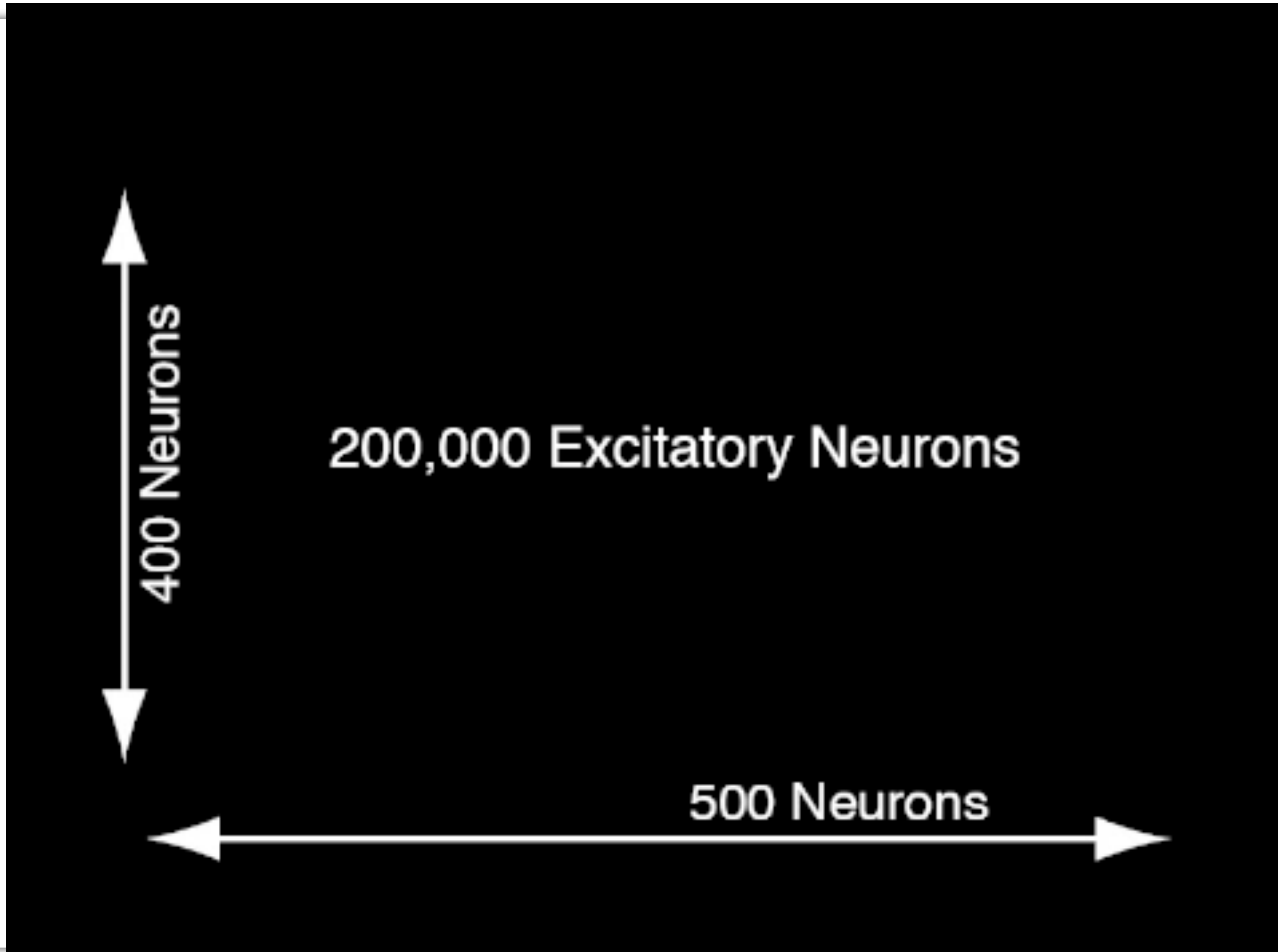
Inhibitory balance in a recurrent neural network



Recurrent network: 200 000 exc. neurons + ~40 000 inh. neurons

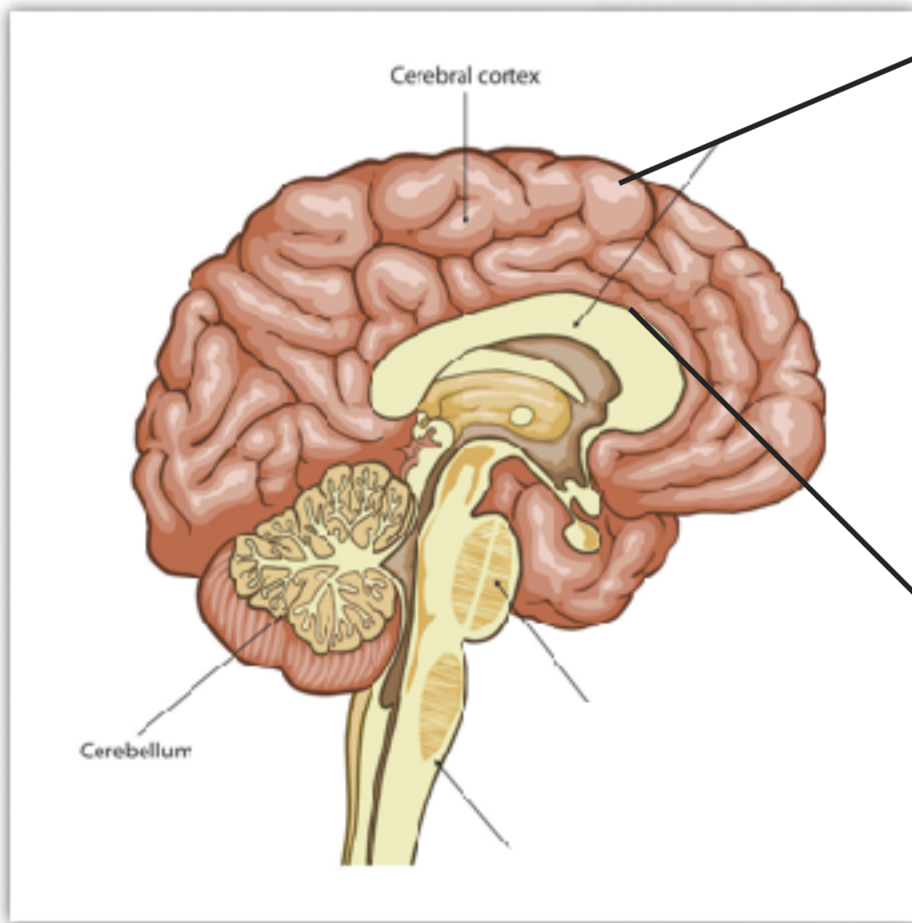
Features: keeps activity under control (homeostasis) and memories hidden

Vogels et al. Science (2011)



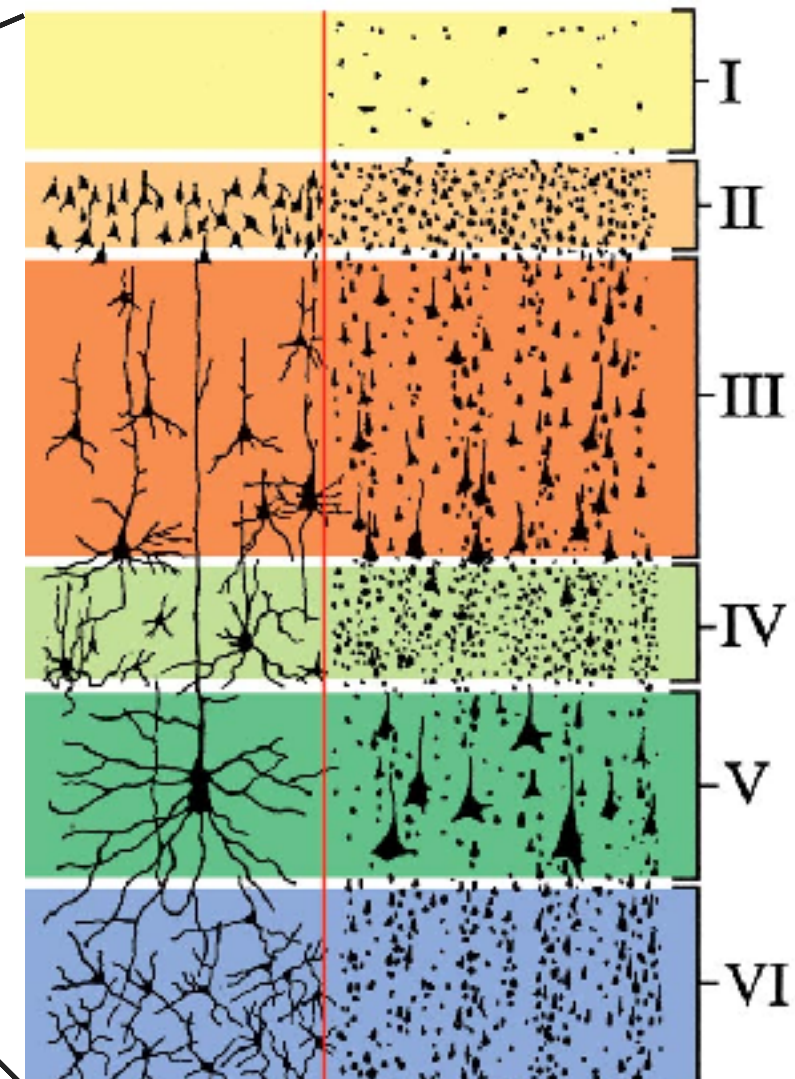
But cortical circuits are way more complicated..

The six neocortical layers



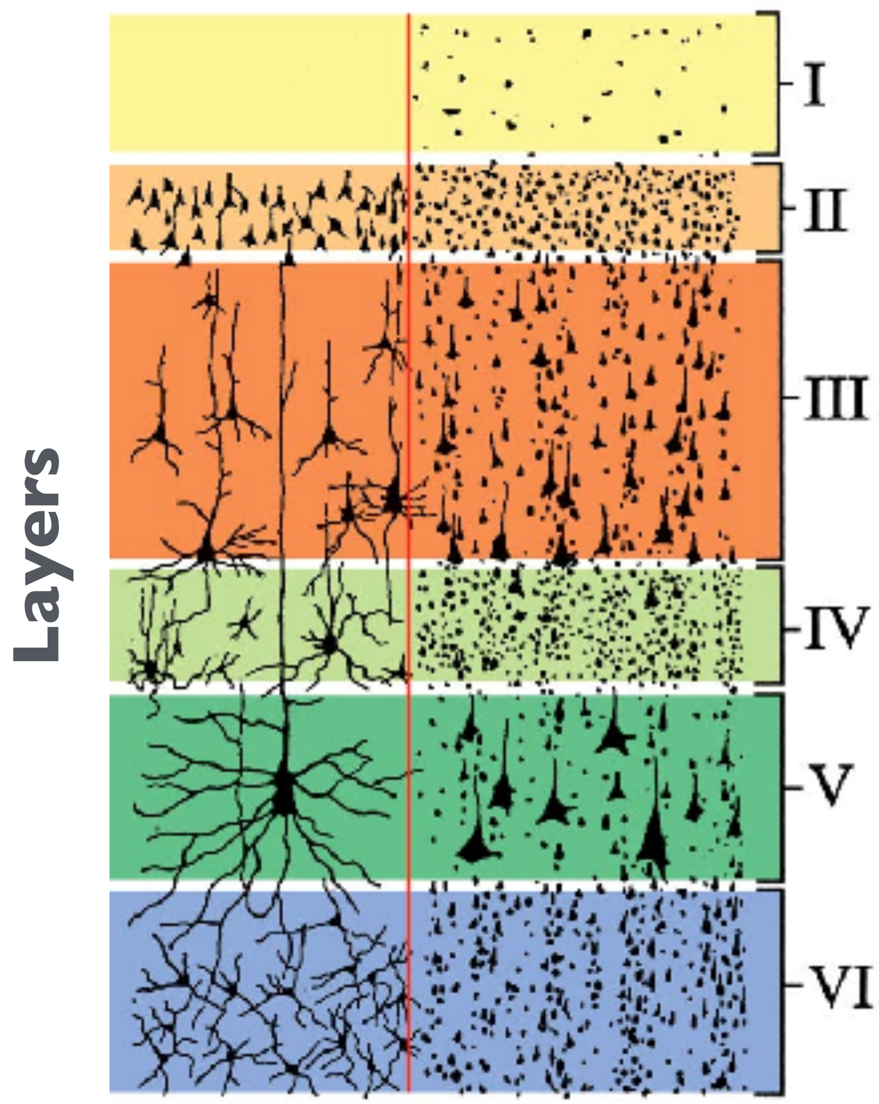
Introduction to Psychology 2015; lib.umn.edu

Neocortical layers



vanat.cvm.umn.edu/brain18

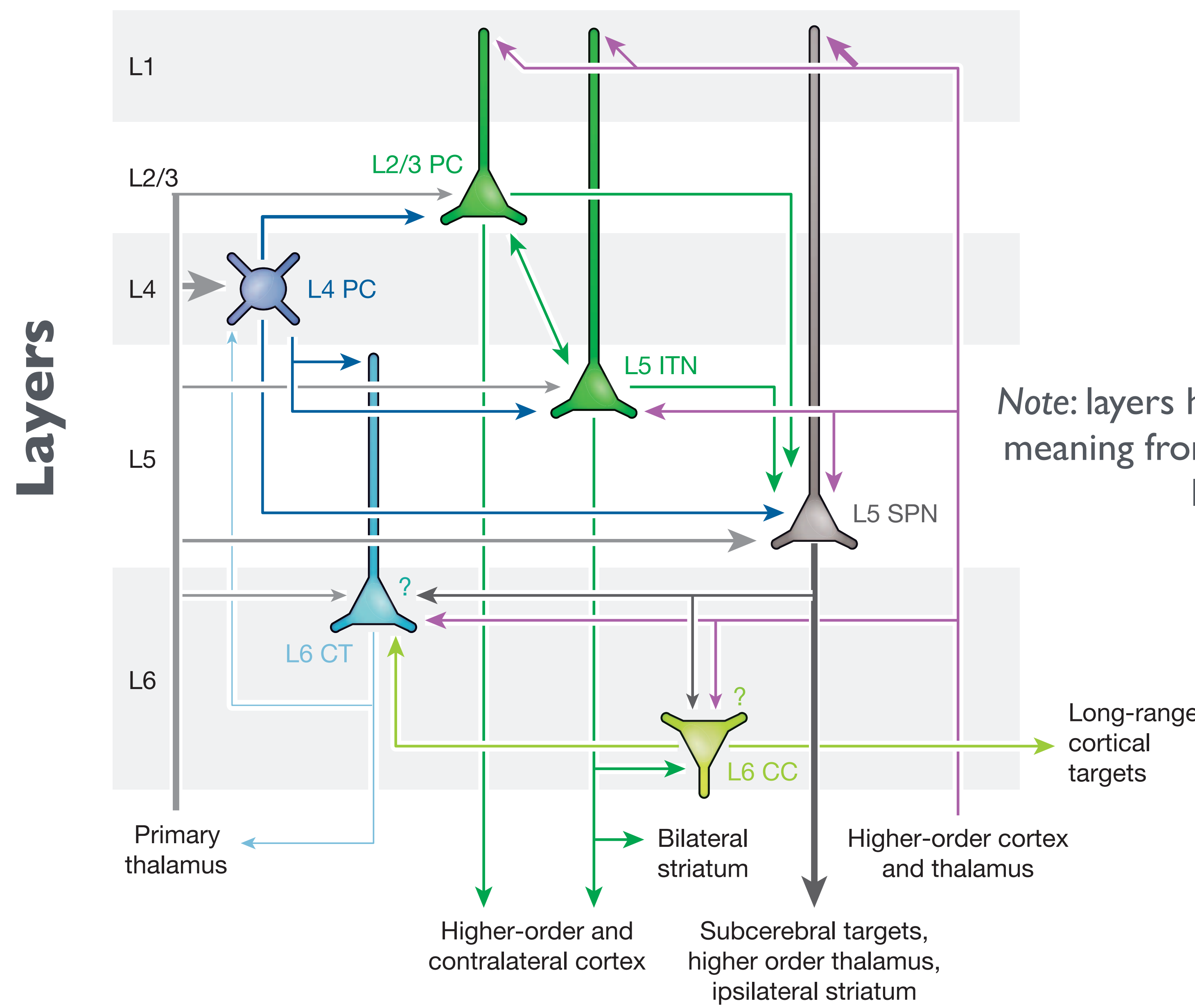
Why so much (apparent) complexity?



Hmm.. what's the neural basis of intelligence?

Ramon y Cajal

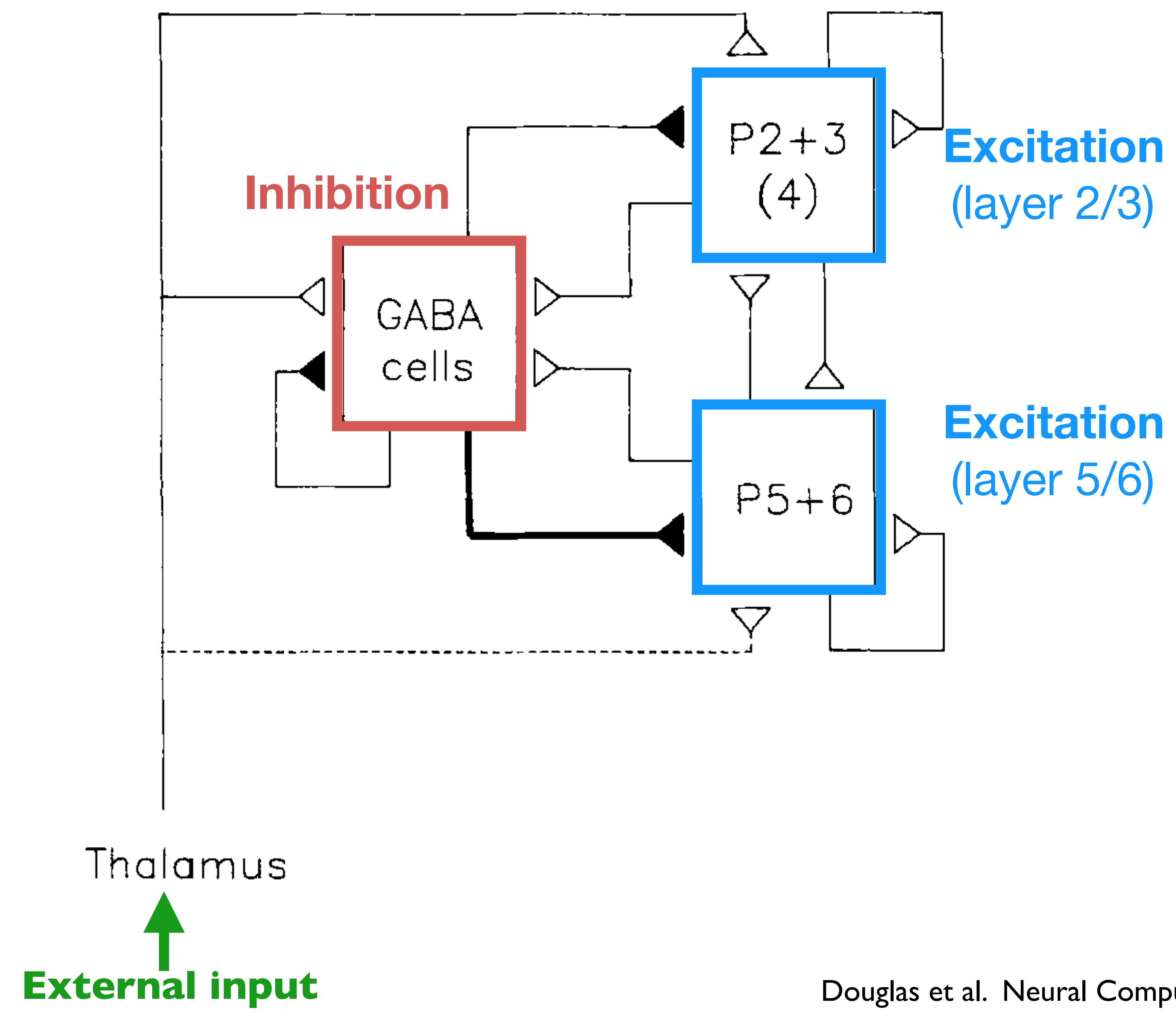
Structure of cortical microcircuits:
excitatory cells



Note: layers here have a different meaning from layers in machine learning

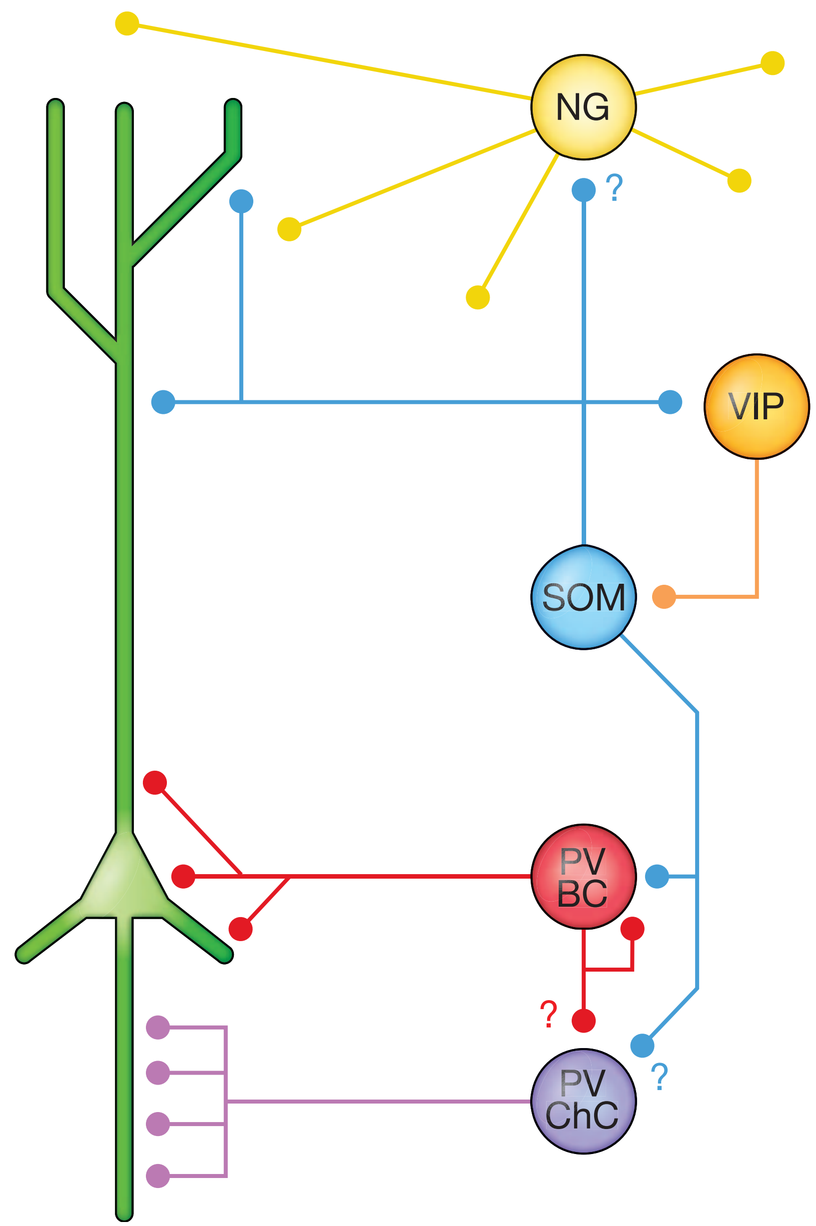
Harris and Mrsic-Flogel. Nature Review 2013

Structure of cortical microcircuits: canonical view



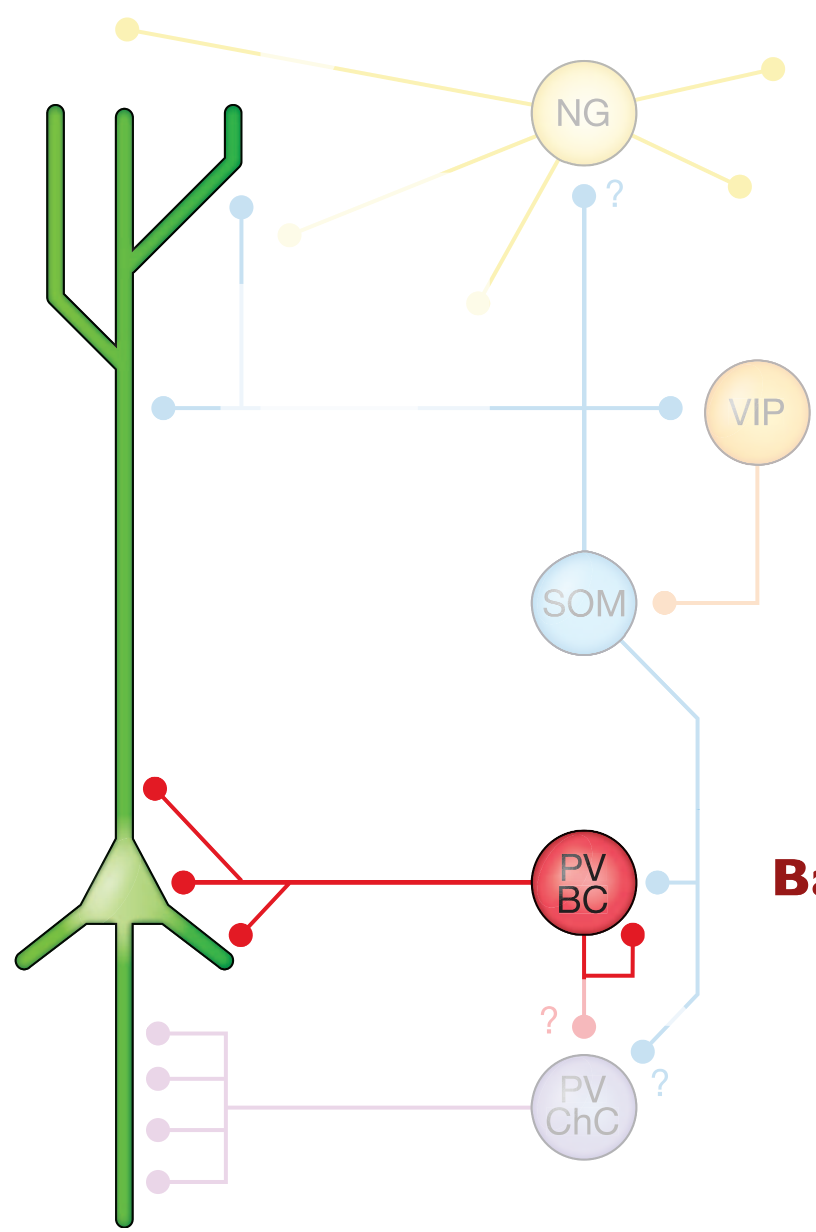
Douglas et al. Neural Computation 1989

Structure of cortical microcircuits:
inhibitory cells (gates)



Harris and Mrsic-Flogel. Nature Review 2013

Structure of cortical microcircuits:
inhibitory cells (gates)



Basket cells (~50% of INs)

Harris and Mrsic-Flogel. Nature Review 2013

Vogels and Abbott NatNeurosci. 2009

Machine learning recurrent neural networks: long short-term memory (LSTM)

- **LSTMs are state-of-the-art** (or close to) **in:**
    - Language modelling (Melis et al. 2017)
    - Caption generation (Lu et al. 2016)
    - Speech recognition (Chan et al. 2016)
    - Machine Translation (Luong et al,. 2015)
    - With new impressive applications every week

Hochreiter and Schmidhuber,
Neural Computation,(1997)

Machine learning recurrent neural networks: long short-term memory (LSTM)

- **LSTMs are state-of-the-art** (or close to) **in:**
  - Language modelling (Melis et al. 2017)
  - Caption generation (Lu et al. 2016)
  - Speech recognition (Chan et al. 2016)
  - Machine Translation (Luong et al,. 2015)
  - With new impressive applications every week

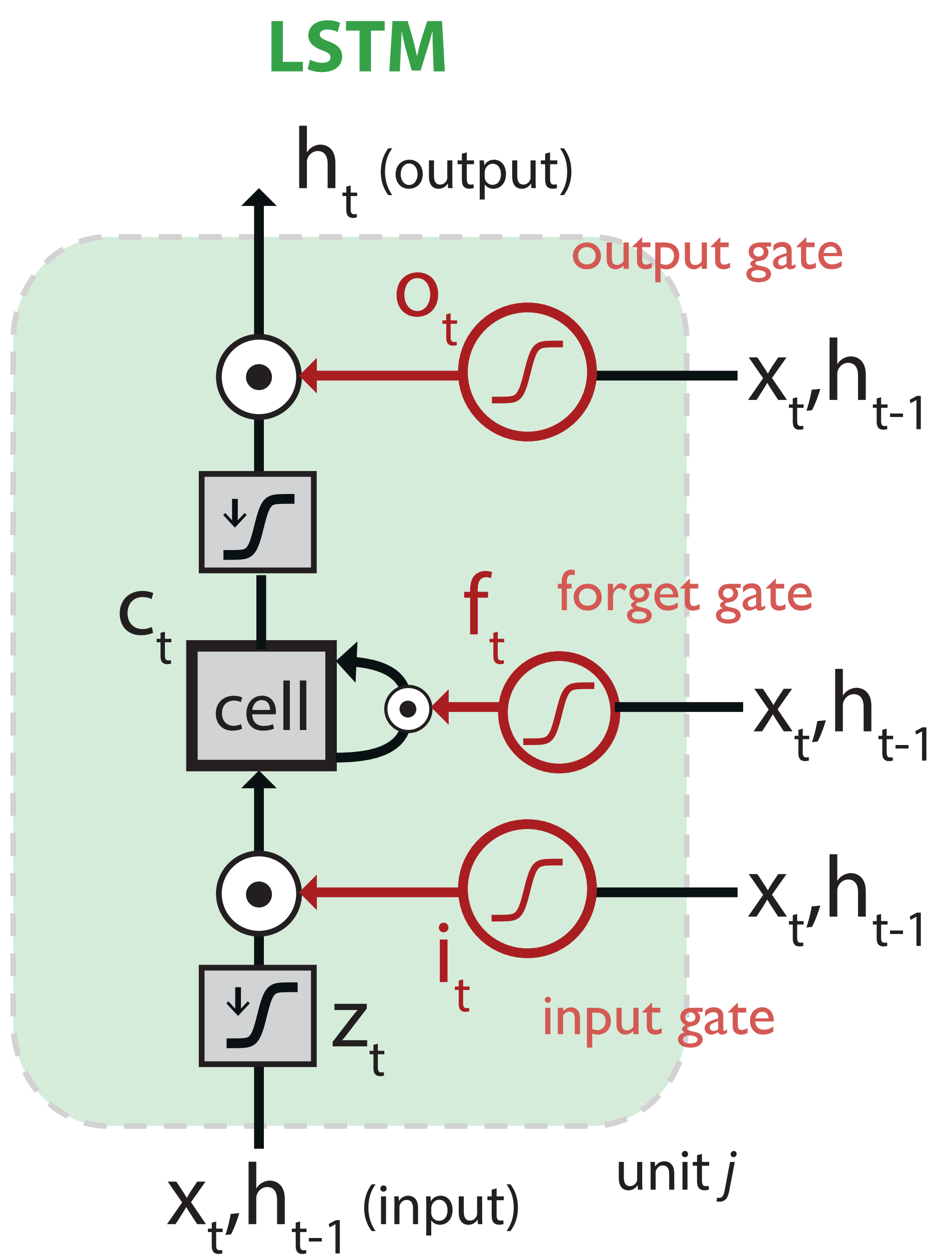
- **At the core of industry applications:**
  - Siri (Apple)
  - Translate (Google)
  - Alexa (Amazon)

Hochreiter and Schmidhuber,
Neural Computation,(1997)

Long short-term memory (LSTM)

Captures long and short-term dependencies!

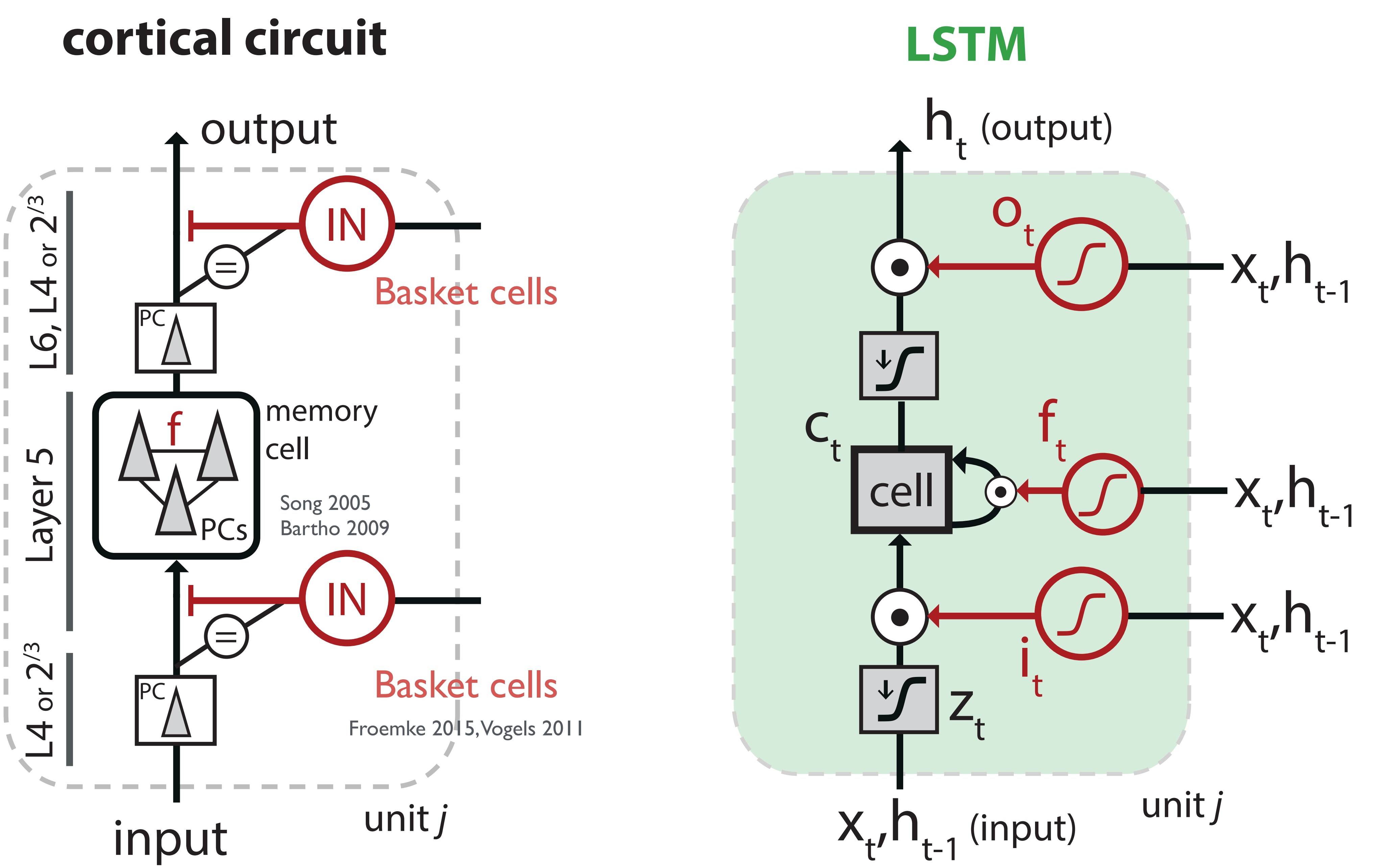
memory cell, c_t



⊙ denote element-wise multiplications

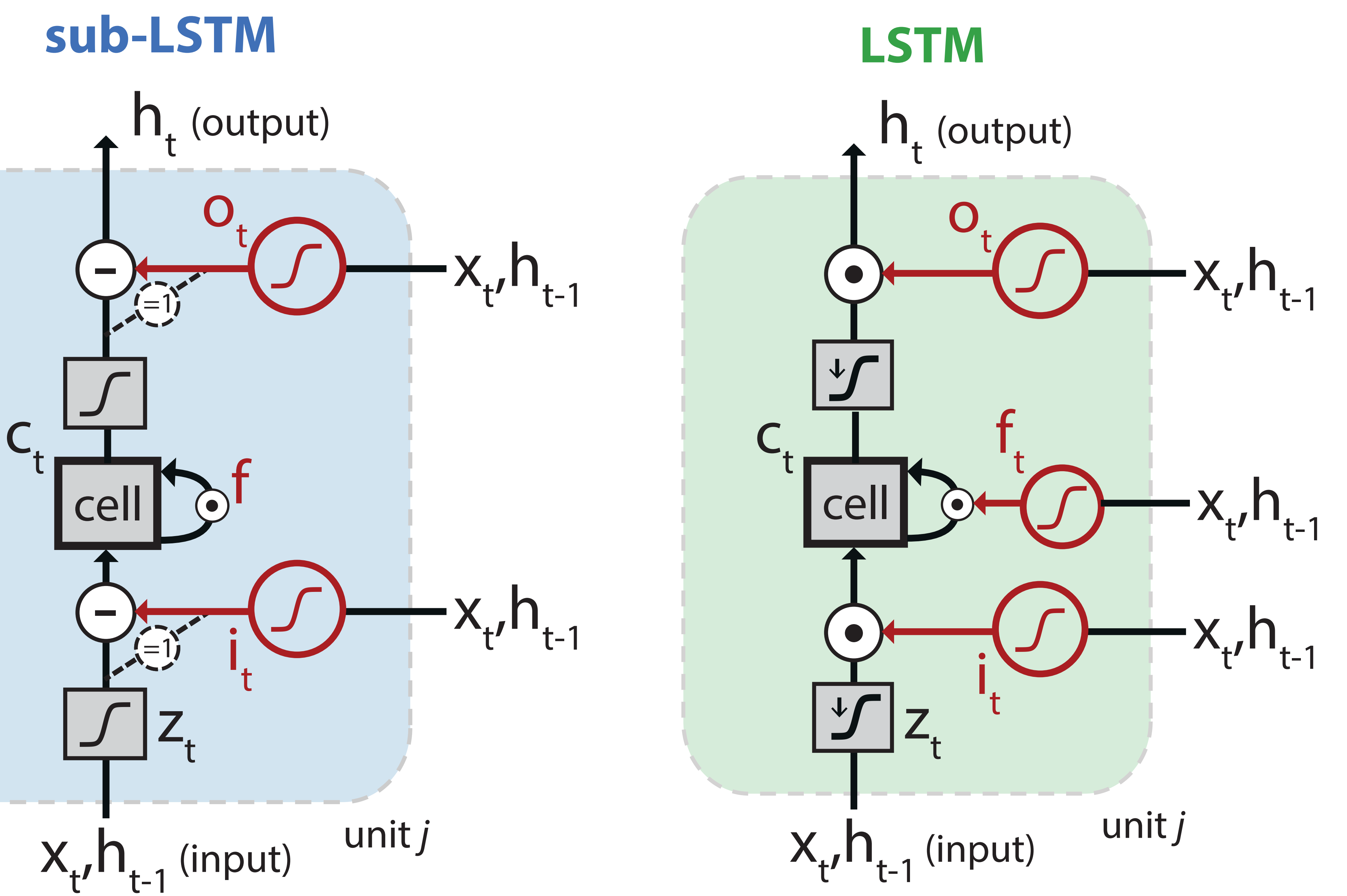
Hochreiter and Schmidhuber, Neural Computation,(1997)

Cortical circuits vs LSTMs



Costa et al. NIPS 2017

Cortical circuits vs LSTMs

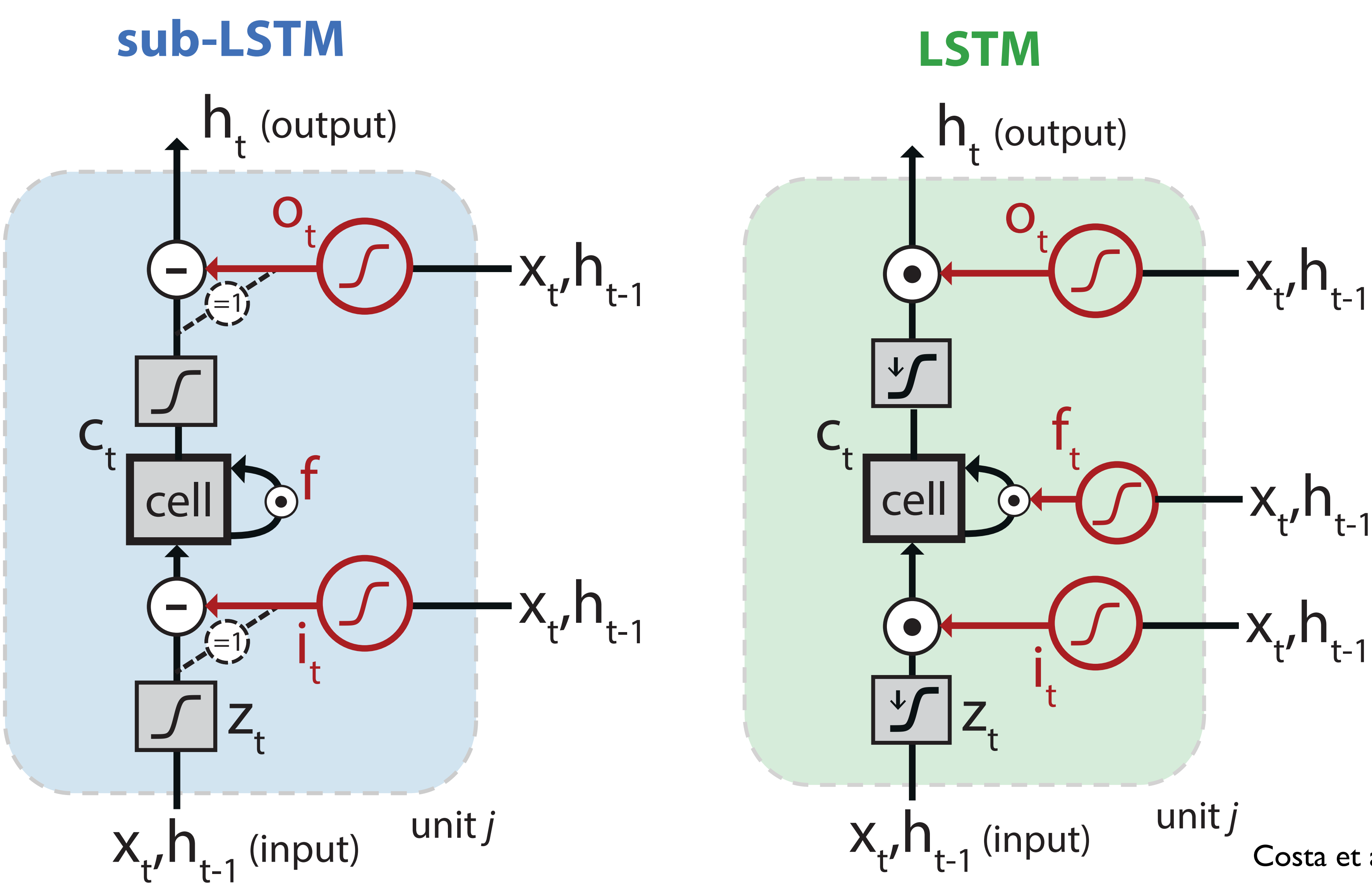


Note: blue now represents subtractive gating

Costa et al. NIPS 2017

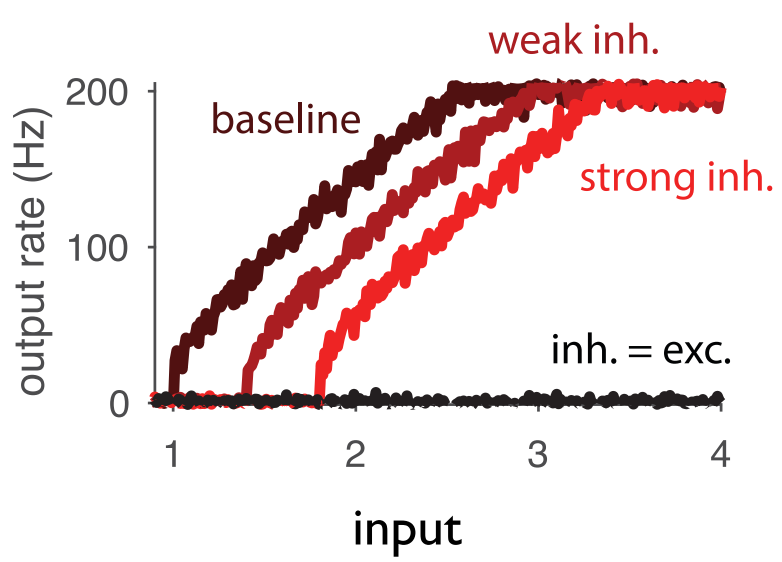
Group discussion
groups of 2-3 (5min)

When would the gates be **closed** and **open** for the **subtractive** and **multiplicative** gating?



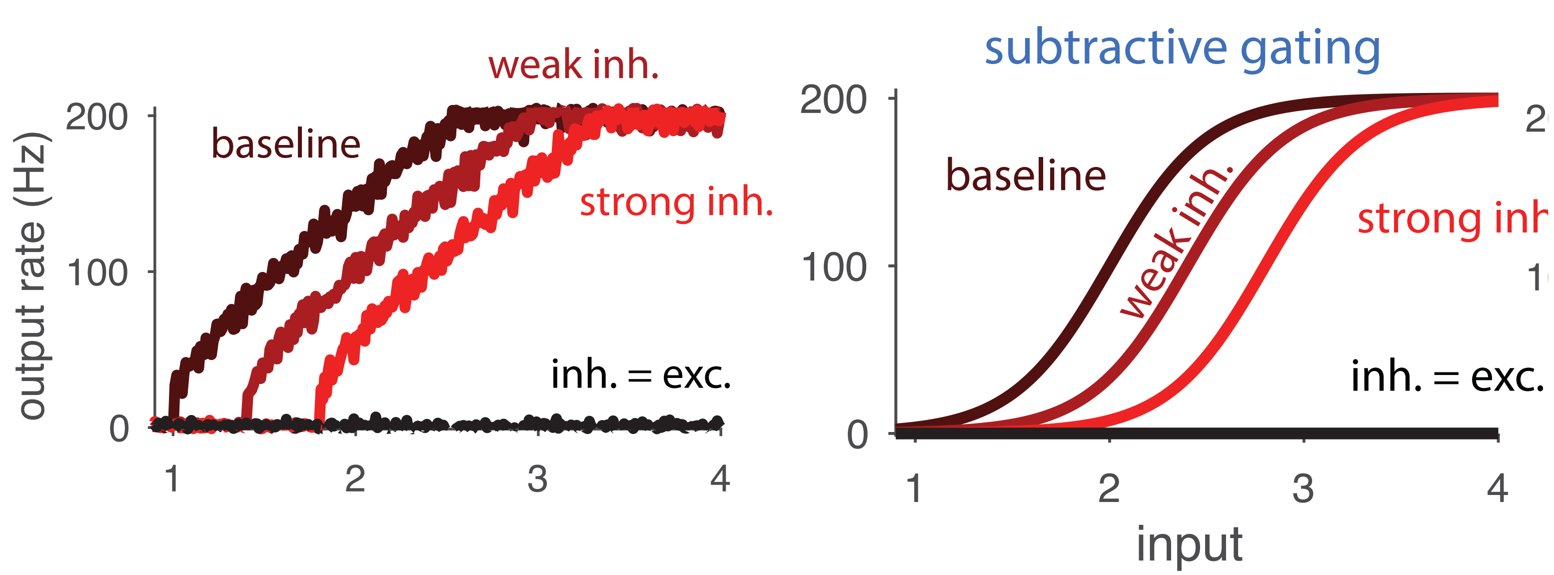
Costa et al. NIPS 2017

Cortical circuits vs LSTMs: input-output



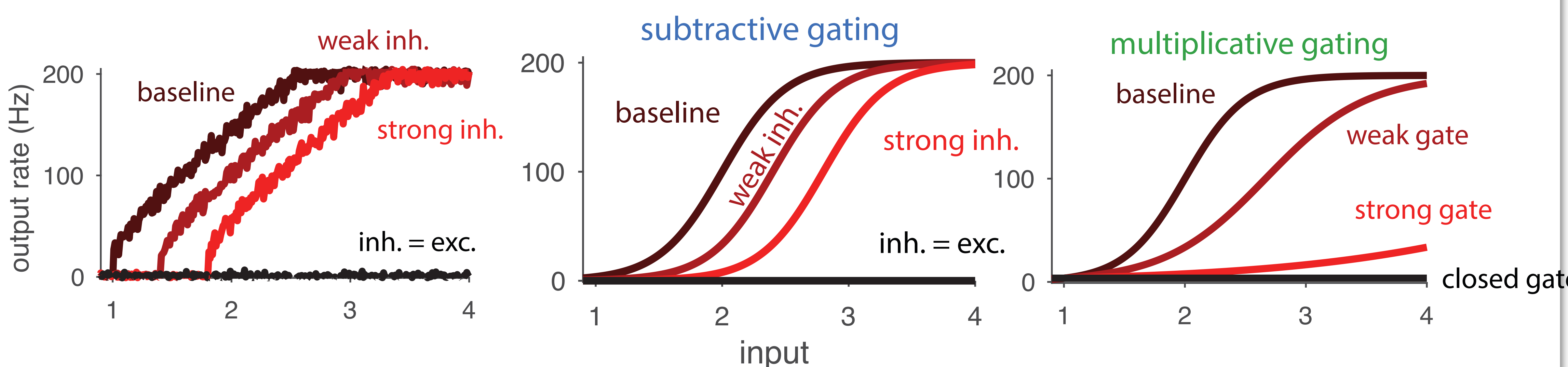
Costa et al. NIPS 2017

Cortical circuits vs LSTMs: input-output



Costa et al. NIPS 2017

Cortical circuits vs LSTMs: input-output

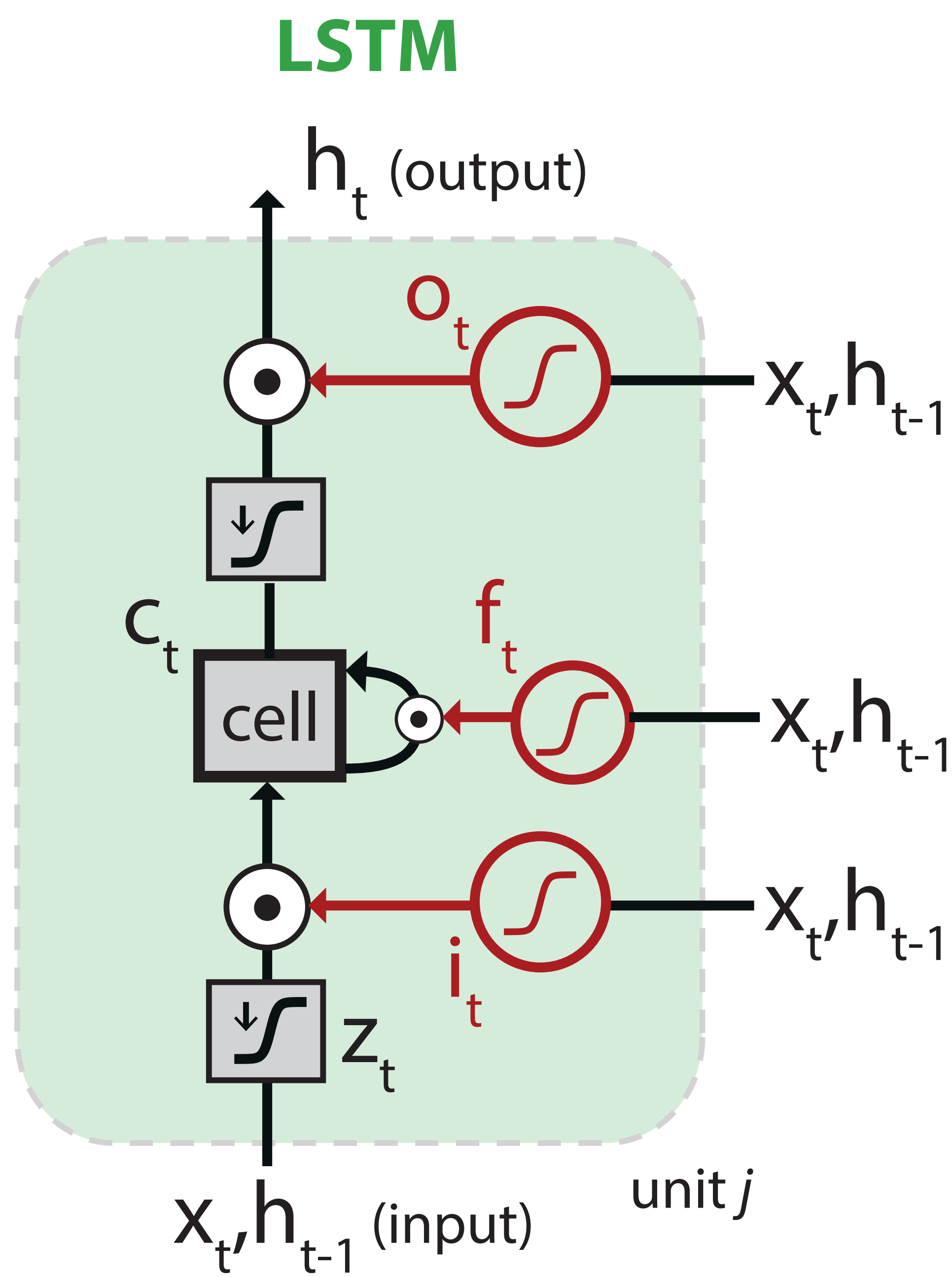


Costa et al. NIPS 2017

subtractive LSTMs

$$\boxed{\textbf{LSTM}}$$

$$[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T = \Bigg| \quad \sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$$

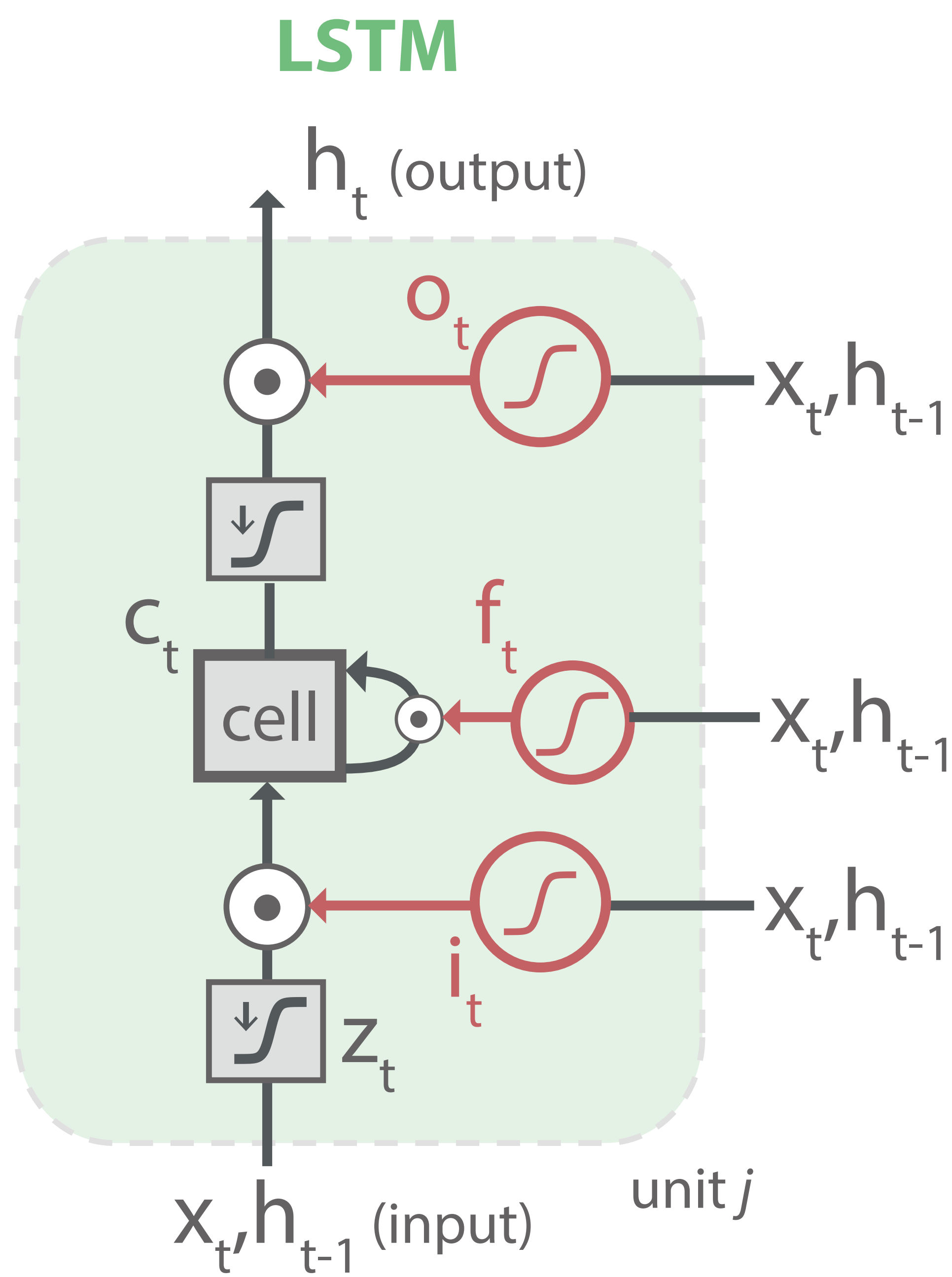


Costa et al. NIPS 2017

subtractive LSTMs

$\boxed{\textbf{LSTM}}$

$$
\begin{array}{r|c}
[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T = & \sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}), \\
\mathbf{z}_t = & \tanh(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),
\end{array}
$$

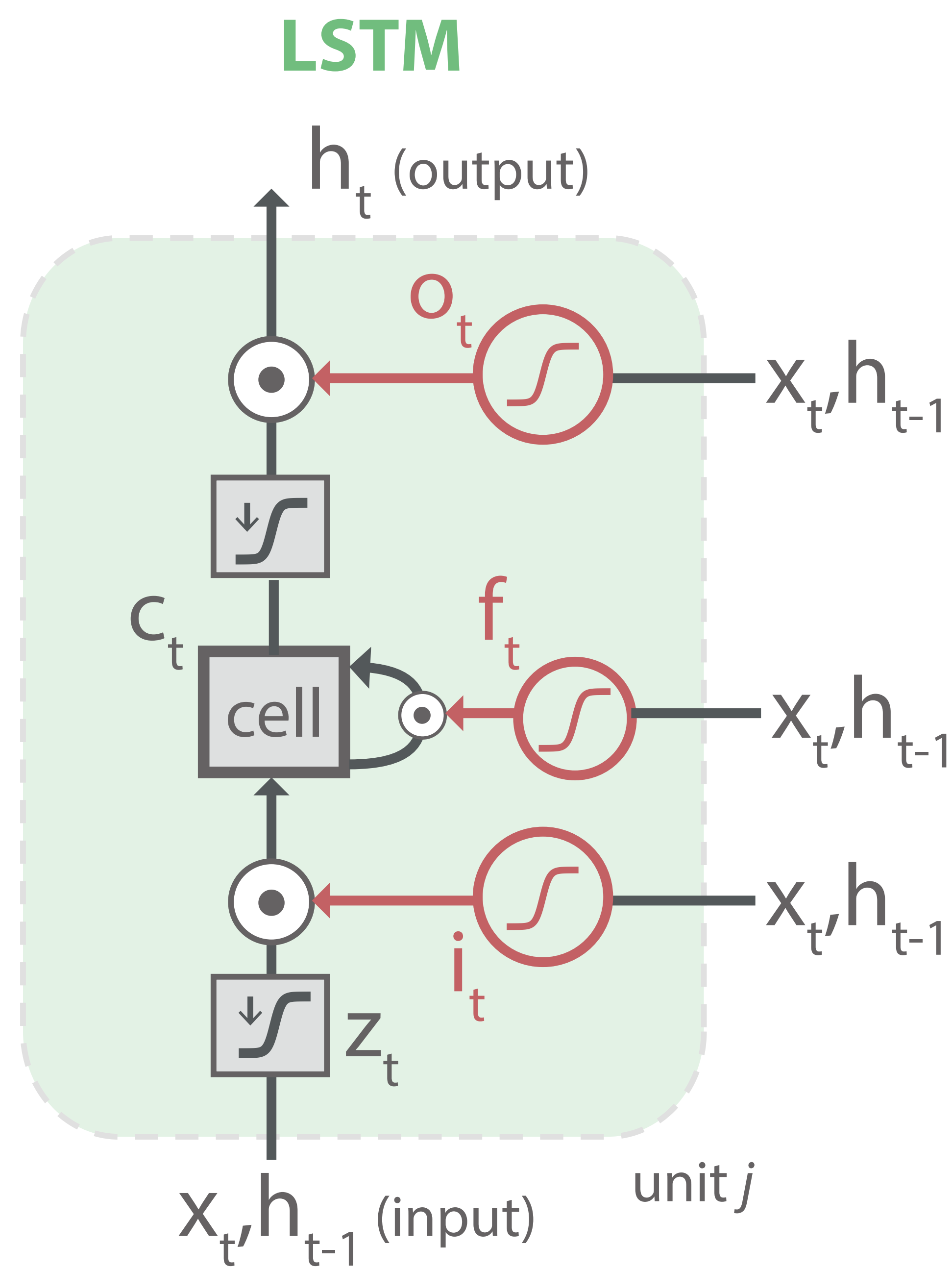


Costa et al. NIPS 2017

subtractive LSTMs

LSTM

$$
\begin{aligned}
[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T = & \quad \sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}), \\
\mathbf{z}_t = & \quad \tanh(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}), \\
\mathbf{c}_t = & \quad \mathbf{c}_{t-1} \odot \mathbf{f}_t + \mathbf{z}_t \odot \mathbf{i}_t,
\end{aligned}
$$



Costa et al. NIPS 2017

subtractive LSTMs

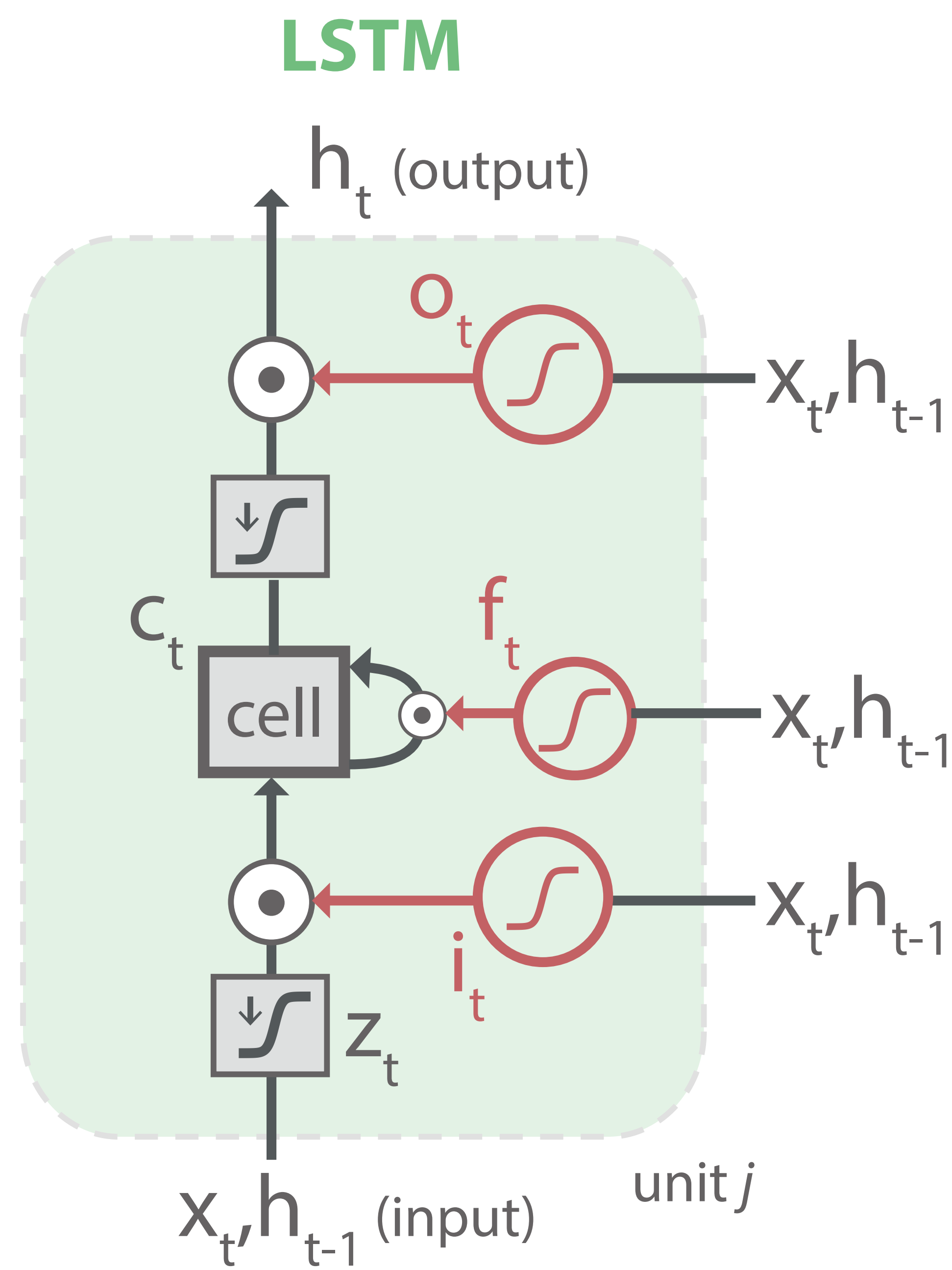
LSTM

$$[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T = \sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$$

$$\mathbf{z}_t = \tanh(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$$

$$\mathbf{c}_t = \mathbf{c}_{t-1} \odot \mathbf{f}_t + \mathbf{z}_t \odot \mathbf{i}_t,$$

$$\mathbf{h}_t = \tanh(\mathbf{c}_t) \odot \mathbf{o}_t.$$



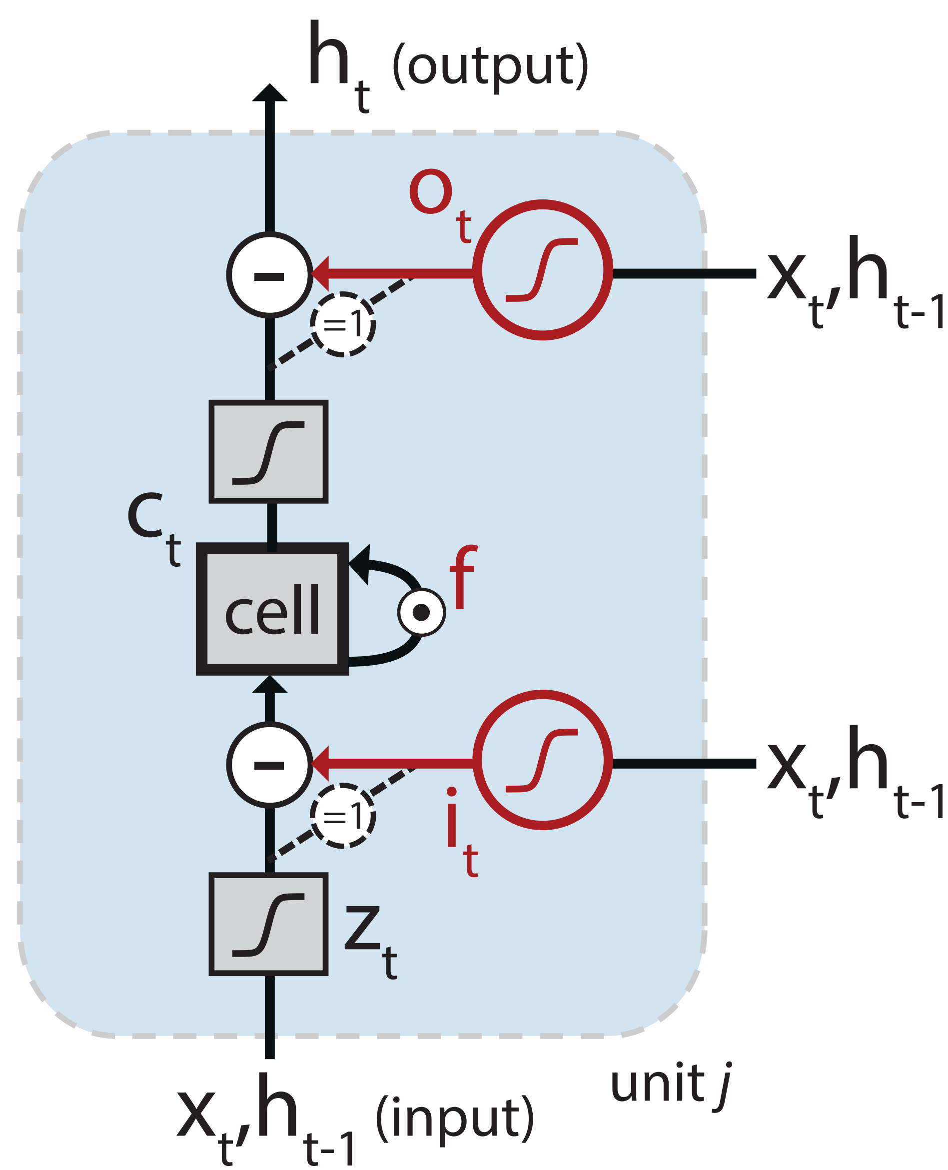
Costa et al. NIPS 2017

subtractive LSTMs

$\boxed{\textbf{subLSTM}}$

$$
\begin{aligned}
[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T &= \sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}), \\
\mathbf{z}_t &= \sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}), \\
\mathbf{c}_t &= \mathbf{c}_{t-1} \odot \mathbf{f}_t + \mathbf{z}_t - \mathbf{i}_t, \\
\mathbf{h}_t &= \sigma(\mathbf{c}_t) - \mathbf{o}_t.
\end{aligned}
$$

sub-LSTM



Costa et al. NIPS 2017

subtractive LSTMs vs LSTMs

	subLSTM	LSTM
$[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T =$	$\sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$	$\sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$
$\mathbf{z}_t =$	$\sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$	$\tanh(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$
$\mathbf{c}_t =$	$\mathbf{c}_{t-1} \odot \mathbf{f}_t + \mathbf{z}_t - \mathbf{i}_t,$	$\mathbf{c}_{t-1} \odot \mathbf{f}_t + \mathbf{z}_t \odot \mathbf{i}_t,$
$\mathbf{h}_t =$	$\sigma(\mathbf{c}_t) - \mathbf{o}_t.$	$\tanh(\mathbf{c}_t) \odot \mathbf{o}_t.$

Costa et al. NIPS 2017

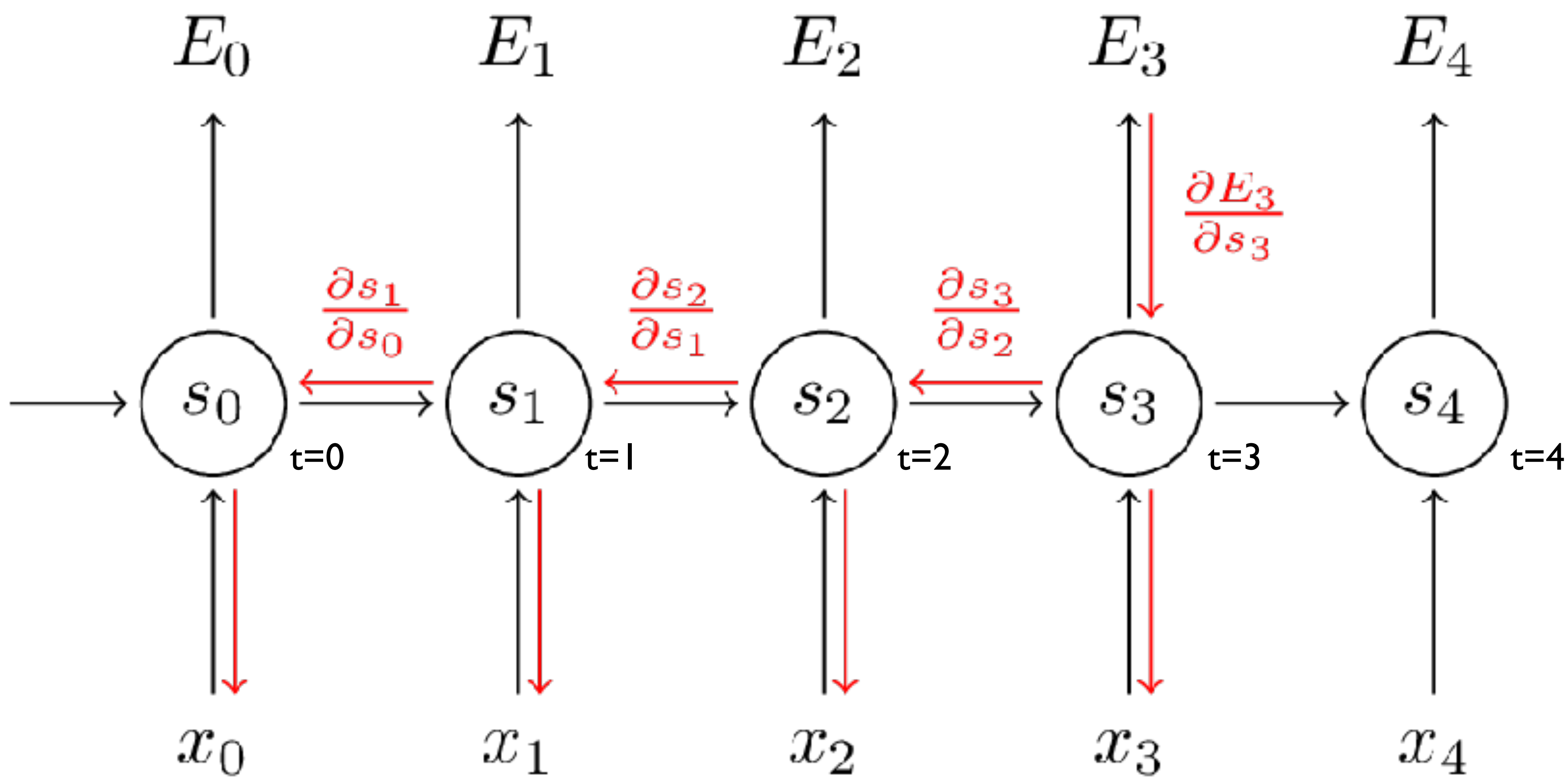
subtractive LSTMs vs LSTMs

	subLSTM	**LSTM**
$[\mathbf{f}_t, \mathbf{o}_t, \mathbf{i}_t]^T =$	$\sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$	$\sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$
$\mathbf{z}_t =$	$\sigma(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$	$\tanh(W\mathbf{x}_t + R\mathbf{h}_{t-1} + \mathbf{b}),$
$\mathbf{c}_t =$	$\mathbf{c}_{t-1} \odot \mathbf{f}_t + \boxed{\mathbf{z}_t - \mathbf{i}_t,}$	$\mathbf{c}_{t-1} \odot \mathbf{f}_t + \boxed{\mathbf{z}_t \odot \mathbf{i}_t,}$
$\mathbf{h}_t =$	$\boxed{\sigma(\mathbf{c}_t) - \mathbf{o}_t.}$	$\boxed{\tanh(\mathbf{c}_t) \odot \mathbf{o}_t.}$

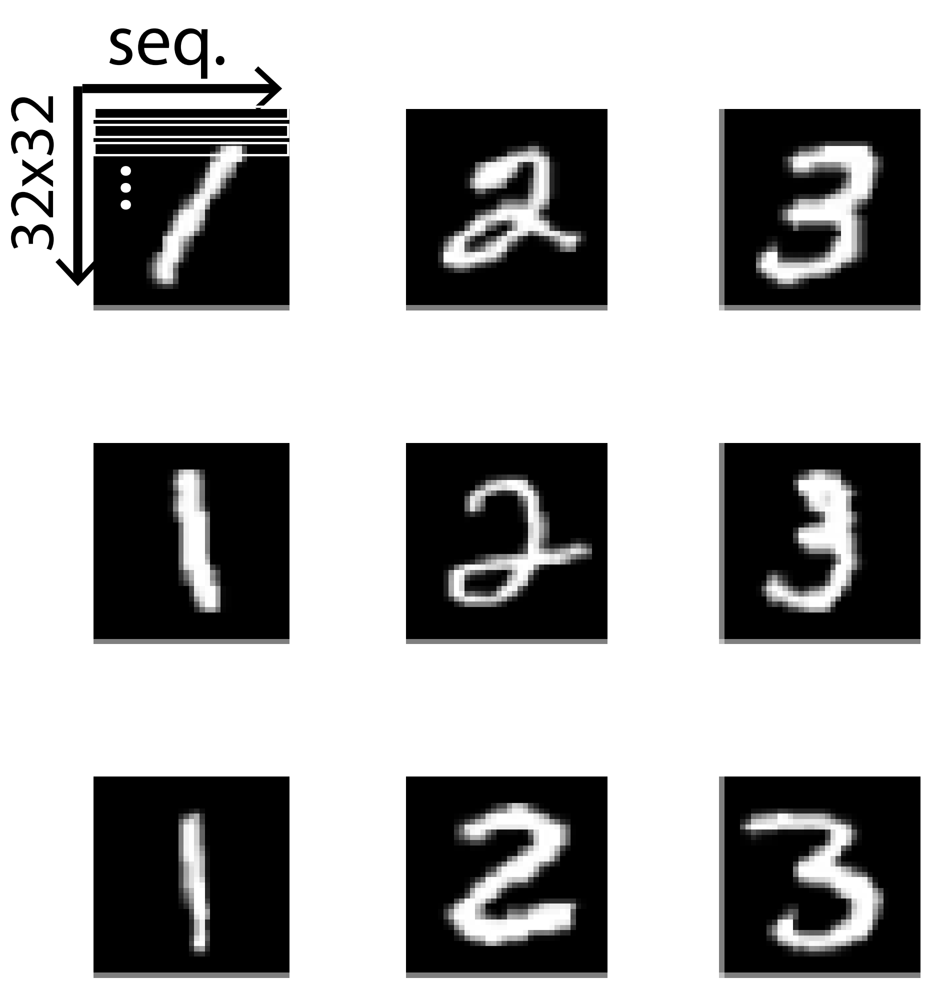
Costa et al. NIPS 2017

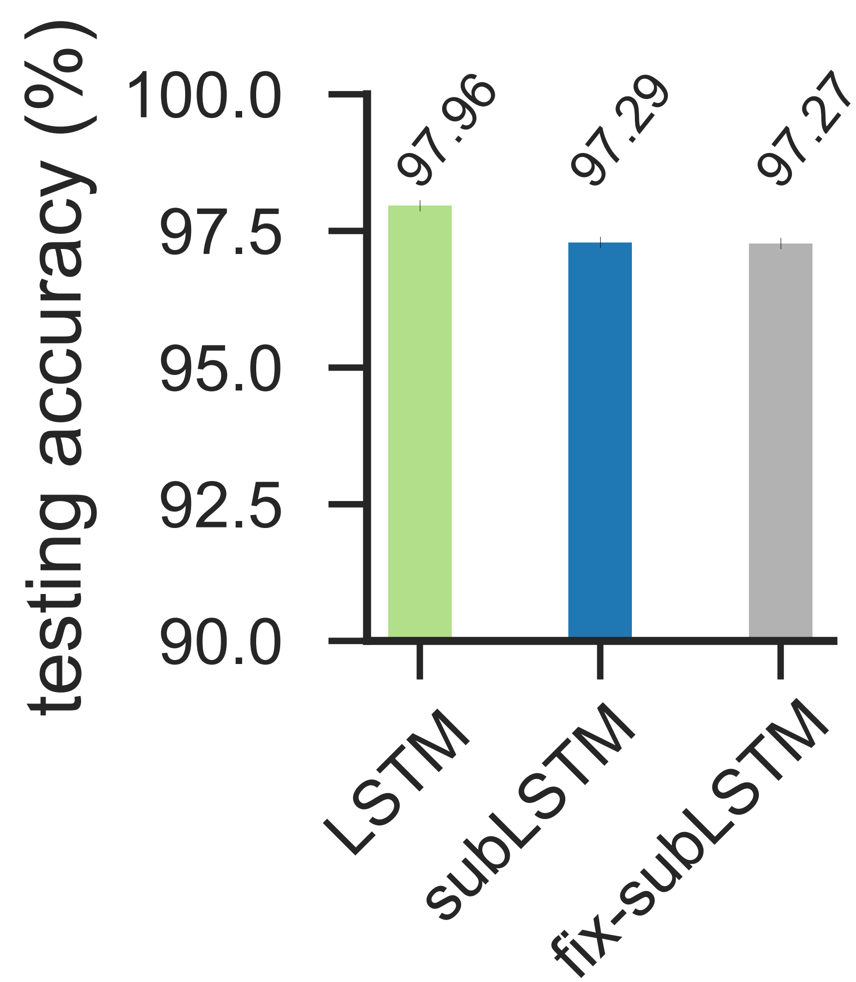
Gated RNNs are usually trained using BackPropagation Through Time (BPTT)

Similar to backprop, but now we unfold the network across time and backprop the gradients 'back in time' (each timestep is a layer).



Task 1: Pixel-by-pixel sequential MNIST (dataset of handwritten digits)





Costa et al. NIPS 2017

Task 2: Language modelling (word-based)
Penn Treebank dataset

Penn Treebank dataset:
Training: 929k; Validation: 73k; Test: 82k; Vocabulary: 10k

"...since then life has changed a lot for X"

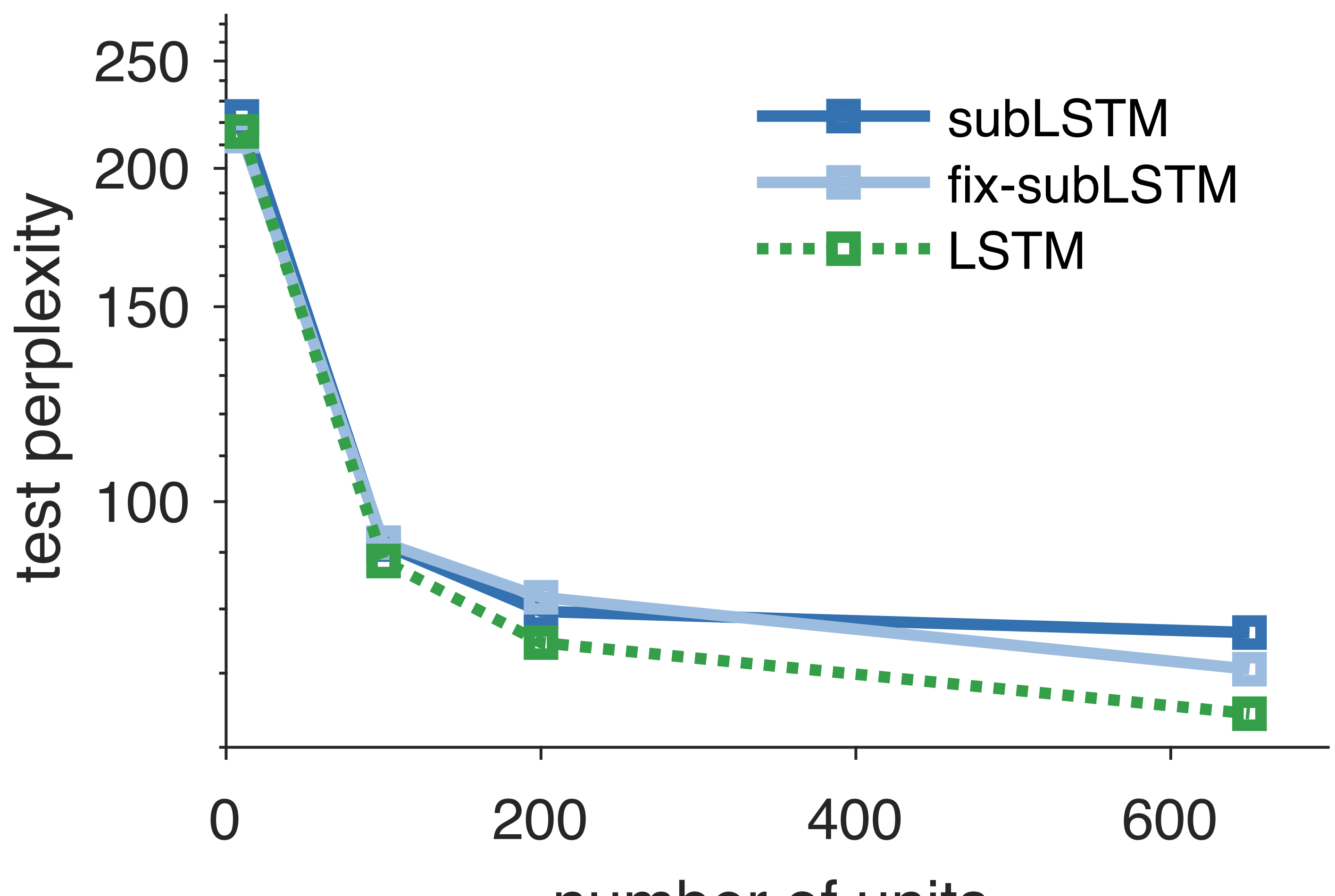
Costa et al. NIPS 2017

Task 2: Language modelling (word-based)
Penn Treebank dataset

Penn Treebank dataset:

Training: 929k; Validation: 73k; Test: 82k; Vocabulary: 10k

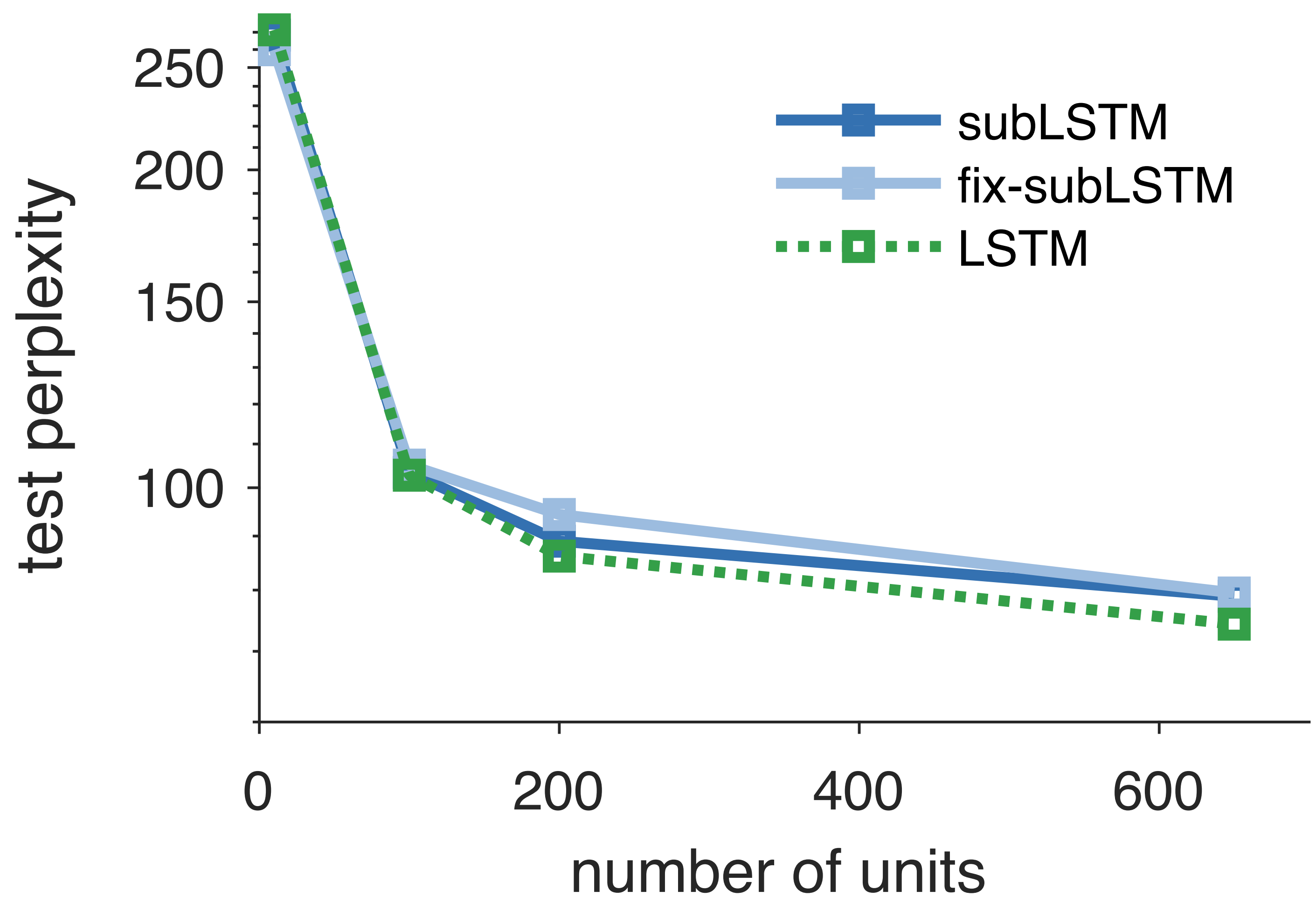
"...since then life has changed a lot for X"



Costa et al. NIPS 2017

Task 3: Language modelling (word-based) Wikitext-2

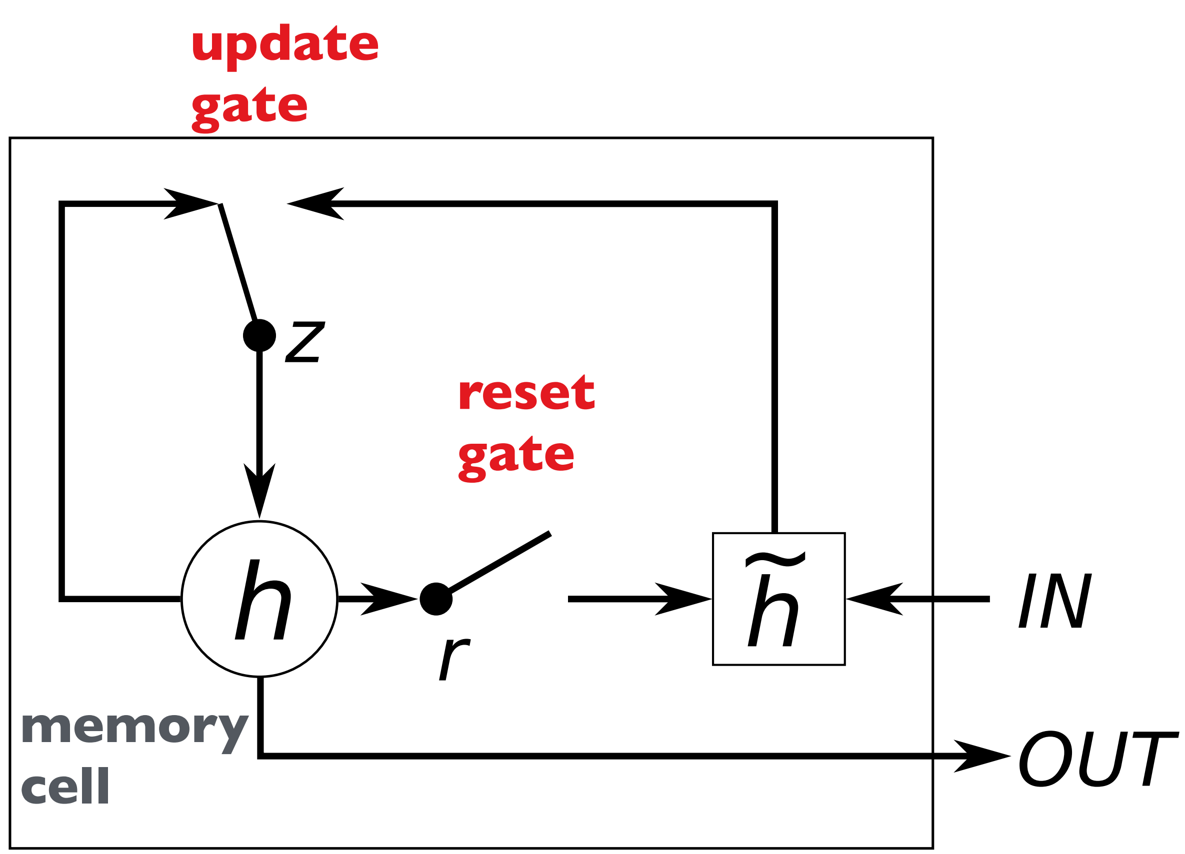
Wikitext-2 dataset:
Training: 2000k; Validation: 217k; Test: 245k; Vocabulary: 33k



Costa et al. NIPS 2017

Gated recurrent units

There are other popular recurrent neural networks, such as the **gated recurrent units** (GRUs):



GRUs are simpler (less parameters) than LSTMs, and obtain competitive results in some tasks.

Chung et al. arXiv 2014

Summary

1. **Multiple excitatory and inhibitory cell types in the brain**

2. **Intricate microcircuits across multiple layers**

3. **Machine learning LSTMs are a form of gated-RNN good for capturing long-term dependencies** (e.g. language modelling)

4. **Cortical microcircuits have similar features to gated-RNNs but** (may) **operate with subtractive gating (subLSTMs)**

References

Text books:

Neuronal Dynamics: Gerstner et al. (2014)

Deep Learning by Courville, Goodfellow and Bengio (2015)

Relevant papers:

- Hennequin et al. Inhibitory Plasticity: Balance, Control, and Codependence. *Annual Review of Neuroscience*, (2017) [review on balanced neural networks]
- Greff et al. LSTM: A Search Space Odyssey, arXiv (2015)
- Costa et al. Cortical Microcircuits as Gated Recurrent Neural Networks. Neural Information Processing (2017) [paper that first introduced the mapping between gated-RNNs and cortical networks]
- Harris and Mrsic-Flogel. Nature Review (2013) [More general review on cortical microcircuits]

Upcoming lectures:
How to relate all these models to experimental data?

- **L17-18: Data science for neural population data** (by Cian O'Donnell)