Deep learning with segregated dendrites

Blake Aaron Richards
Learning in Neural Circuits Laboratory
Department of Biological Sciences
University of Toronto Scarborough

LiNC LAB
linclab.org

UNIVERSITY OF TORONTO SCARBOROUGH

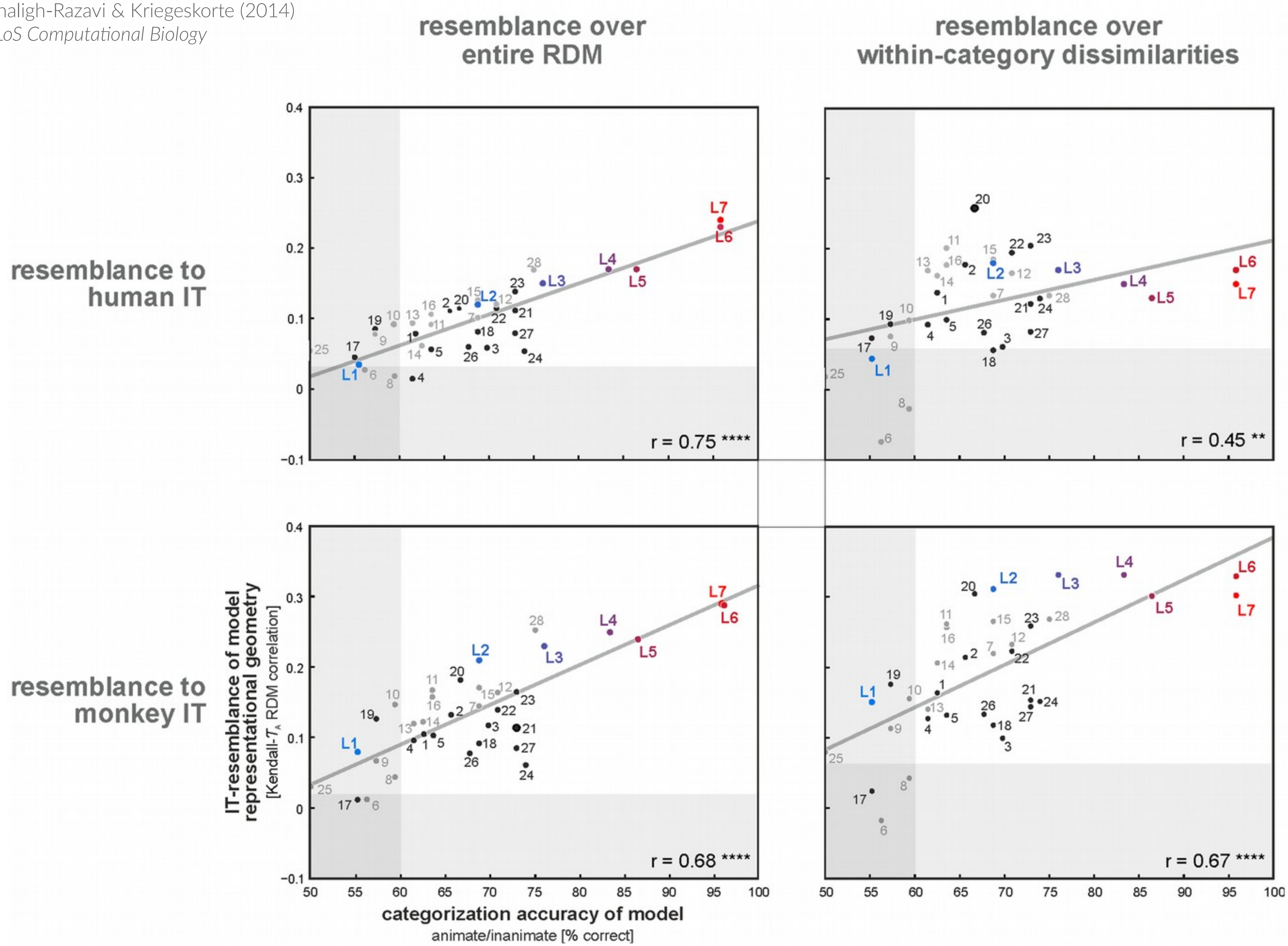
Deep learning has revolutionized artificial intelligence

I think the real brain engages in deep learning

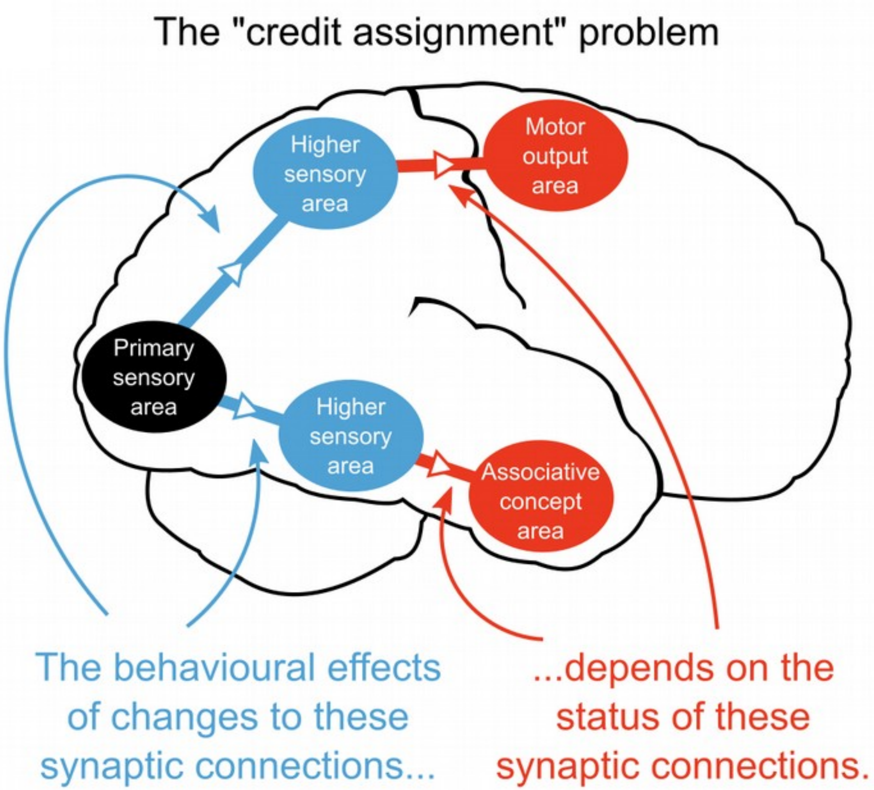
By this I mean the following:
I believe that the brain has mechanisms that allow it to ***solve the credit assignment problem*** in order to optimize ***global cost functions***

1. Deep learning in biological neural networks

Khaligh-Razavi & Kriegeskorte (2014)
PLoS Computational Biology



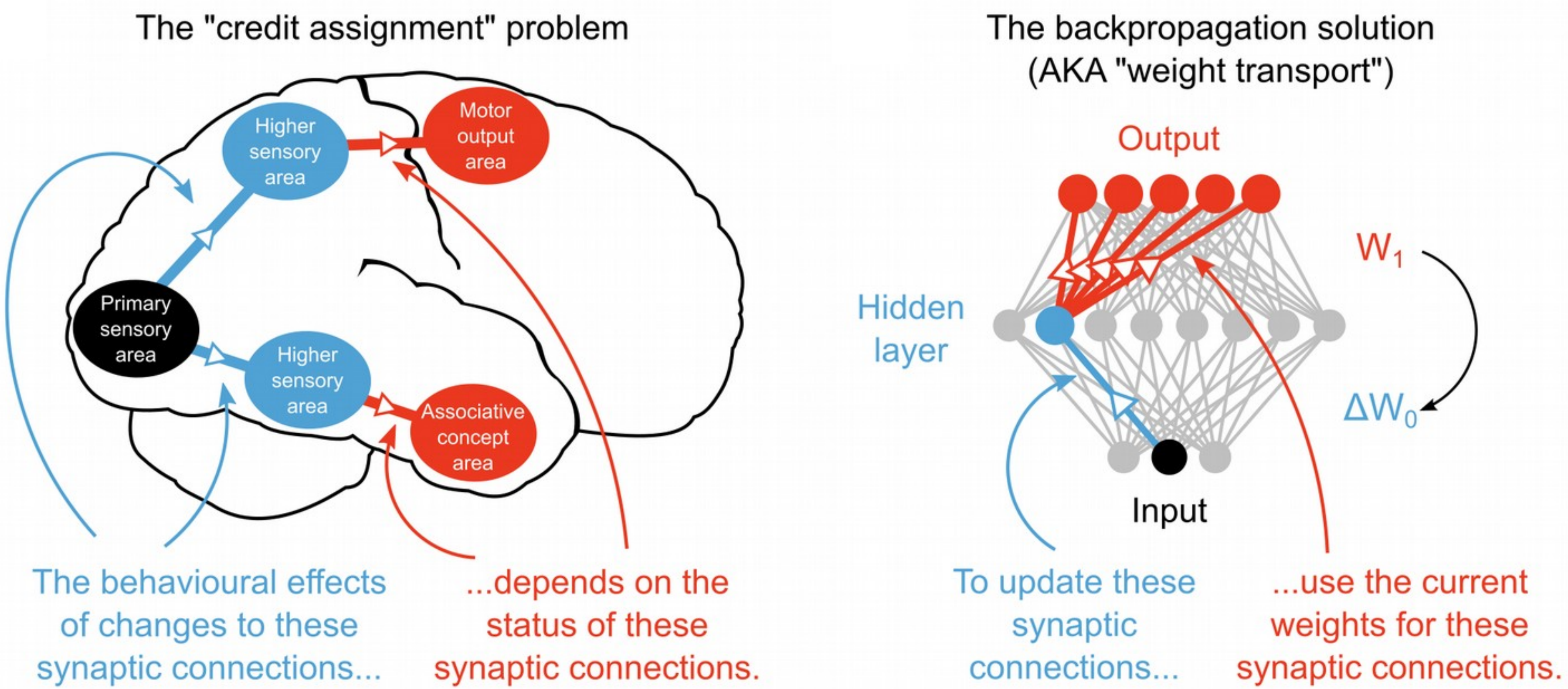
1. Deep learning in biological neural networks



Neuroscientists have almost completely ignored the credit assignment problem for decades

(Or more accurately, they have assumed that reinforcement signals are sufficient for global cost function optimization)

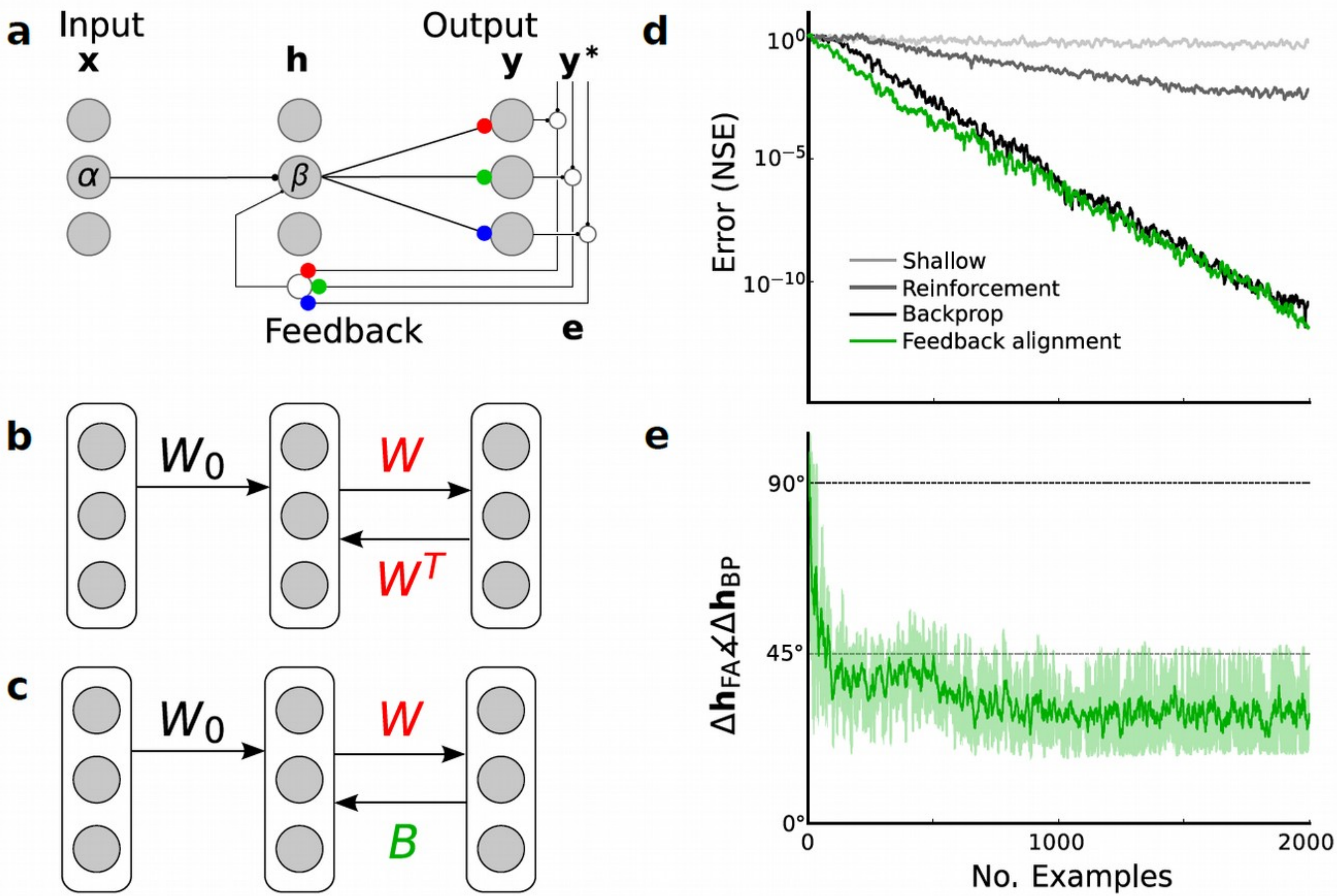
1. Deep learning in biological neural networks



Neuroscientists have almost completely ignored the credit assignment problem for decades

(Or more accurately, they have assumed that reinforcement signals are sufficient for global cost function optimization)

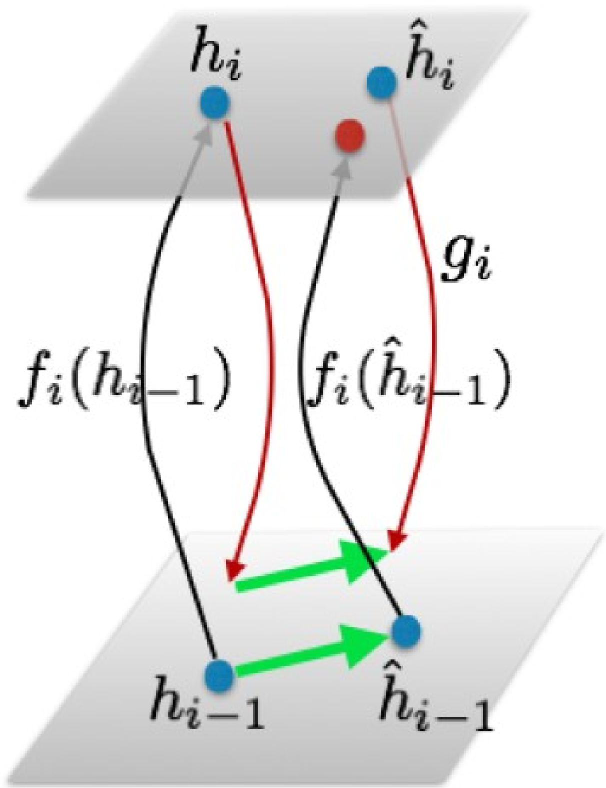
2. Deep learning with random feedback weights



Surprisingly, credit assignment can be done with random feedback weights

Lillicrap et al. (2016)
Nature Communications

2. Deep learning with random feedback weights



If $\hat{\mathbf{h}}_i$ is the target activity at layer i, define:

$$\hat{\mathbf{h}}_{i-1} = \mathbf{h}_{i-1} + g_i(\hat{\mathbf{h}}_i) - g_i(\mathbf{h}_i)$$

This ensures that:

$$\mathbf{h}_i = \hat{\mathbf{h}}_i \;\Rightarrow\; \mathbf{h}_{i-1} = \hat{\mathbf{h}}_{i-1}$$

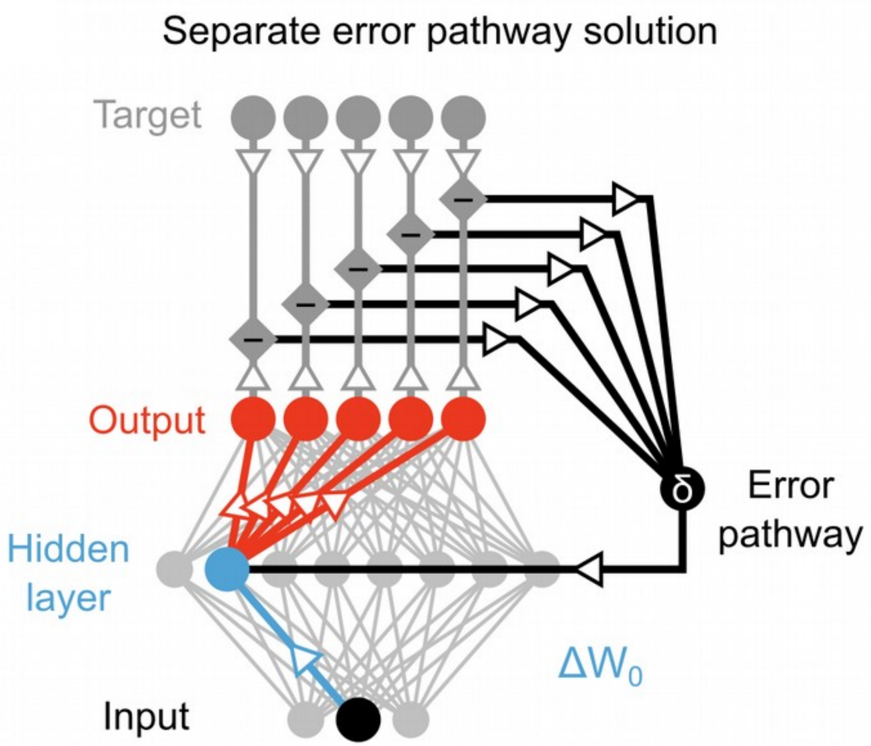
and

$$||\hat{\mathbf{h}}_i - f_i(\hat{\mathbf{h}}_{i-1})||_2^2 < ||\hat{\mathbf{h}}_i - \mathbf{h}_i||_2^2$$

Something similar can be achieved if we use random feedback weights to define local targets

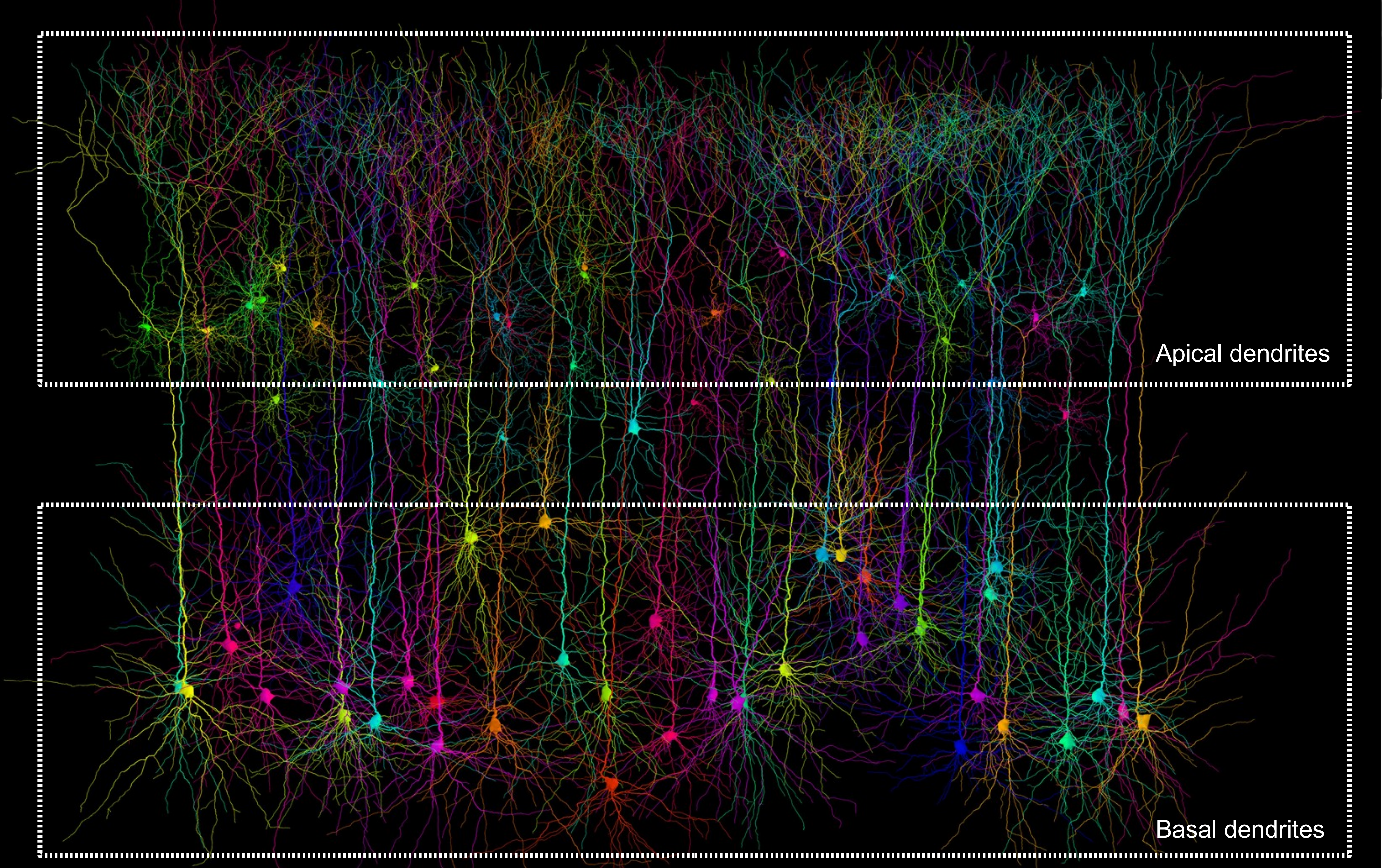
Lee et al. (2015)
Joint Euro. Conf. on M. Learn. and Know. Disc. in Databases

2. Deep learning with random feedback weights



The difficulty with these solutions is that they involve an implicit separate feedback pathway, in order to preserve feedforward calculations

This is possible in the brain, but there is no evidence for anything like it



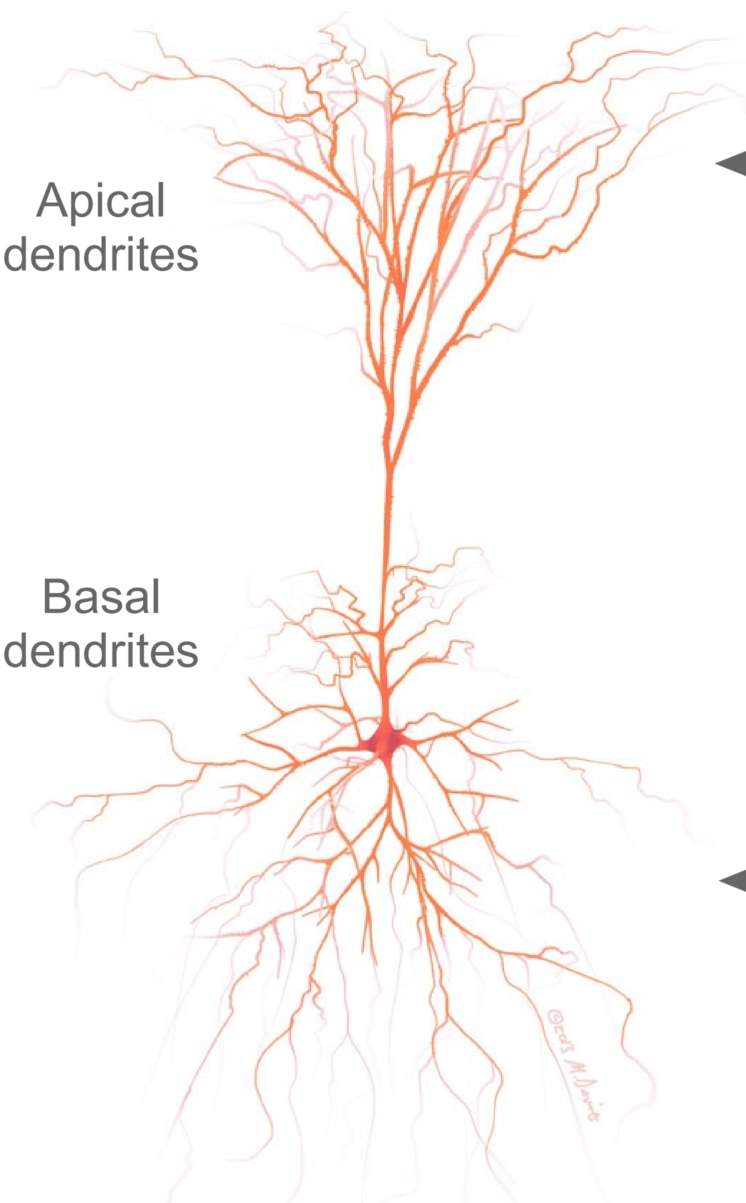
Real neurons in the neocortex have a more complicated structure than the abstract ones used in machine learning.

Perhaps the solution lies here…

3. Pyramidal neuron morphology

Apical dendrites

Basal dendrites

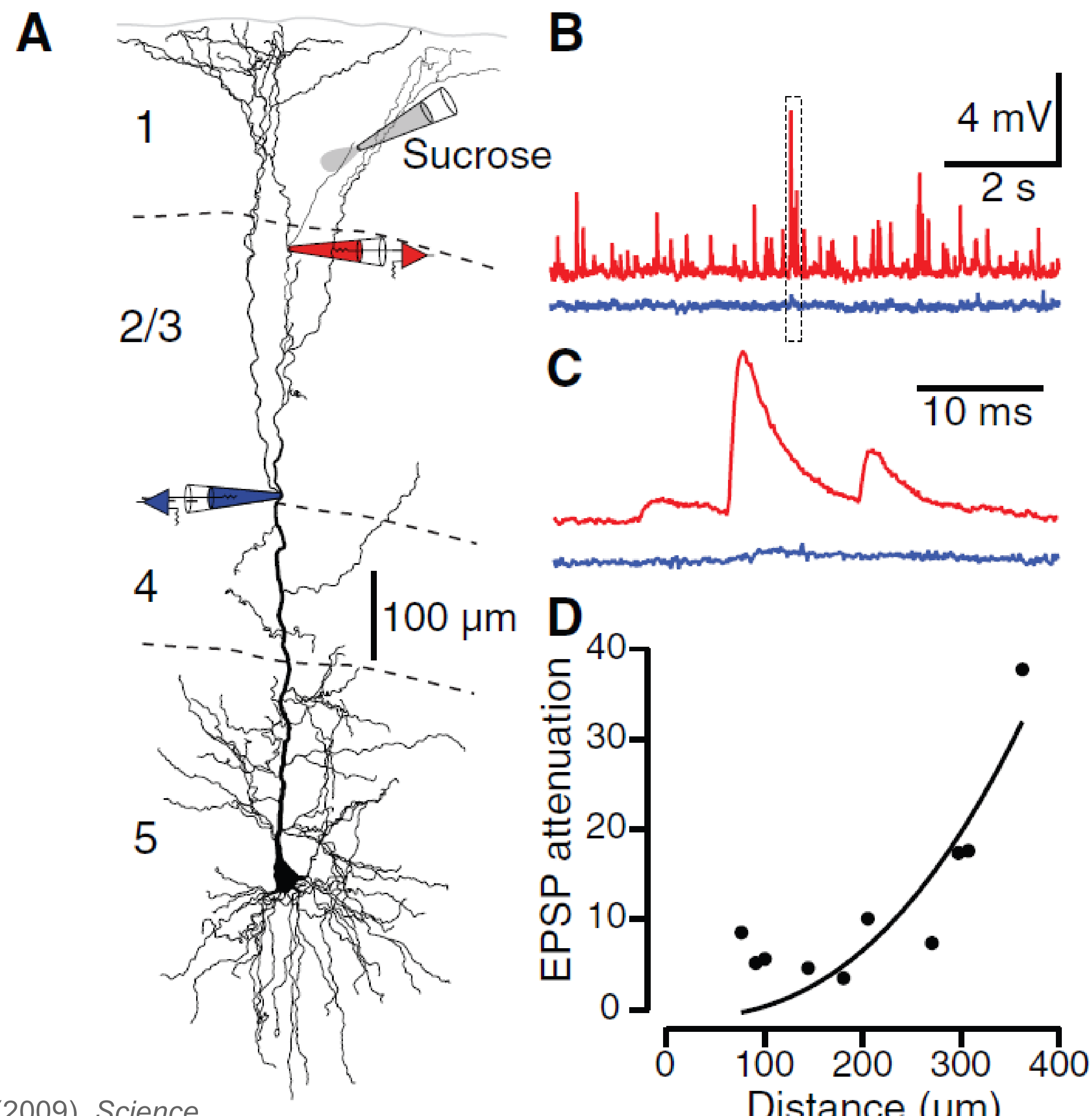


Feedback from other layers and regions of neocortex (and from associative thalamus)

Feedforward thalamic sensory inputs and within layer recurrent connections

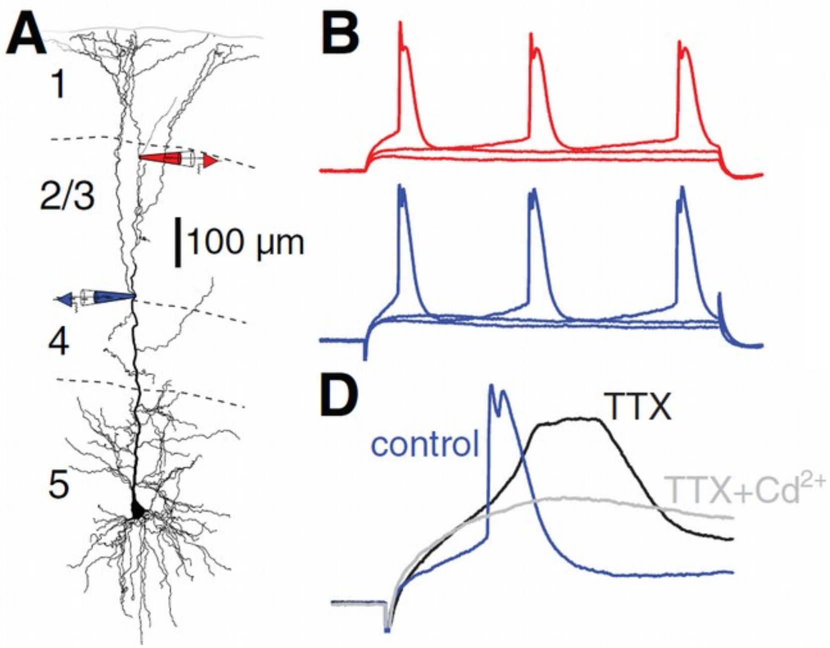
From mdavisstudios.com

3. Pyramidal neuron morphology

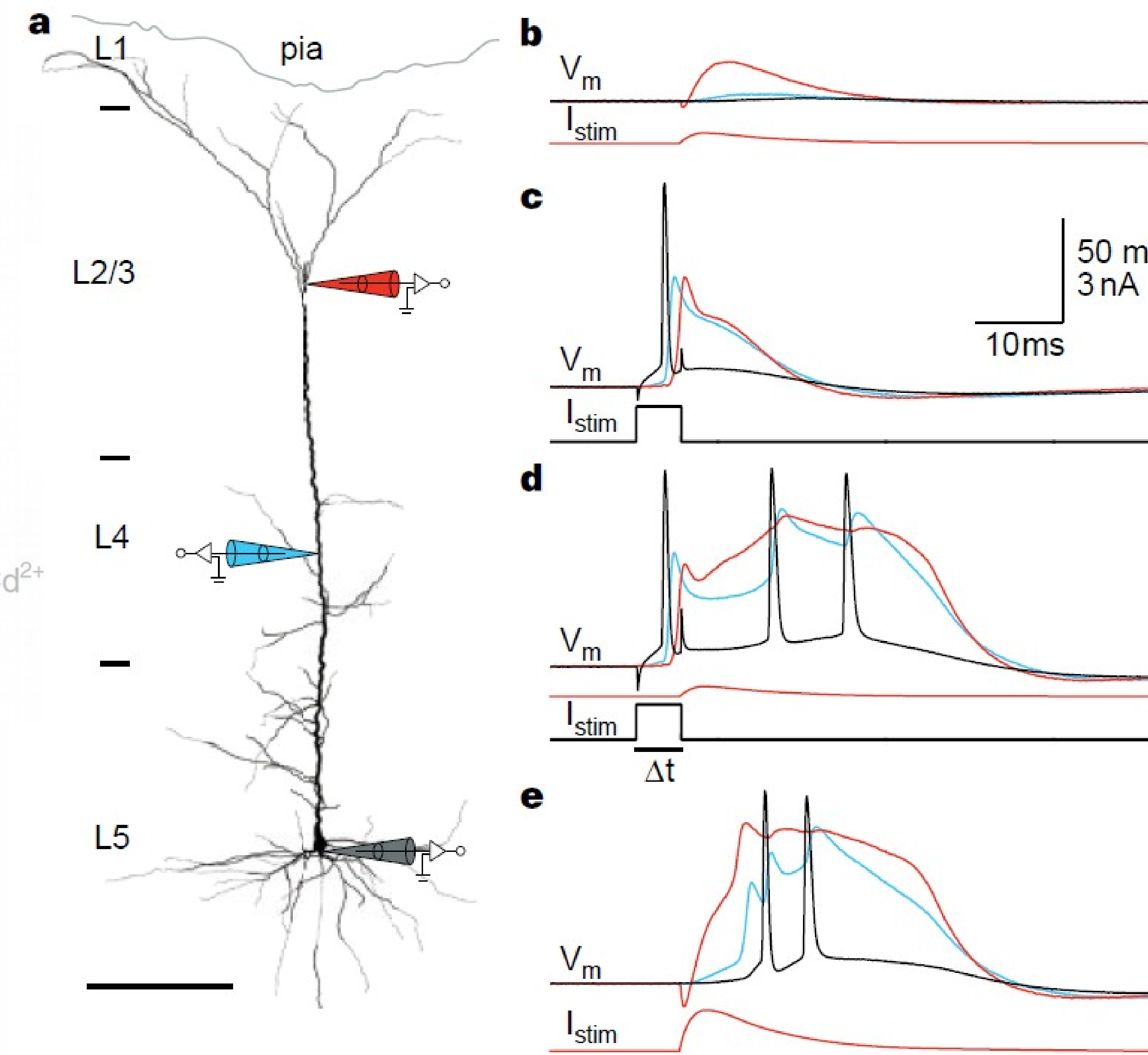


Larkum et al. (2009), *Science*

3. Pyramidal neuron morphology

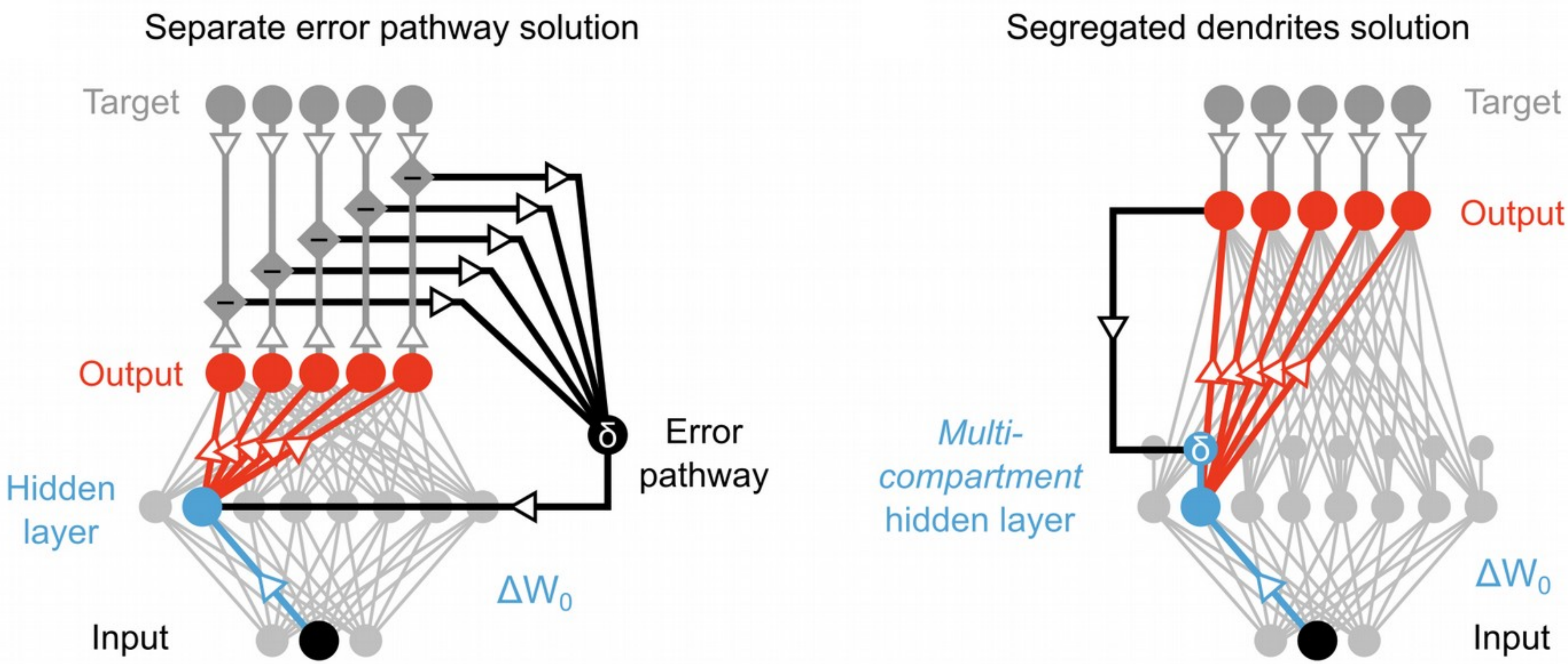


Larkum et al. (2009), *Science*



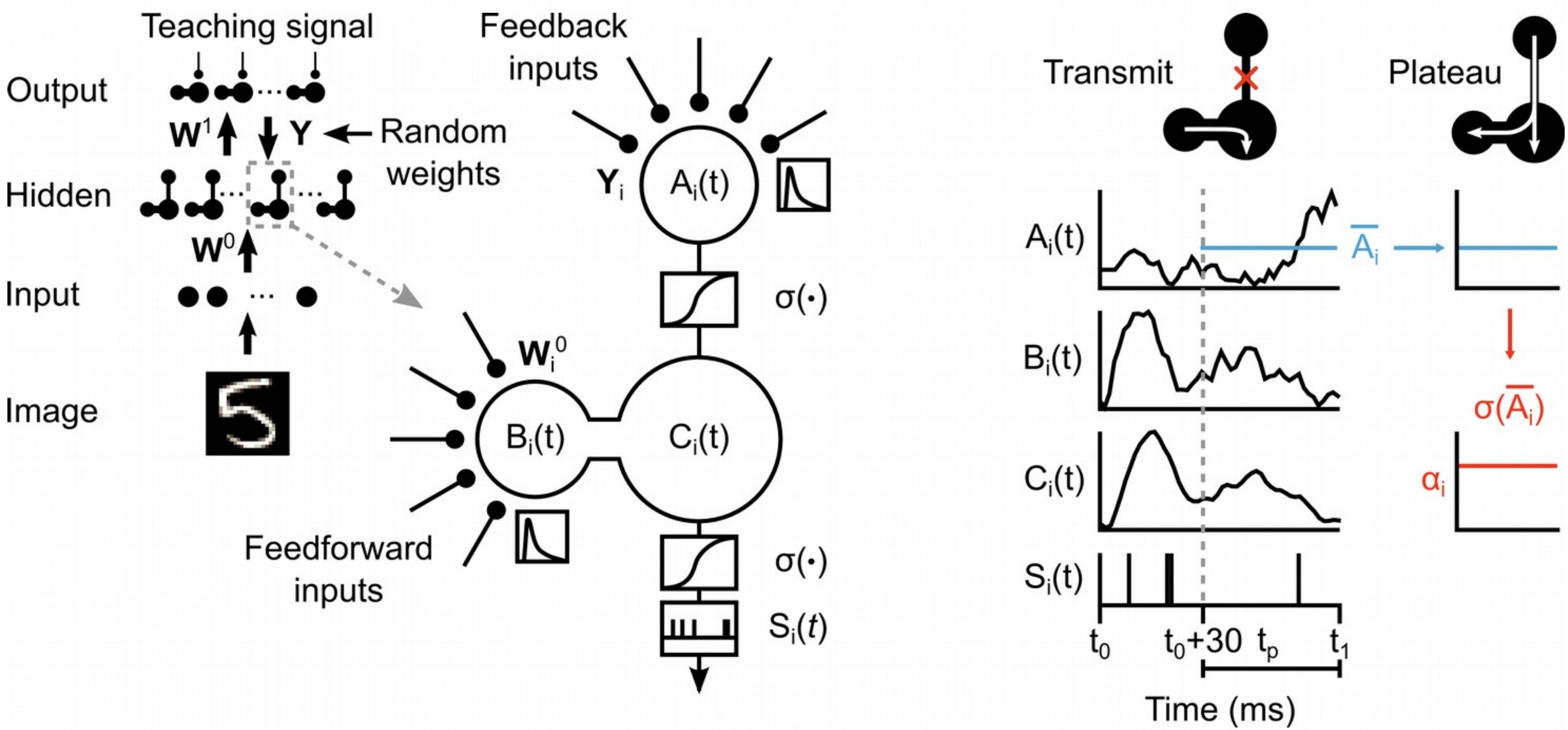
Larkum et al. (1999), *Nature*

3. Pyramidal neuron morphology



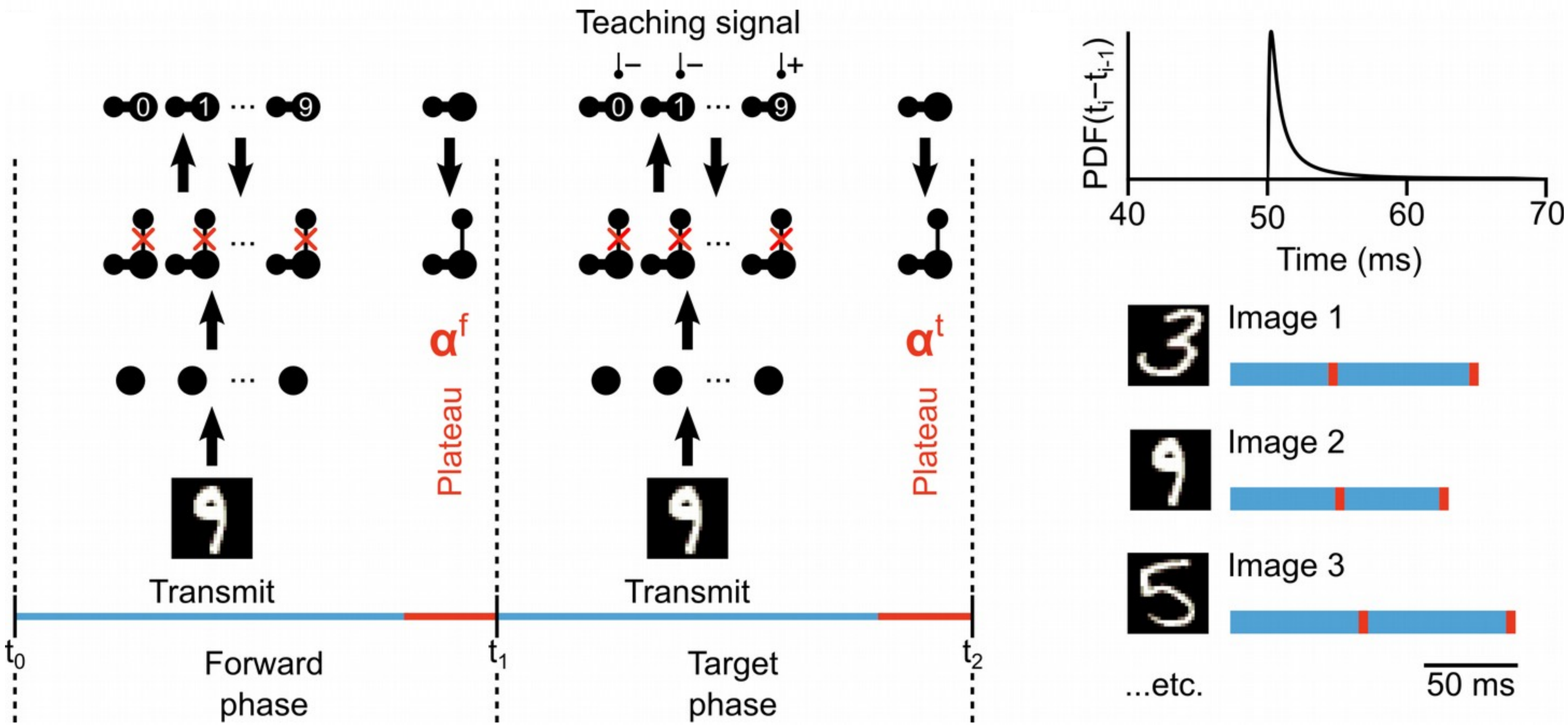
Hypothesis: neocortical neurons segregate feedback in distal apical compartments in order to preserve feedforward calculations while still receiving feedback for credit assignment

4. A multi-compartment deep learning model



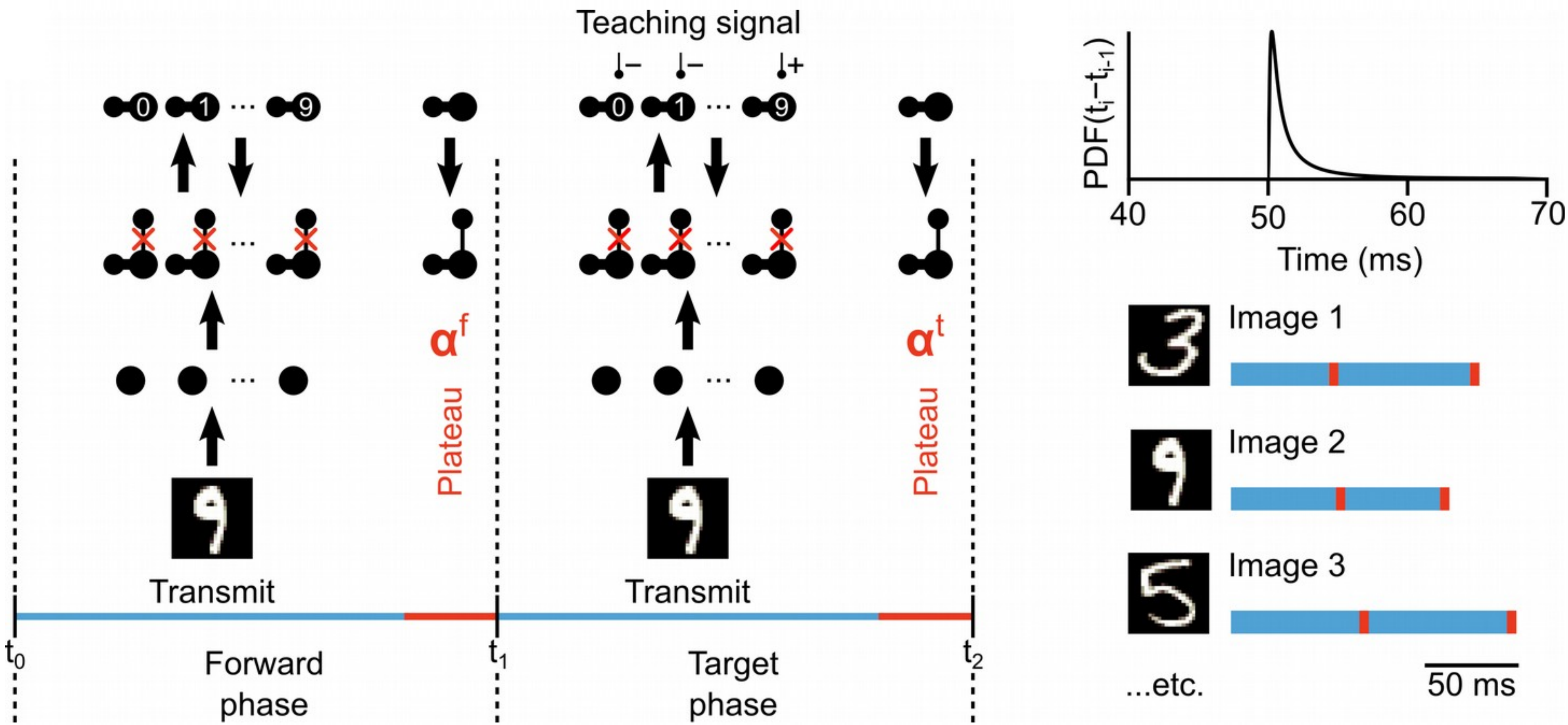
Guerguiev, Lillicrap and Richards (2016) arXiv: 1610.00161

4. A multi-compartment deep learning model



Guerguiev, Lillicrap and Richards (2016) arXiv: 1610.00161

4. A multi-compartment deep learning model



We define a target firing rate for neuron *i* in the hidden layer using the plateau potentials:

$$\hat{\lambda}_i^C = \overline{\lambda_i^C} + \alpha_i^t - \alpha_i^f$$

Average firing rate

4. A multi-compartment deep learning model

We define the target firing rate for the output neurons to be the average firing rate during the target phase:

$$\hat{\lambda}_i^U = \overline{\lambda_i^{Ut}}$$

Output target

$$\hat{\lambda}_i^C = \overline{\lambda_i^C} + \alpha_i^t - \alpha_i^f$$

Hidden target (from previous slide)

To train the network we perform stochastic gradient descent on these loss functions:

$$L^1 = ||\hat{\boldsymbol{\lambda}}^U - \lambda_{max}\sigma(\overline{\boldsymbol{U}^f})||_2^2$$

$$L^0 = ||\hat{\boldsymbol{\lambda}}^C - \lambda_{max}\sigma(\overline{\boldsymbol{C}^f})||_2^2$$

max firing rate

average voltages during forward phase

There are two important things to understand about this approach

(1) The weight updates in the hidden layers use (spatially) local information for credit assignment, i.e. the difference between the plateaus:

$$\Delta \boldsymbol{W}^{0} \propto \boldsymbol{\alpha}^{t} - \boldsymbol{\alpha}^{f}$$

(2) Like Lee et al.'s (2015) Difference Target Propagation algorithm we can prove that our target coordinates learning across layers:

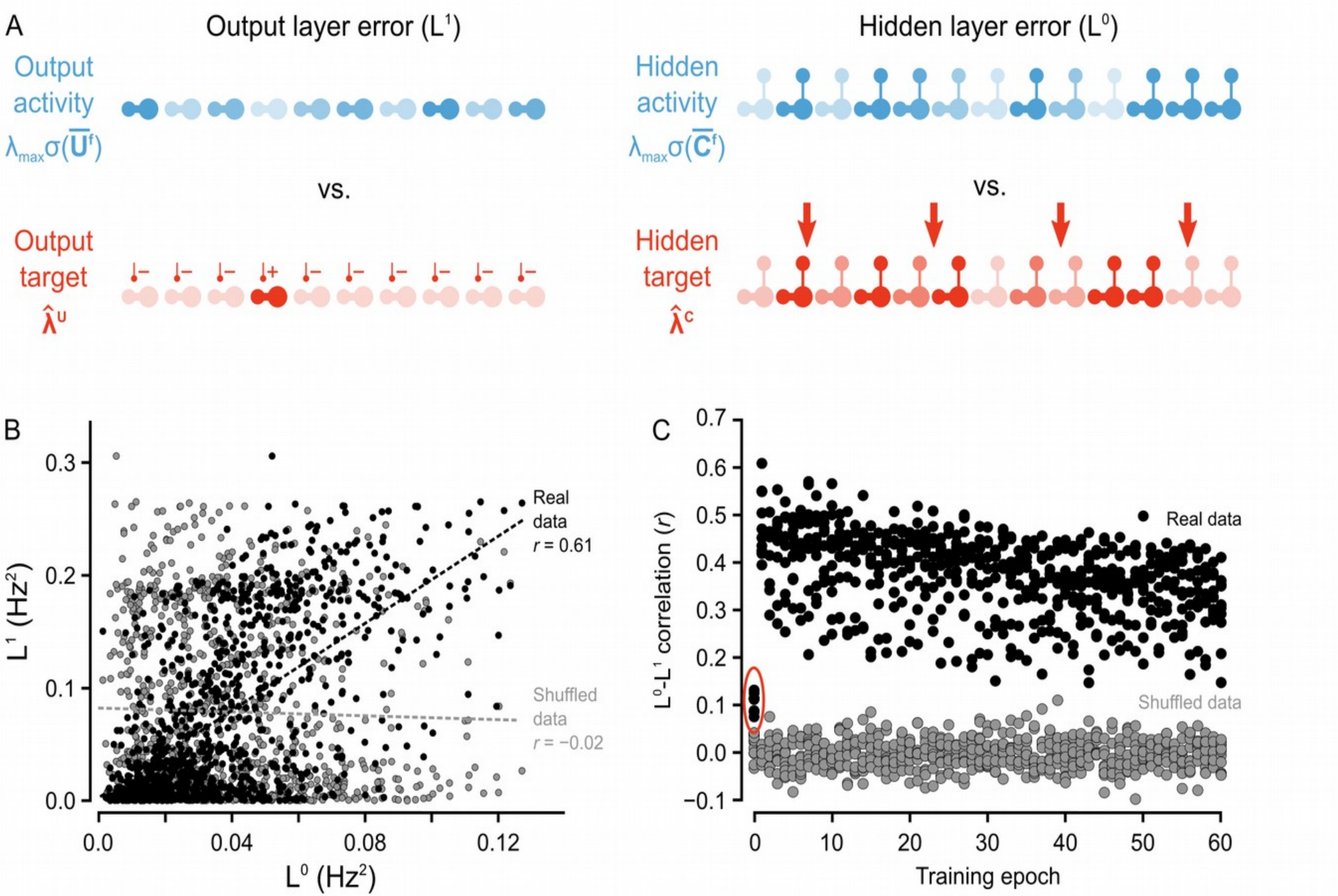
$$\|\hat{\boldsymbol{\lambda}}^{U} - \underline{\lambda_{max}\sigma(k_D \boldsymbol{W}^{1} \hat{\boldsymbol{\lambda}}^{C}}) \|_2^2 < \|\hat{\boldsymbol{\lambda}}^{U} - \underline{\lambda_{max}\sigma(E[\bar{\boldsymbol{U}}^{f}]}) \|_2^2$$

The expected rate of fire of the output neurons if the hidden layer was at its target

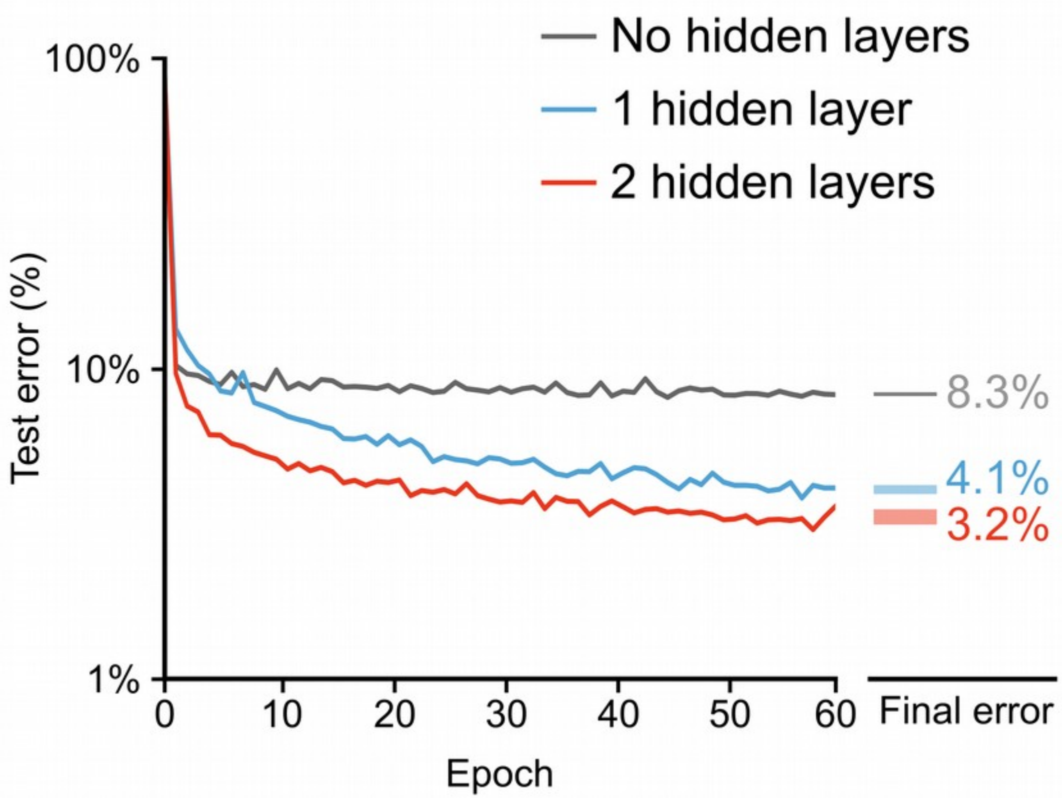
The expected rate of fire of the output neurons without the teaching signal

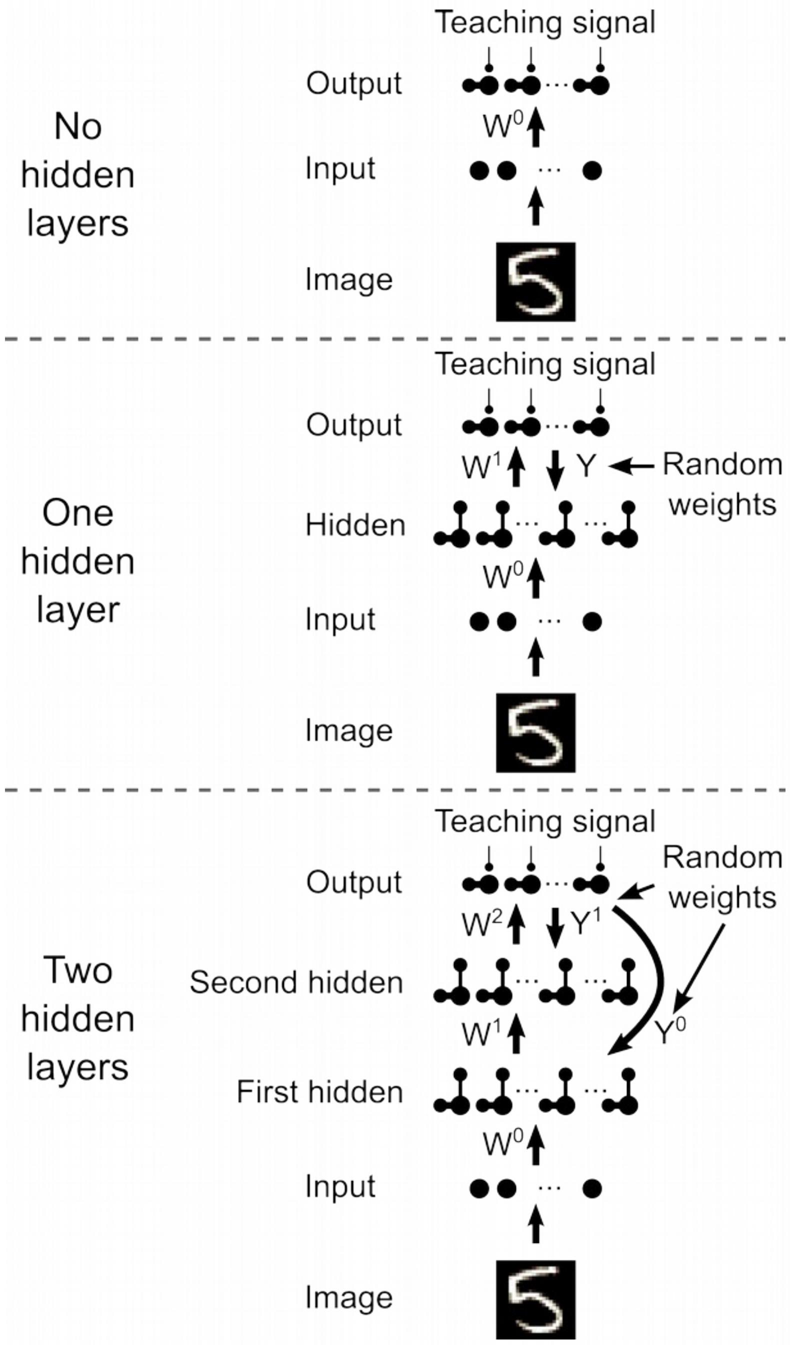
4. A multi-compartment deep learning model

Support of proof: when the hidden layer has a low error, so does the output layer



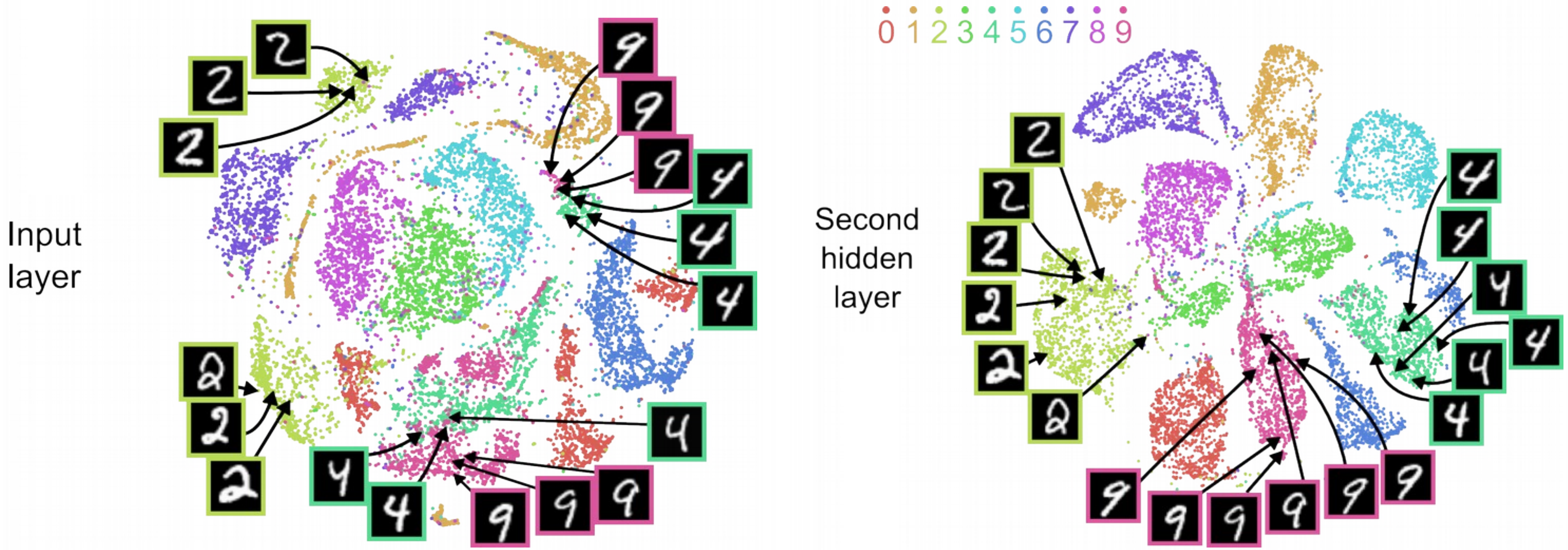
Our model exhibits deep learning (light), because **adding hidden layers improves performance**





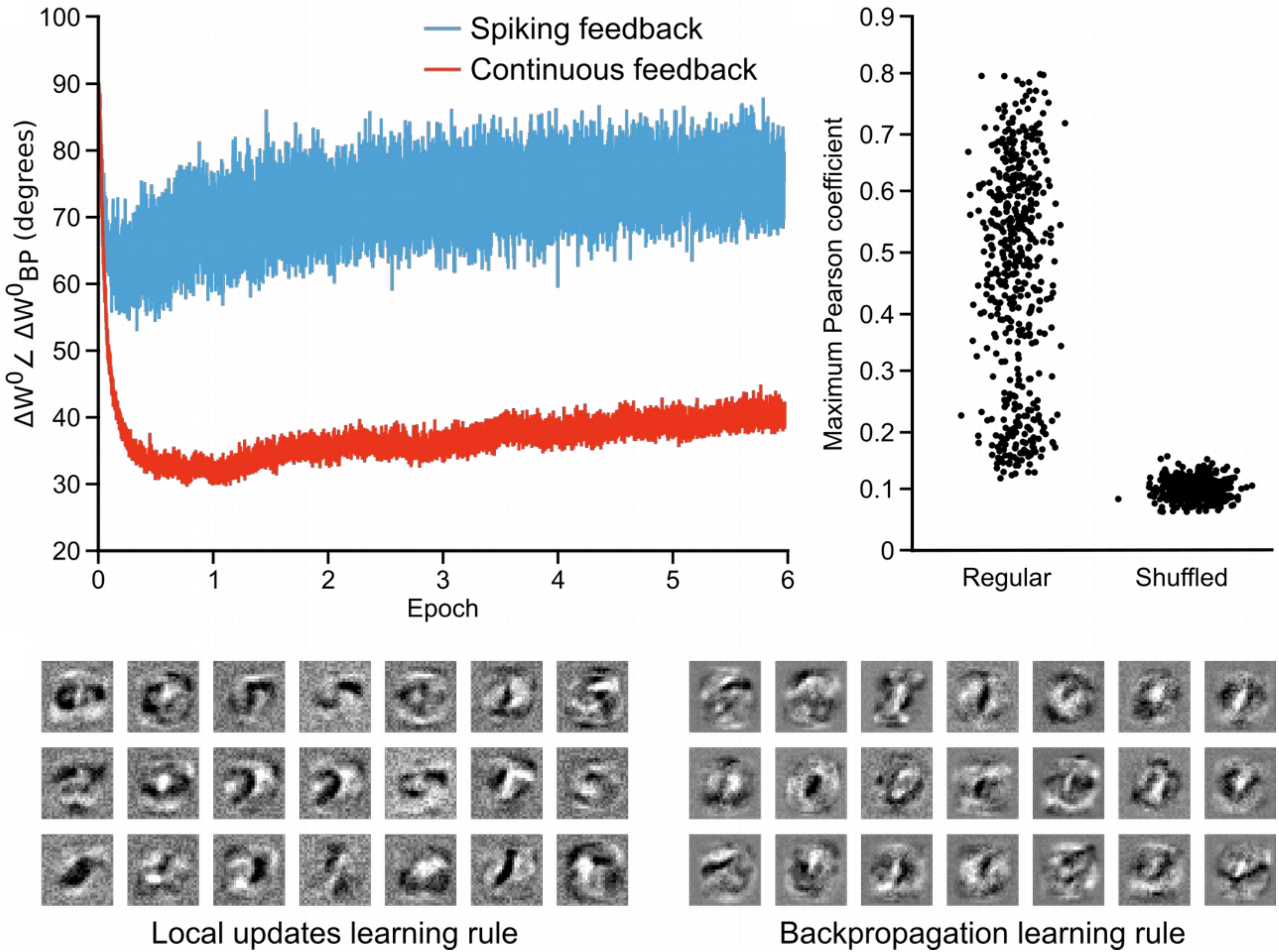
4. A multi-compartment deep learning model

The network learns abstract, category-based representations in its upper layers



Results of t-SNE applied to network activity

5. Relationship to feedback alignment



5. Relationship to feedback alignment

The coordination between the hidden layer and the output layer error was based on this inequality:

$$\|\hat{\boldsymbol{\lambda}}^{U} - \lambda_{max}\sigma(k_D \boldsymbol{W}^1 \hat{\boldsymbol{\lambda}}^{C})\|_2^2 < \|\hat{\boldsymbol{\lambda}}^{U} - \lambda_{max}\sigma(E[\bar{\boldsymbol{U}}^f])\|_2^2$$

The proof for this has two conditions:

(1) The output error is low (easy)

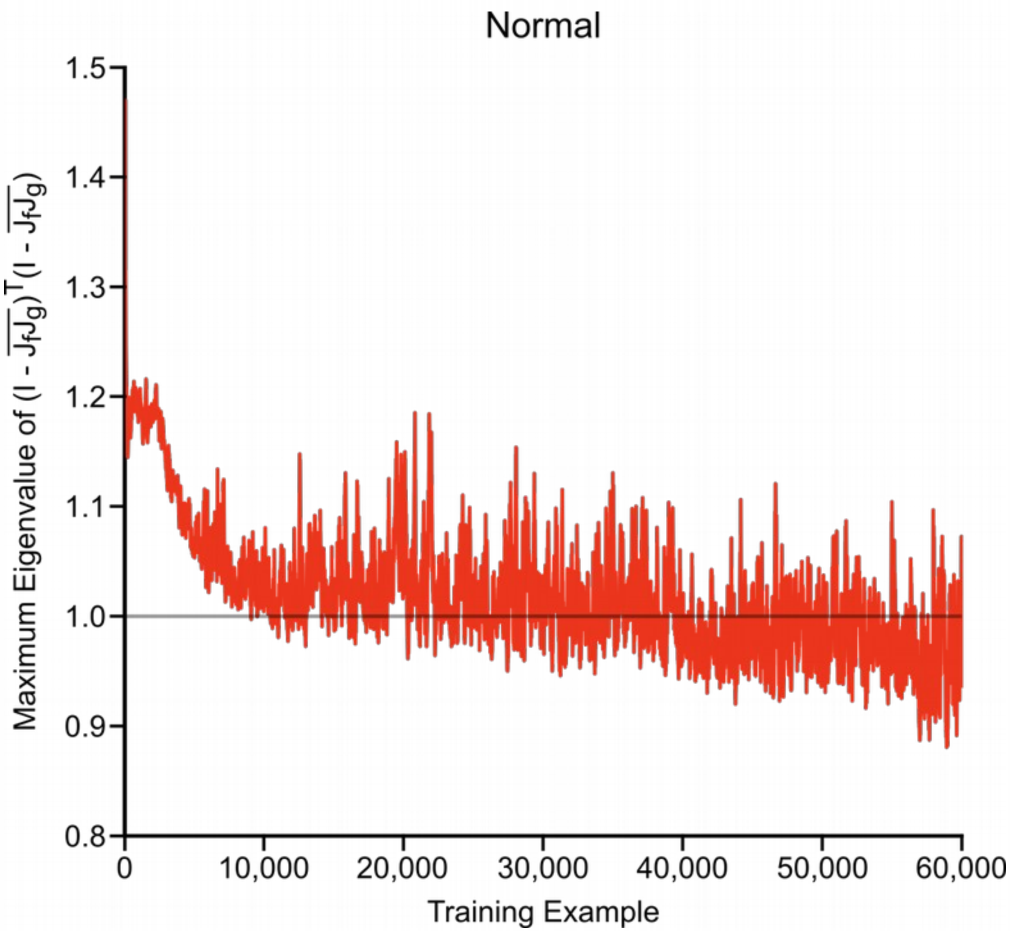
(2) The maximum eigenvalue of this product is less than 1:

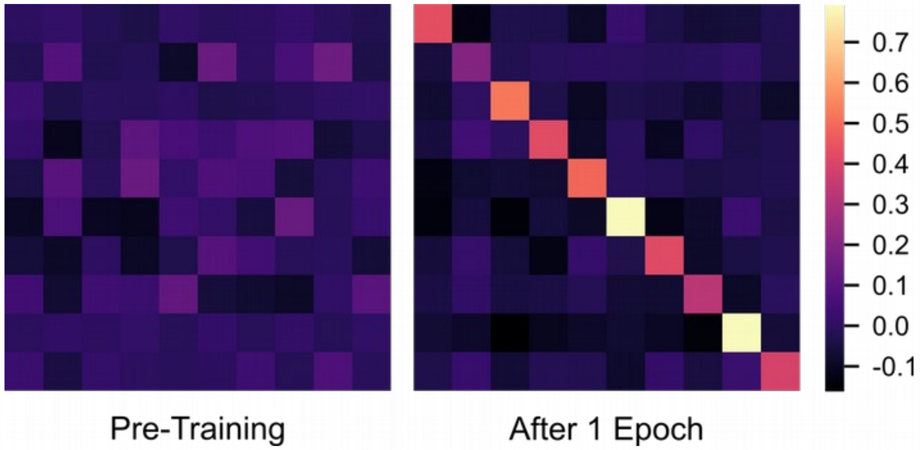
$$(\boldsymbol{I} - \boldsymbol{J_f}\boldsymbol{J_g})(\boldsymbol{I} - \boldsymbol{J_f}\boldsymbol{J_g})^T$$

where J_f and J_g are the forwards and backwards Jacobians, respectively

This requires alignment of the forward and backward weights

But, unlike Lee et al. (2015), here we don't do any learning on the backwards weights, so what happens?



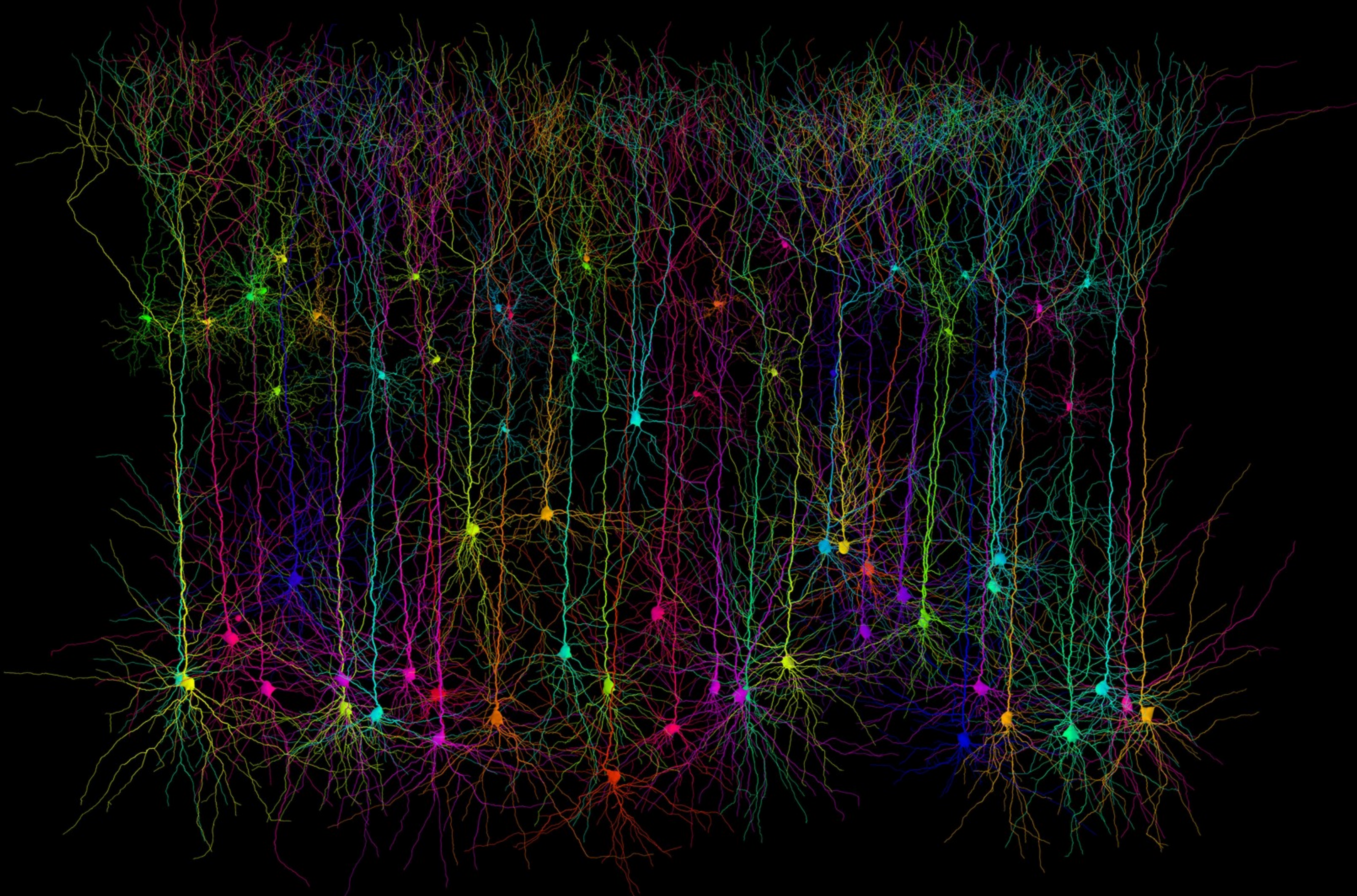


$J_f J_g$

We have demonstrated that we can achieve deep learning (light) in a simulation of neocortical pyramidal neurons

The details of our model are undoubtedly not accurate, but we believe that there is an important principle that our model illustrates that can inform our understanding of biological deep learning

One way to accomplish credit assignment without separate forward and backward pathways is to use electrotonically segregated dendritic compartments *like those observed in the neocortex*



My wager: the complex, beautiful morphology of pyramidal neurons is an important component of nature's solution to deep learning

Jordan Guerguiev
PhD Candidate

Timothy Lillicrap
Google DeepMind

Financial support

NSERC CRSNG

HFSP

Google

CIFAR | ICRA
CANADIAN INSTITUTE FOR ADVANCED RESEARCH | INSTITUT CANADIEN DE RECHERCHES AVANCÉES

LiNC LAB
linclab.org

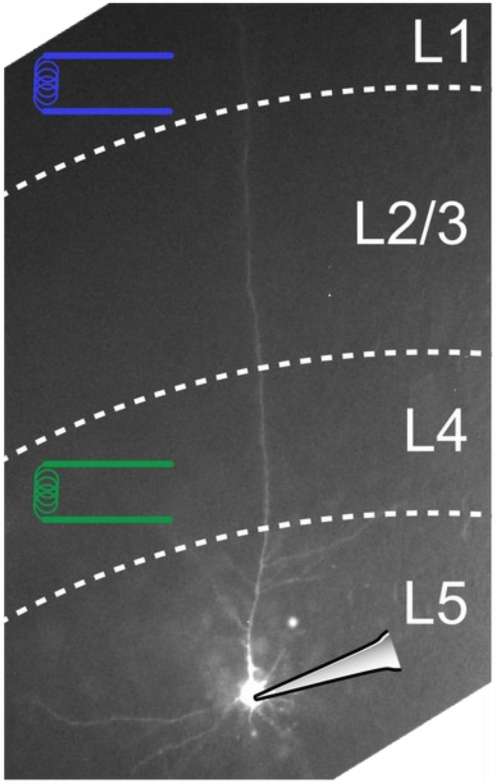
UNIVERSITY OF TORONTO SCARBOROUGH

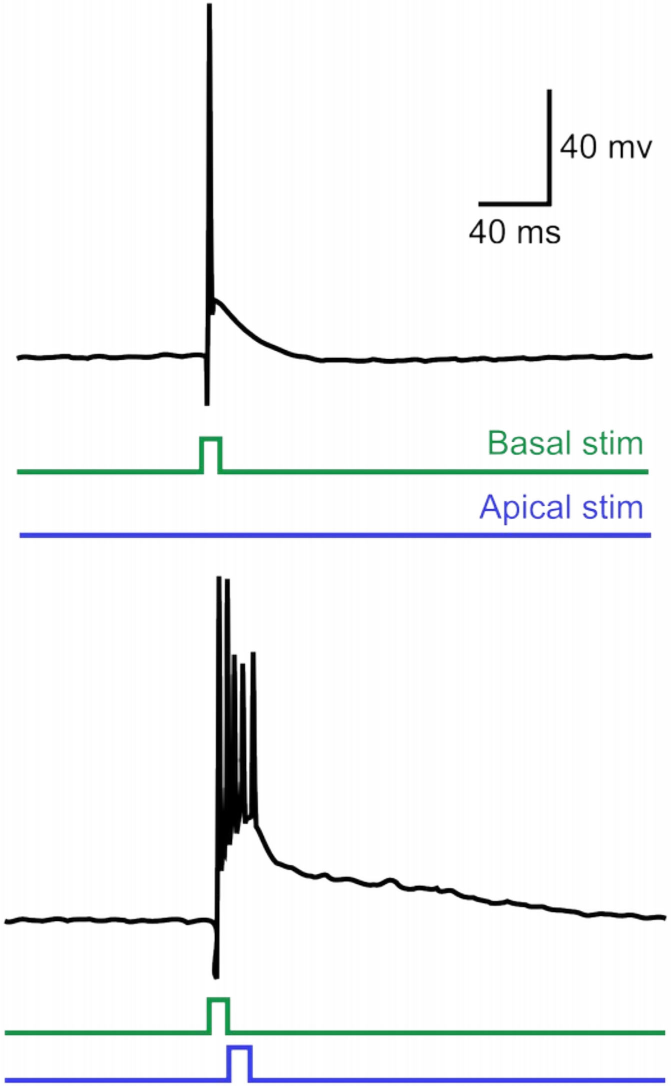
7. Next steps: experimental explorations

We are beginning experiments to determine whether or not plateau potentials driven by apical inputs can modulate the ***sign*** of feedforward plasticity

$$\Delta W^0 \propto \alpha^t - \alpha^f$$

Matthew Tran
PhD Candidate





7. Next steps: experimental explorations

We are beginning experiments to determine whether or not plateau potentials driven by apical inputs can modulate the ***sign*** of feedforward plasticity

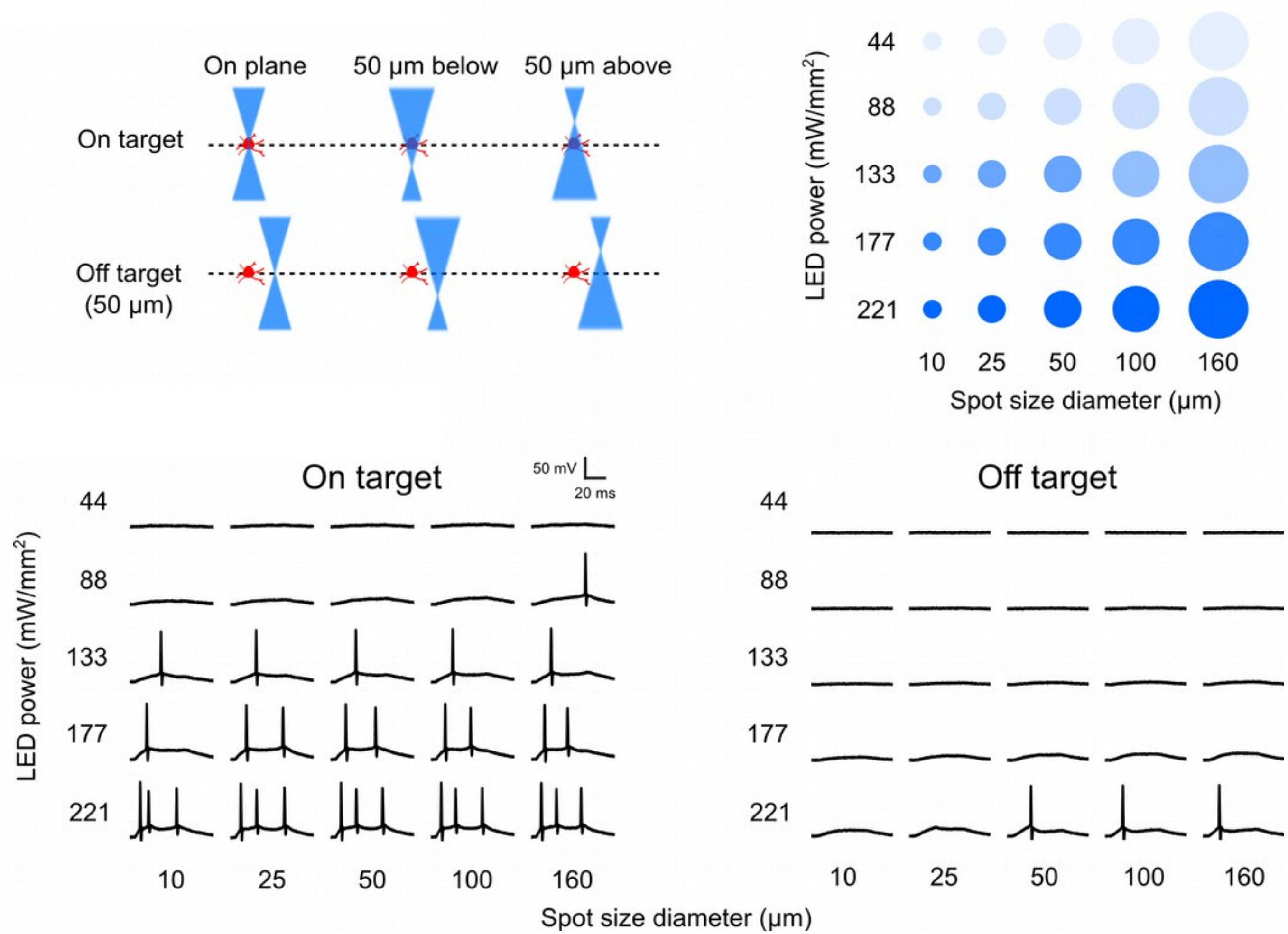
$$\Delta W^0 \propto \alpha^t - \alpha^f$$

We can use optogenetics paired with a digital multimirror device DMD to ensure we restrict ourselves to top-down vs. bottom-up input

Matthew Tran
PhD Candidate

We are beginning experiments to determine whether or not plateau potentials driven by apical inputs can modulate the ***sign*** of feedforward plasticity

$$\Delta W^0 \propto \alpha^t - \alpha^f$$



Matthew Tran
PhD Candidate

7. Next steps: experimental explorations

We are beginning experiments to determine whether or not plateau potentials driven by apical inputs can modulate the ***sign*** of feedforward plasticity

$$\Delta \boldsymbol{W}^0 \propto \boldsymbol{\alpha}^t - \boldsymbol{\alpha}^f$$

We can use optogenetics paired with a digital multimirror device DMD to ensure we restrict ourselves to top-down vs. bottom-up input

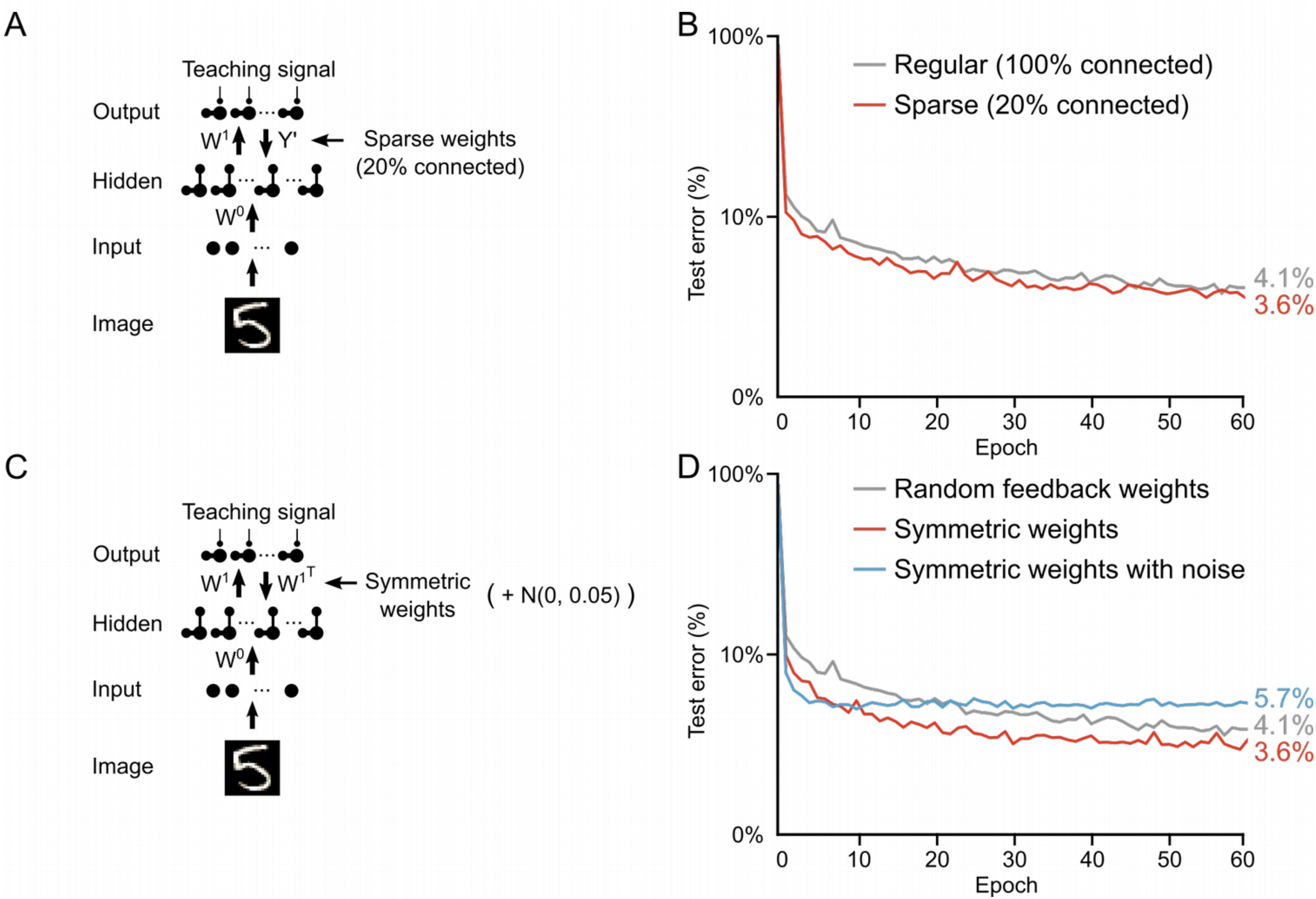
(1) Infect downstream circuits that provide top-down inputs to sensory area (e.g. V1 or S1) to control top-down projections

(2) Infect Scnn1a-Tg3-Cre mice with floxed, red-shifted opsin

(3) Use DMD and light colour to restrict activation to top-down apical projections or bottom-up basal projections

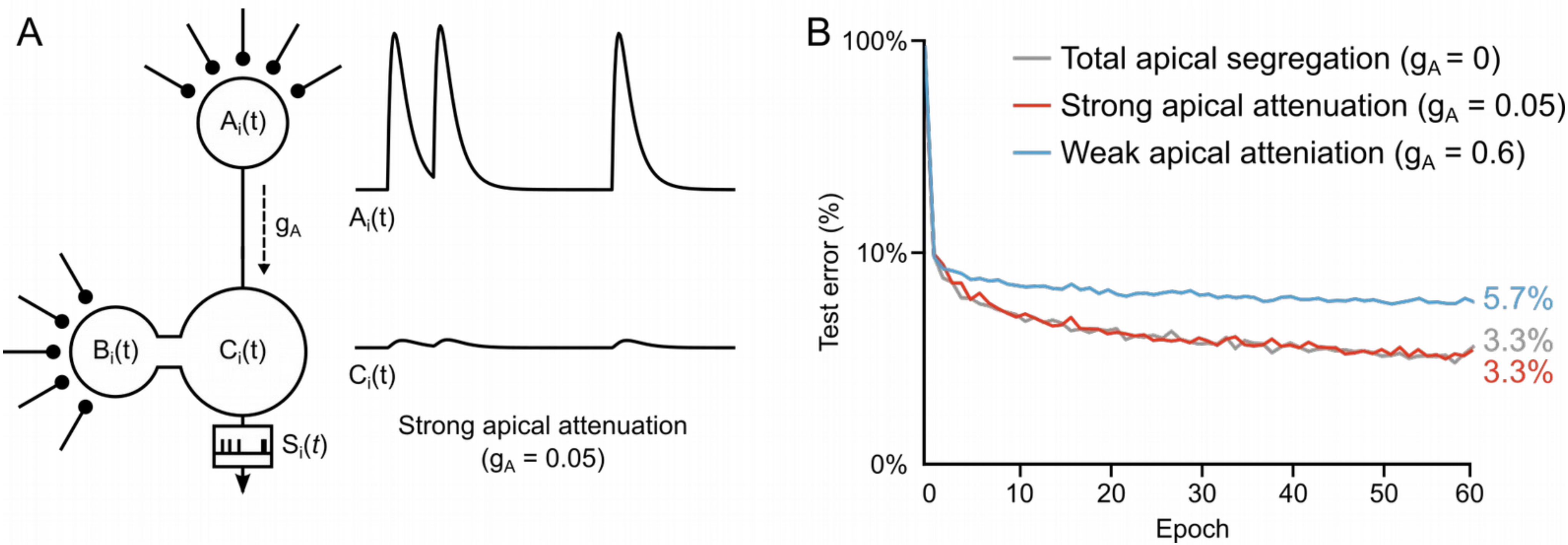
Matthew Tran
PhD Candidate

8. Feedback requirements



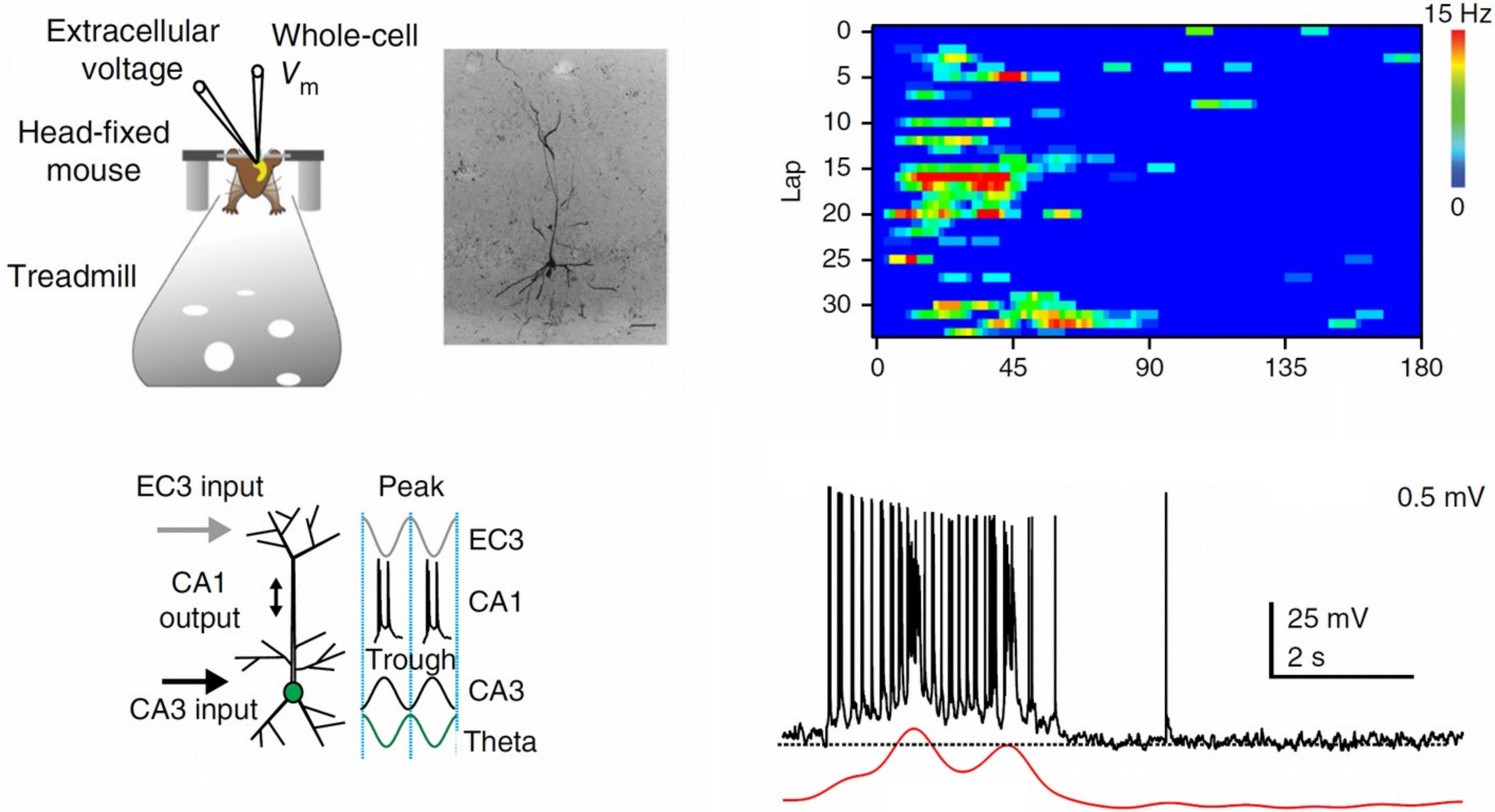
8. Feedback requirements

Segregation of the apical dendrites is critical for learning, but biologically realistic levels of attenuation are fine



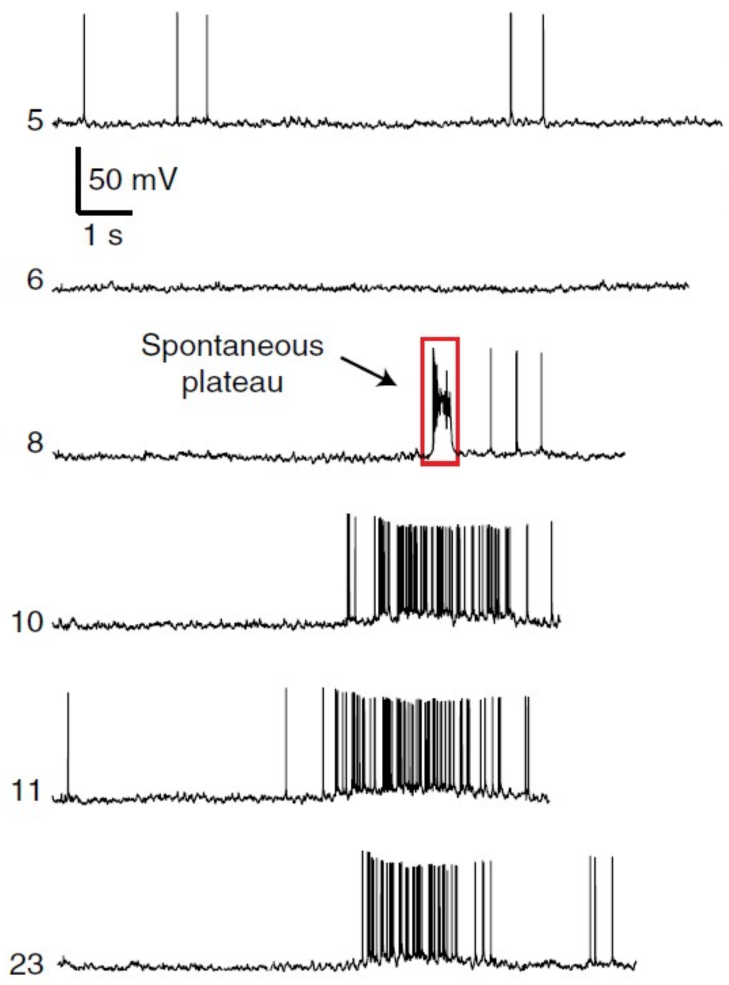
9. Next steps: experimental explorations

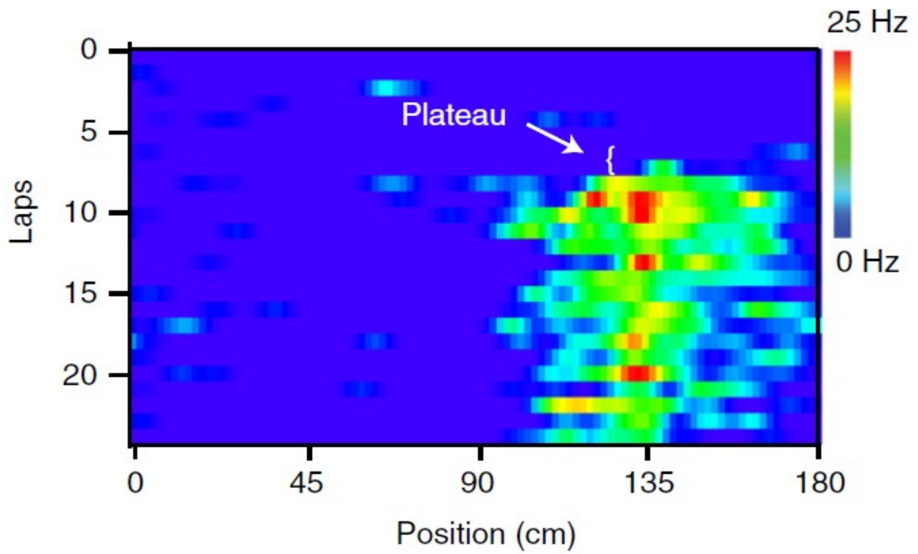
There is some evidence that apical dendrite inputs guide basal dendrite weight updates



Bittner et al. (2015), *Nature Neuroscience*

There is some evidence that apical dendrite inputs guide basal dendrite weight updates



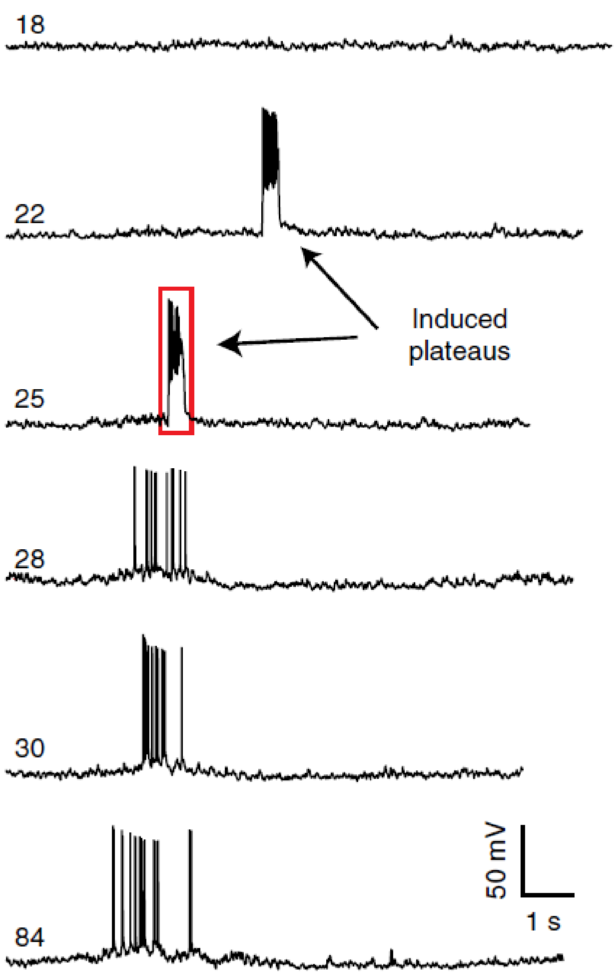


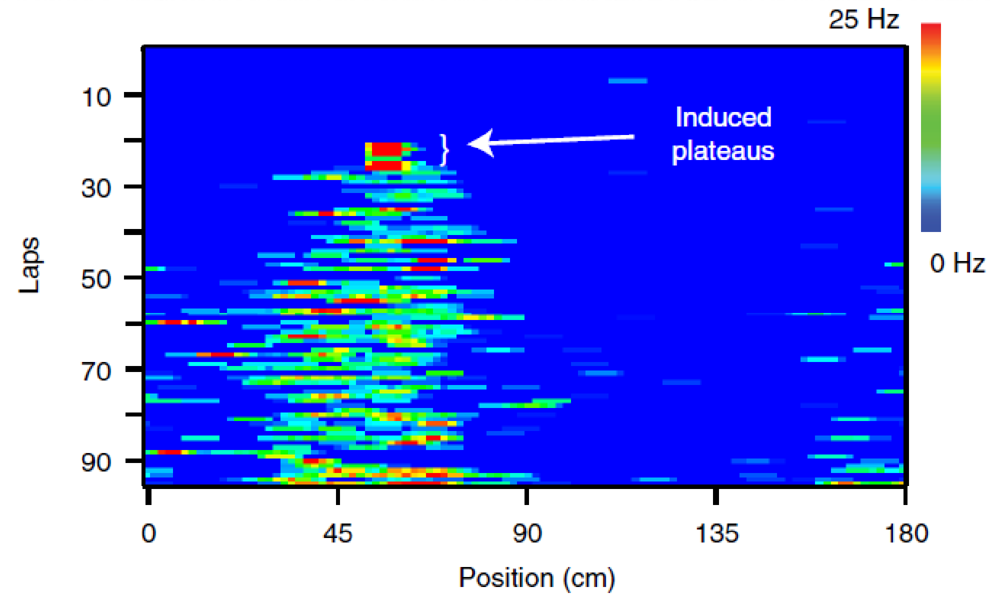
Plateau potentials driven by apical inputs determine where place fields form

Bittner et al. (2015), *Nature Neuroscience*

9. Next steps: experimental explorations

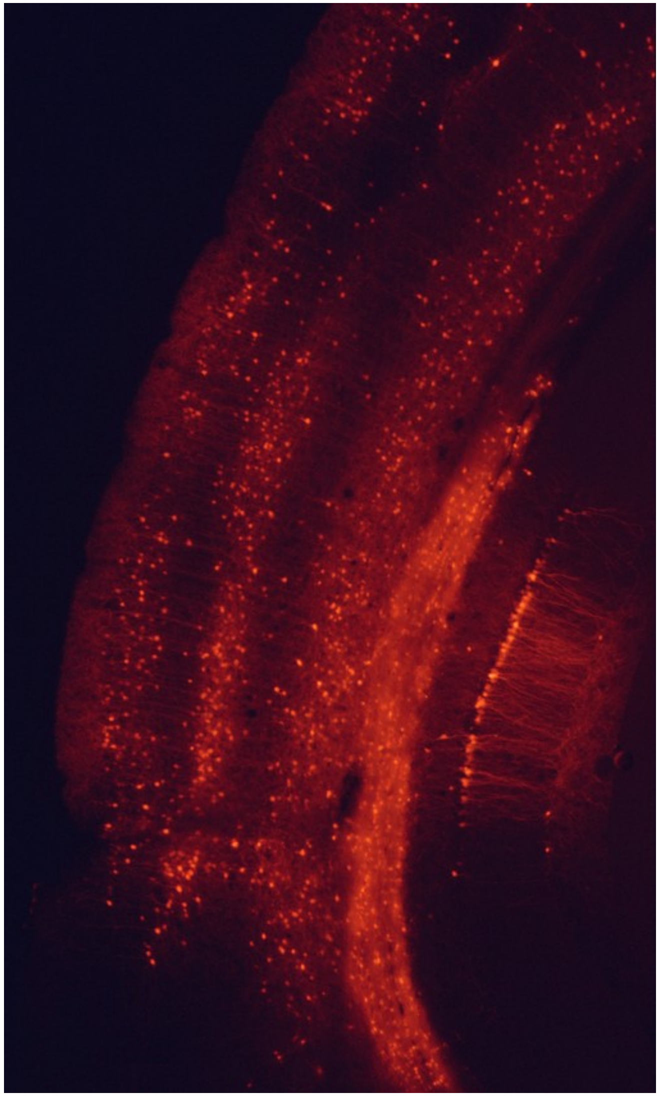
There is some evidence that apical dendrite inputs guide basal dendrite weight updates

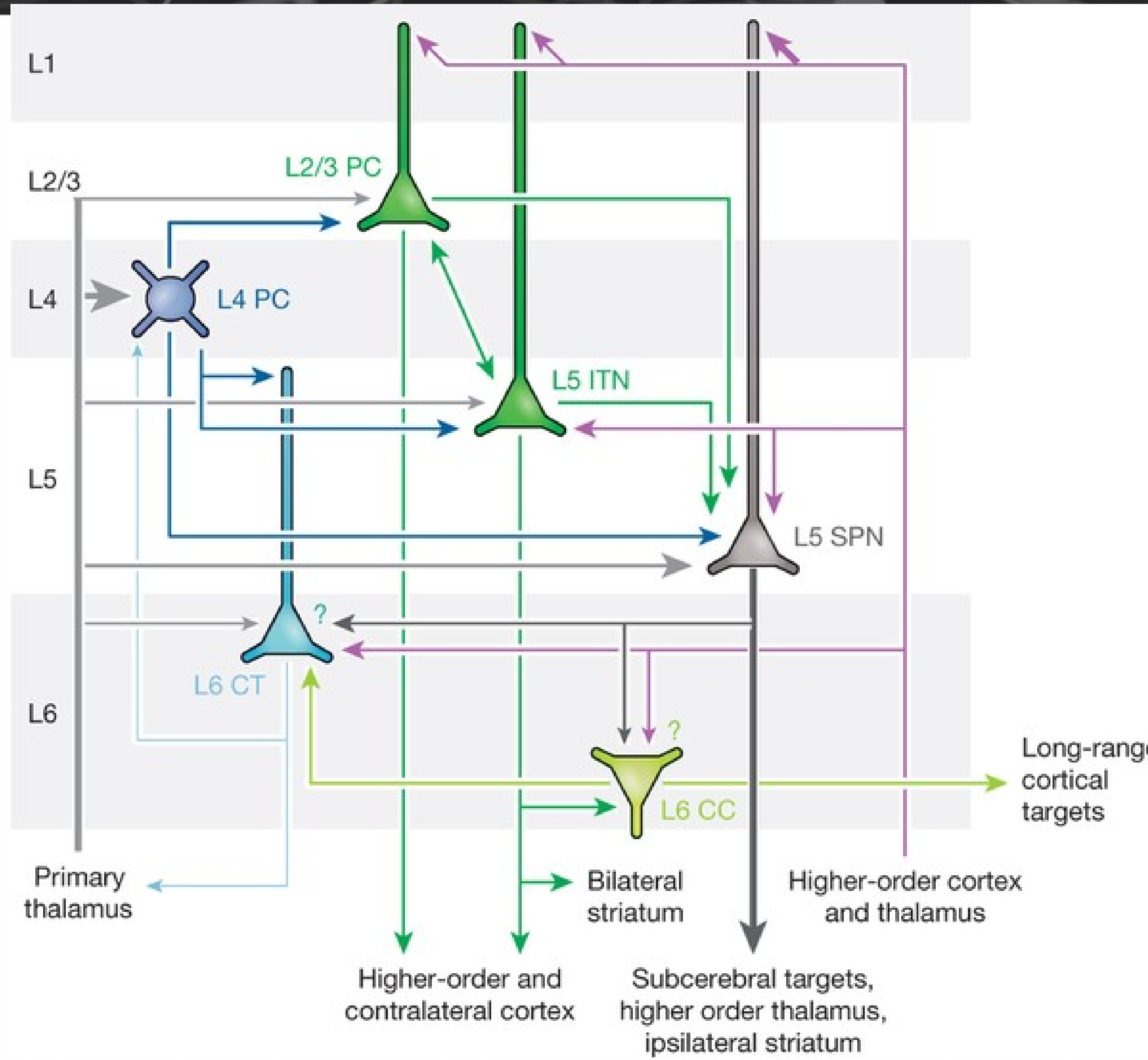




Plateau potentials driven by apical inputs determine where place fields form

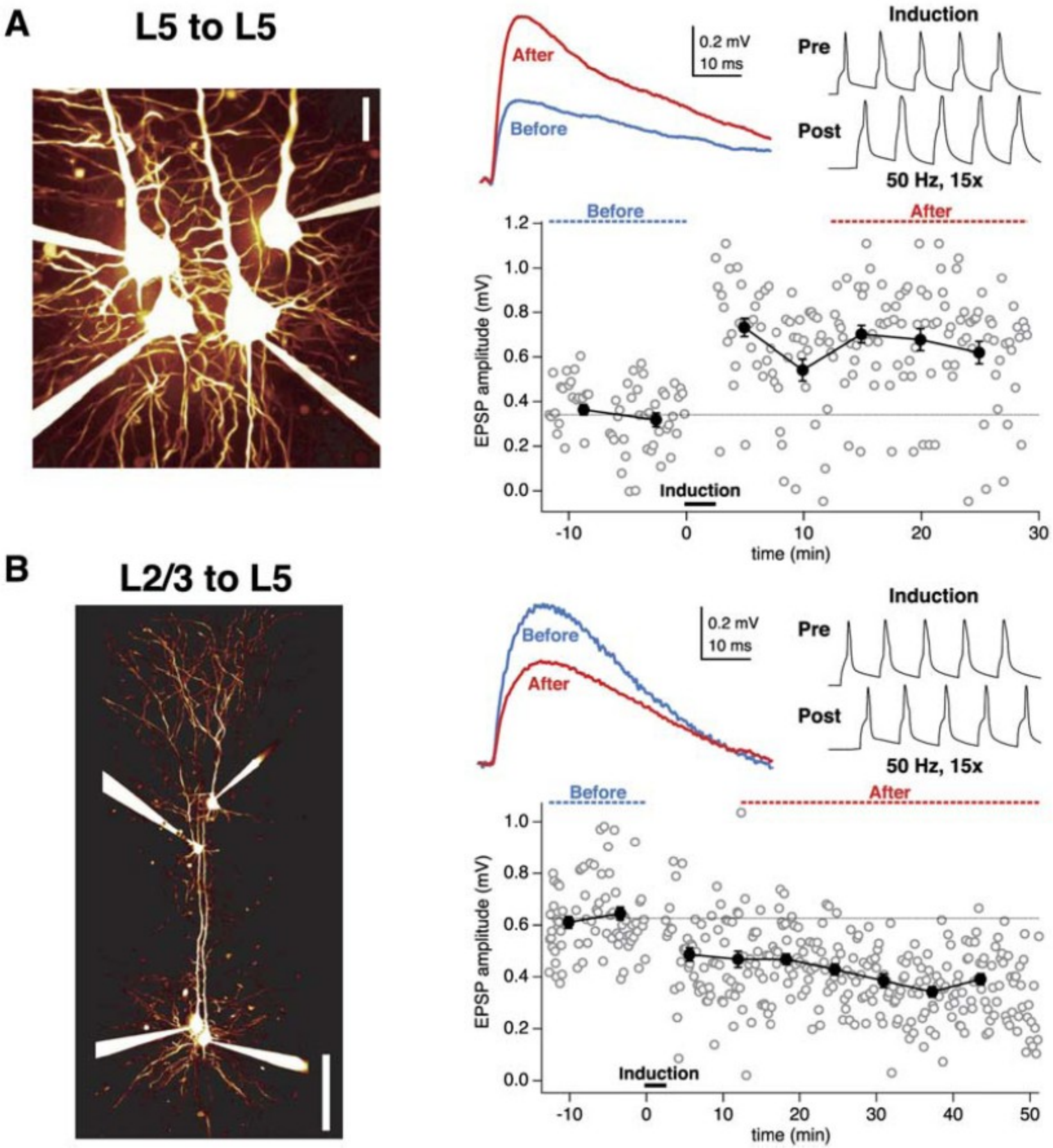
Bittner et al. (2015), *Nature Neuroscience*

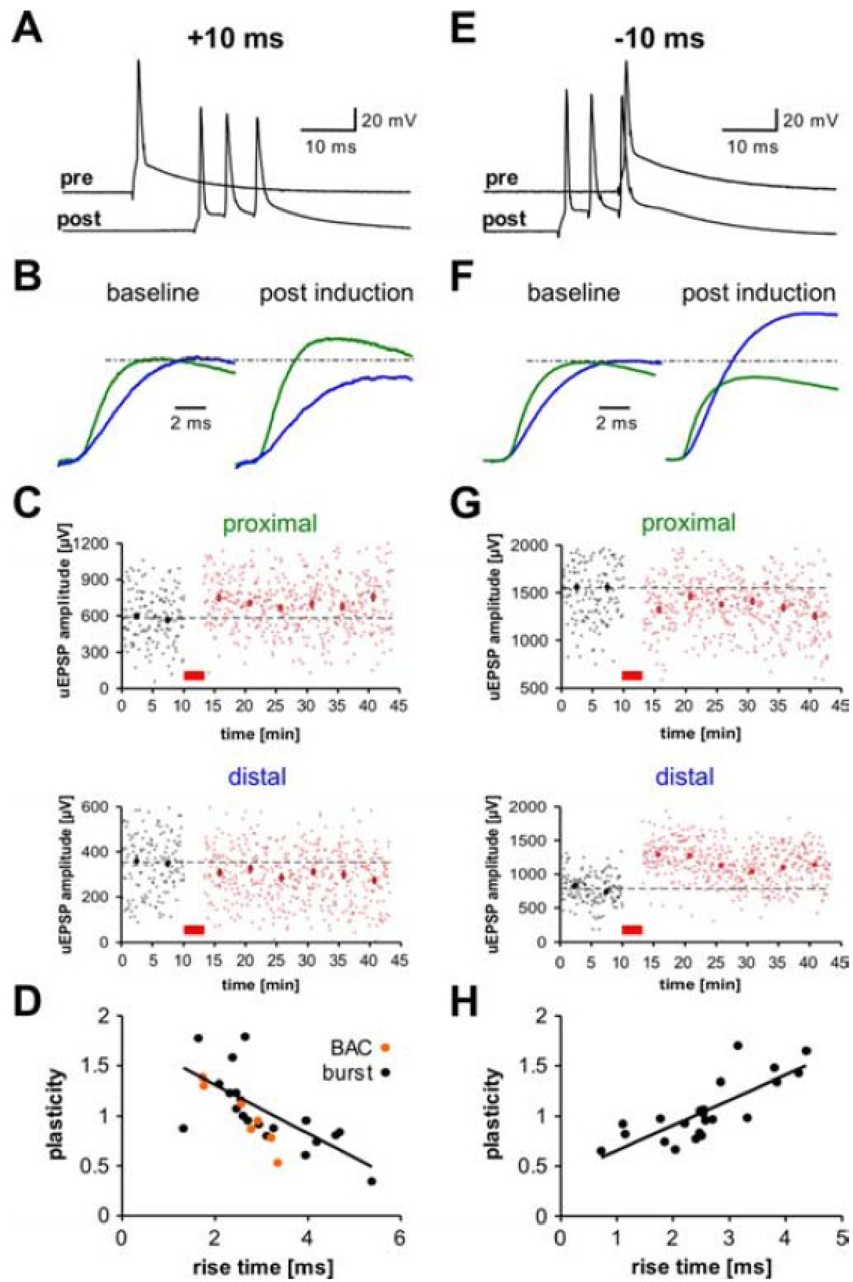




Harris & Mrsic-Flogel (2013)

Synaptic plasticity rules at apical and basal dendrites appear to be different

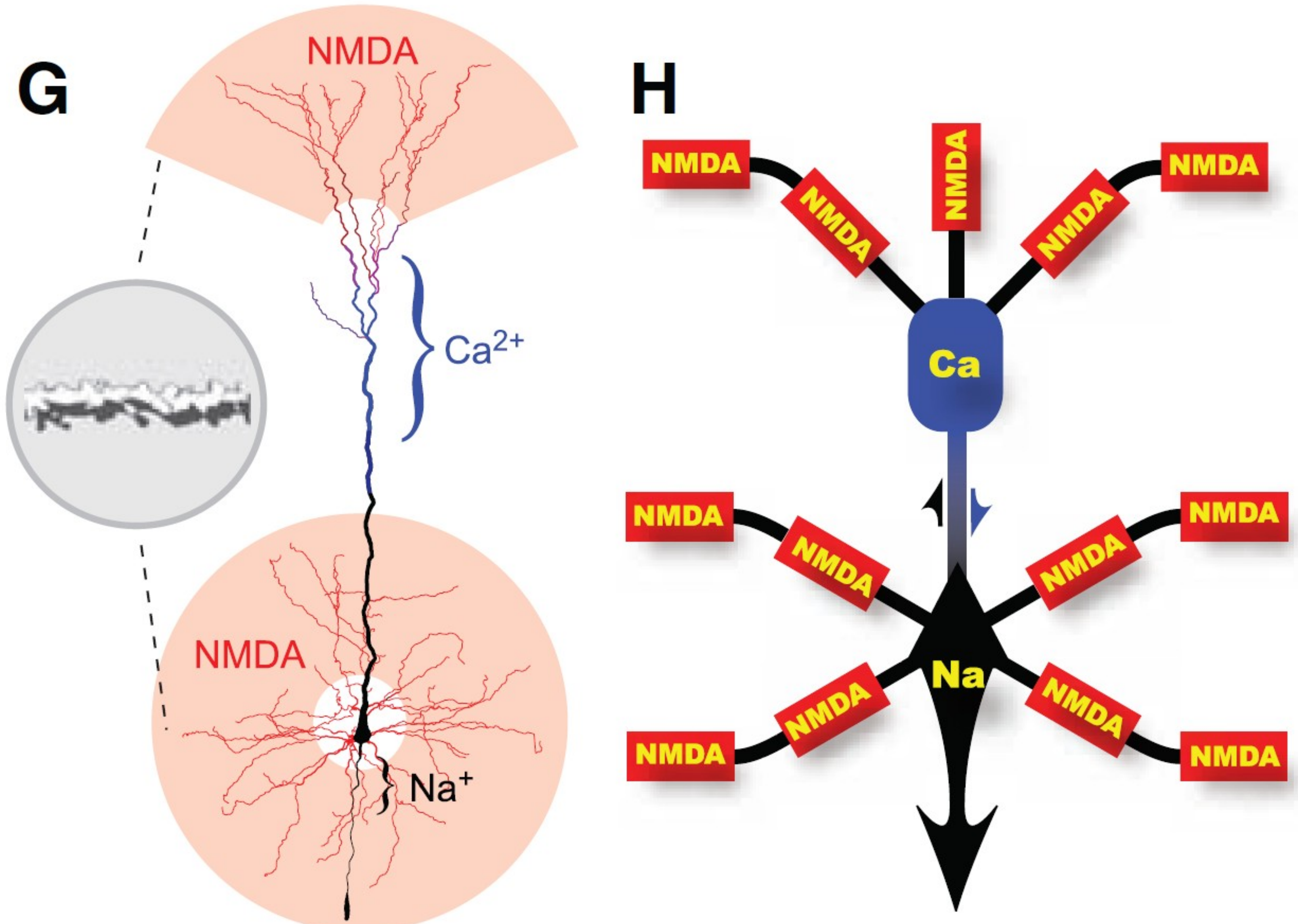




Sjöström & Häusser (2006), *Neuron*

Letzkus et al. (2006), *J. Neurosci.*

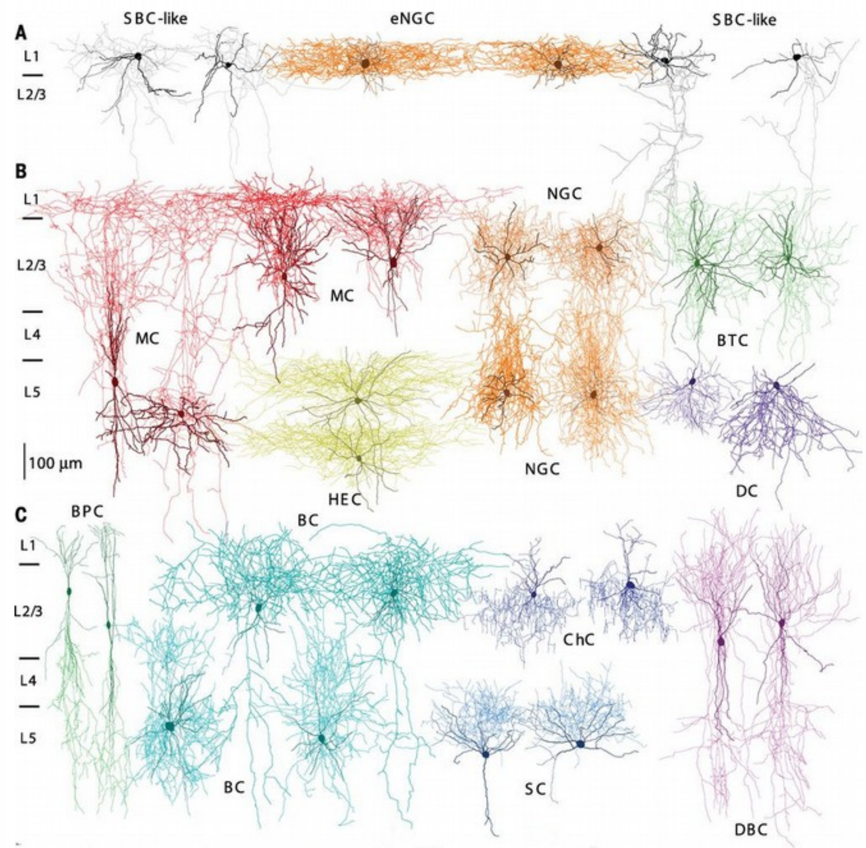
The current model of pyramidal cell integration...

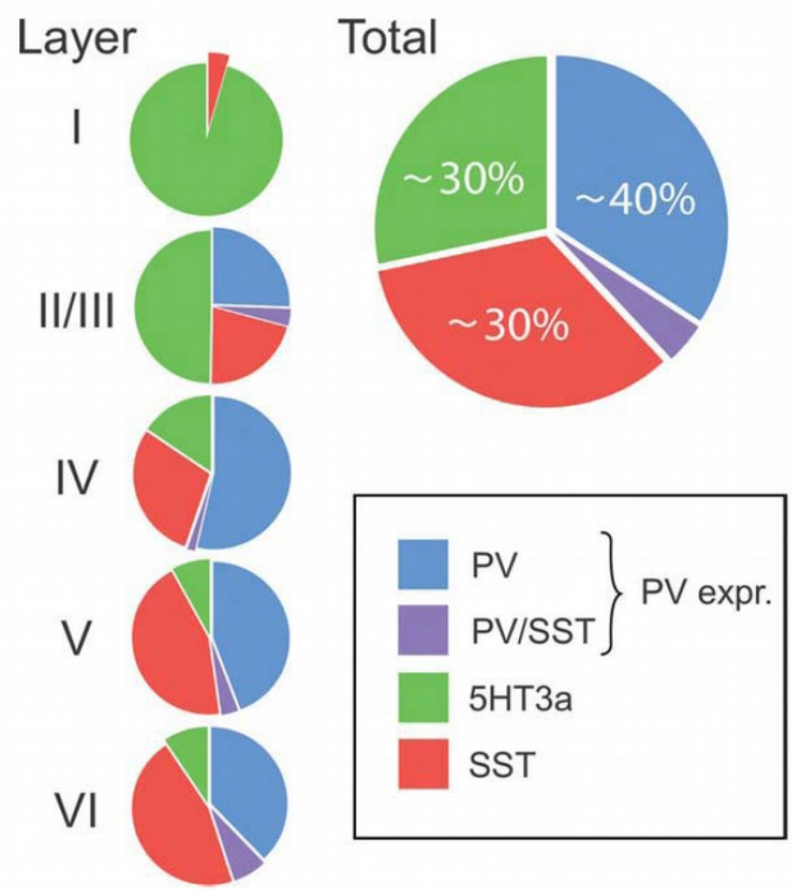


Larkum et al. (2009), *Science*

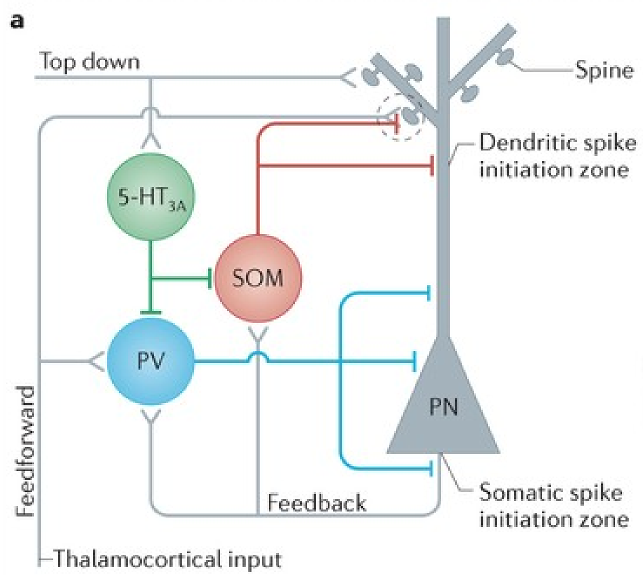
The other 30% of cells in the neocortex are gamma-amino butyric acid (**GABA**) releasing ***interneurons*** (which are usually inhibitory)

Current estimates are that there are around 15 types of GABAergic interneurons, though they all fall into 3 non-overlapping categories based on the expression of particular proteins: parvalbumin (**PV** or **FS**), somatostatin (**SST** or **SOM**) and the **5HT3a** serotonin receptor

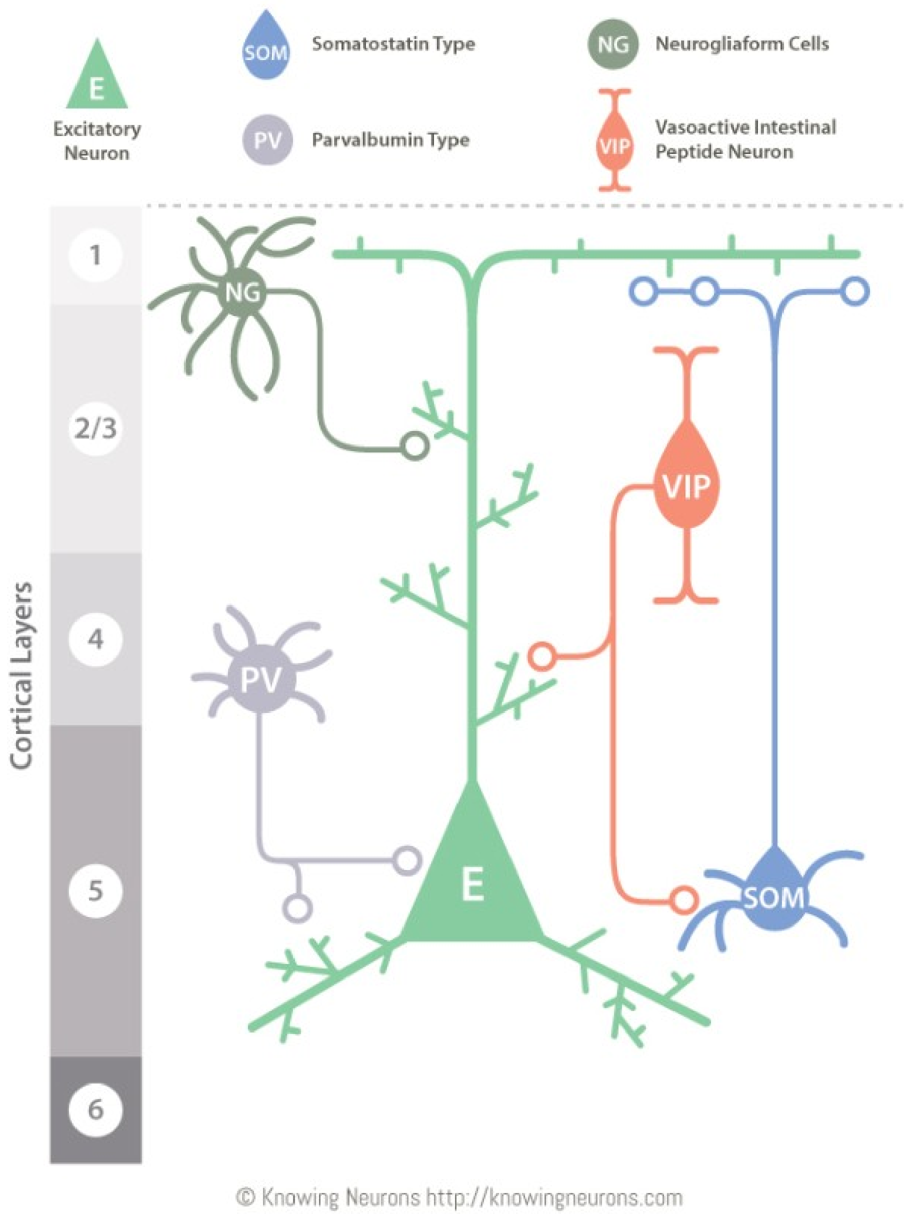




The 3 different major classes of interneurons have different targets/effects

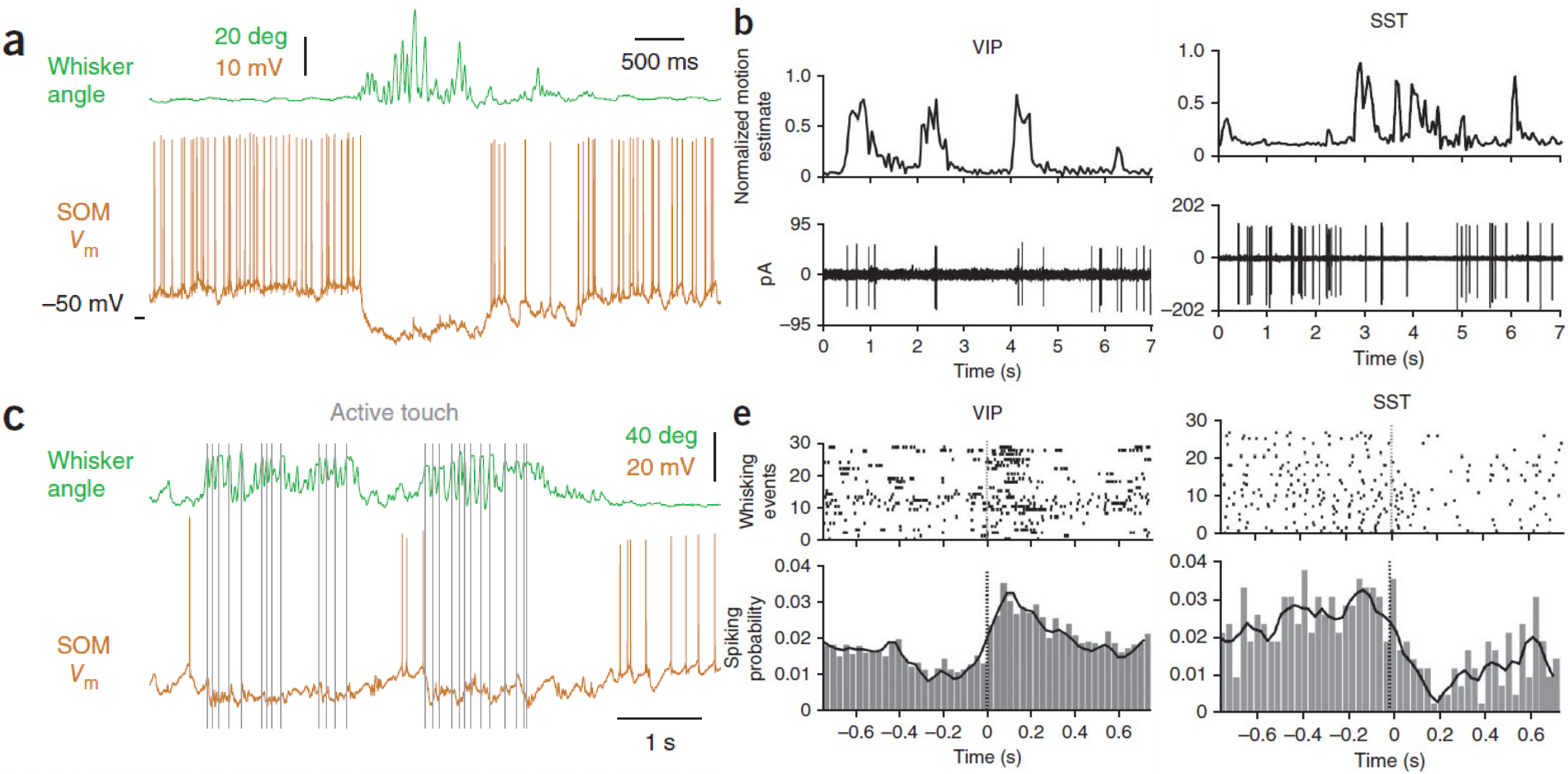


Higley (2014), *Nat. Rev. Neurosci*



Note: roughly half of 5HT3a interneurons are vasoactive intestinal poly peptide (**VIP**) interneurons

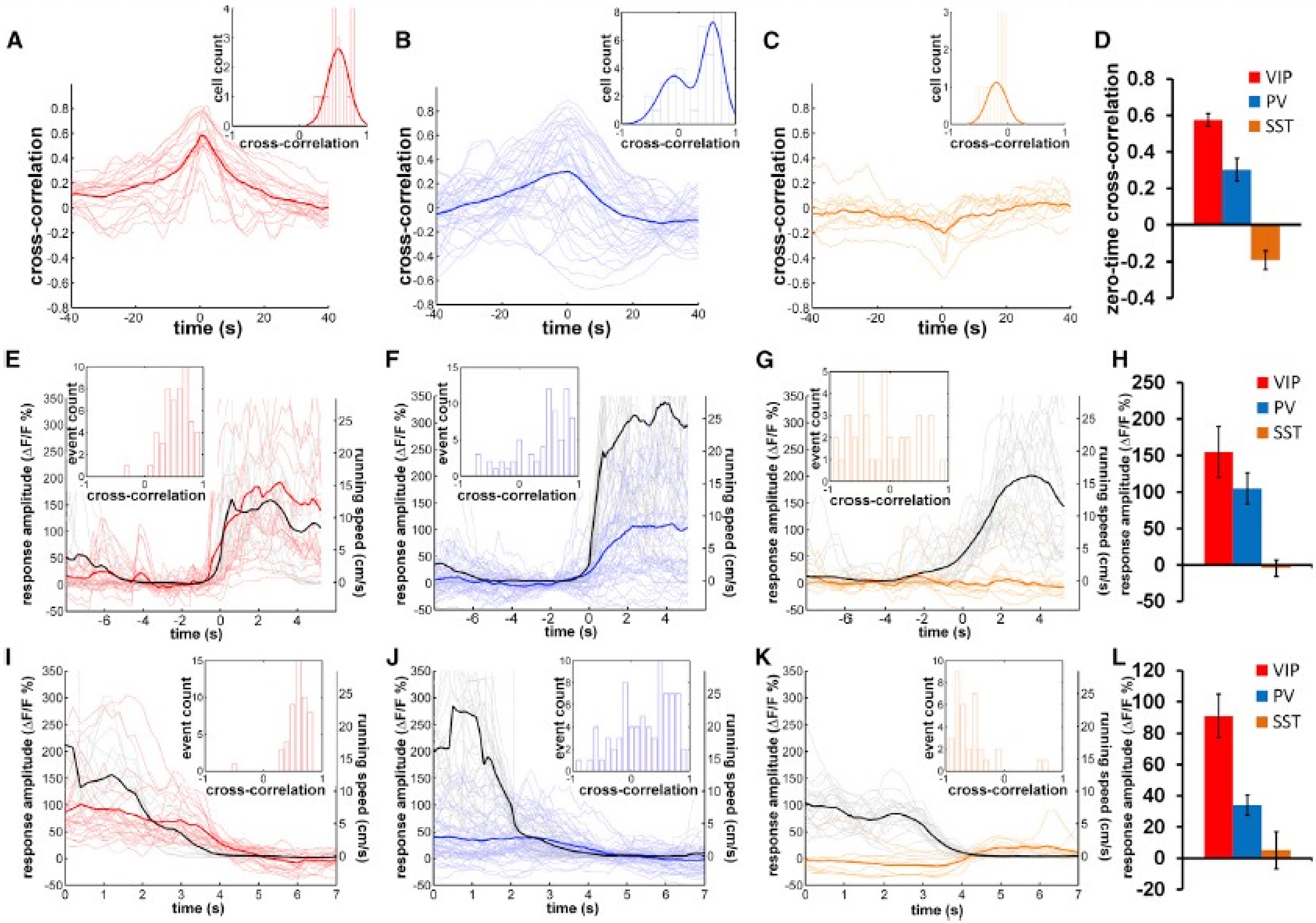
SST interneurons are inhibited by VIP interneurons during motor feedback



Gentet et al. (2012), *Nat. Neurosci.*

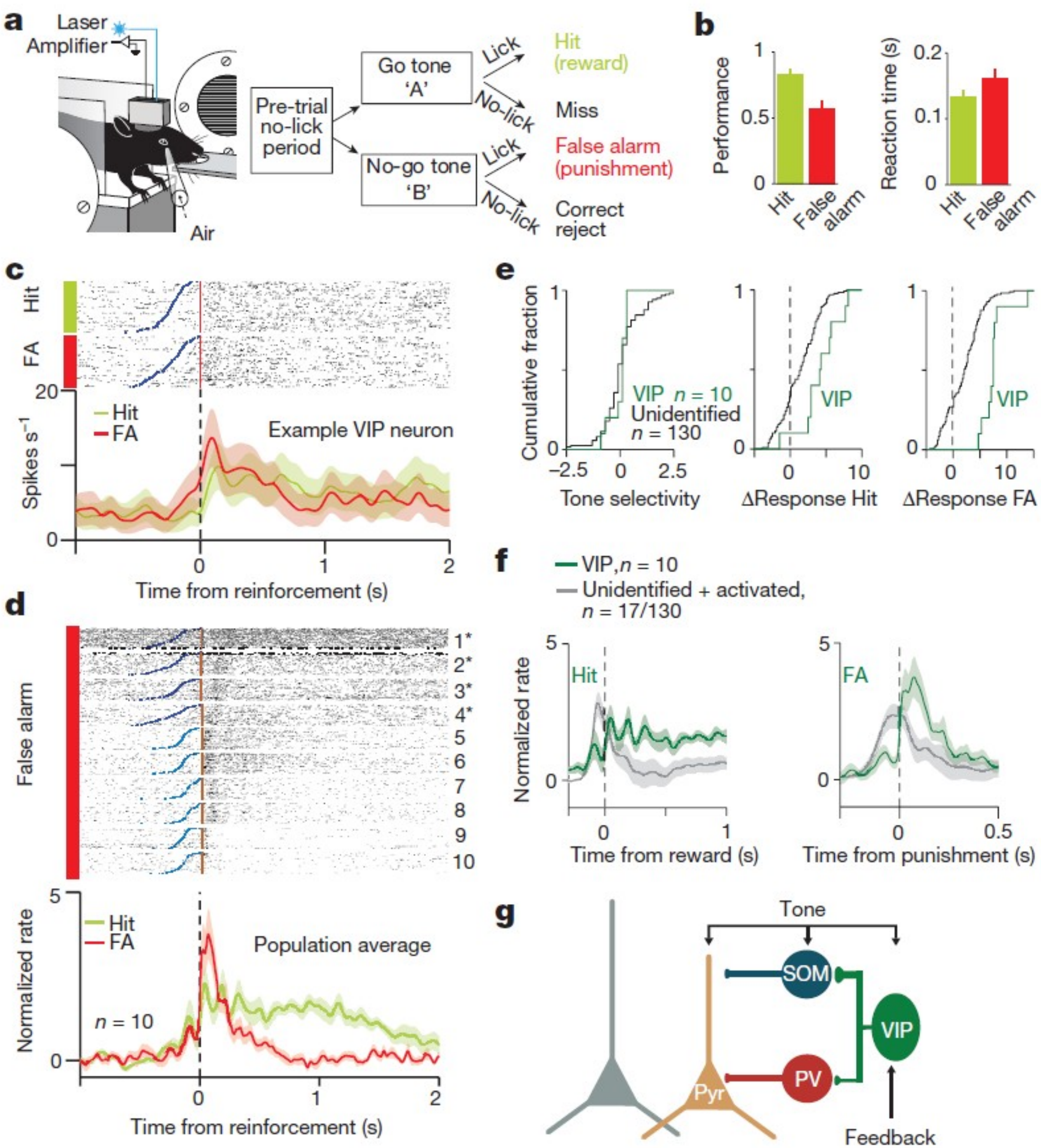
Lee et al. (2013), *Nat. Neurosci.*

SST interneurons are inhibited by VIP interneurons during motor feedback



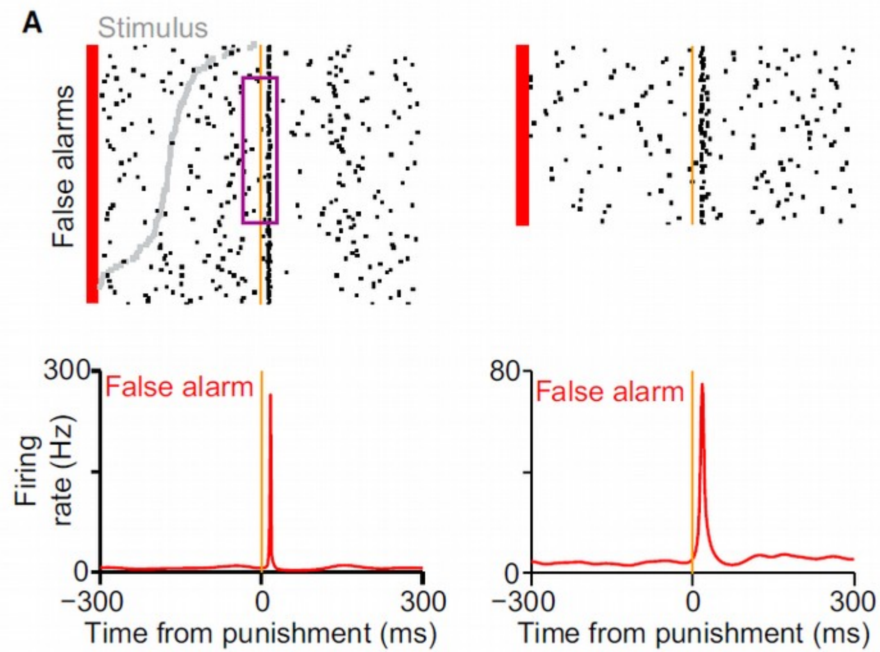
Fu et al. (2014), *Neuron*

VIP interneurons can be rapidly activated by reinforcement via acetylcholine inputs



Pi et al. (2013), *Nature*

Acetylcholine neurons



Hangya et al. (2015), *Cell*