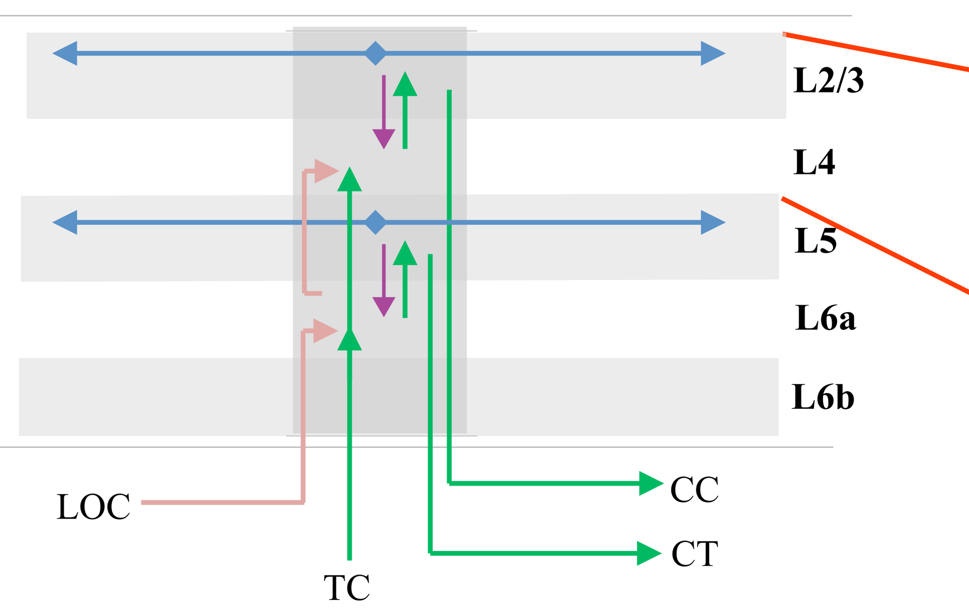
3D Object Learning with Cortical Columns

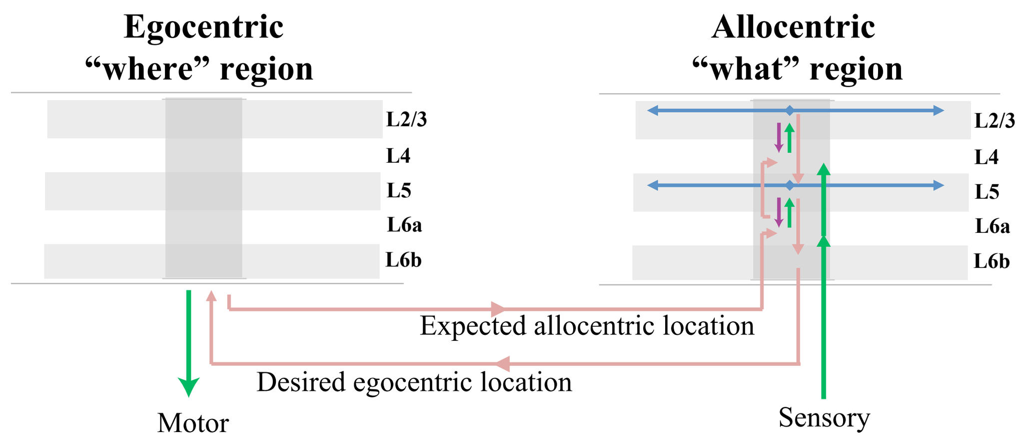
Jeff Hawkins, Yuwei Cui, Subutai Ahmad, Nathanael Romano, and Marcus Lewis

jhawkins@numenta.com, ycui@numenta.com, sahmad@numenta.com naromano@stanford.edu, and mlewis@numenta.com

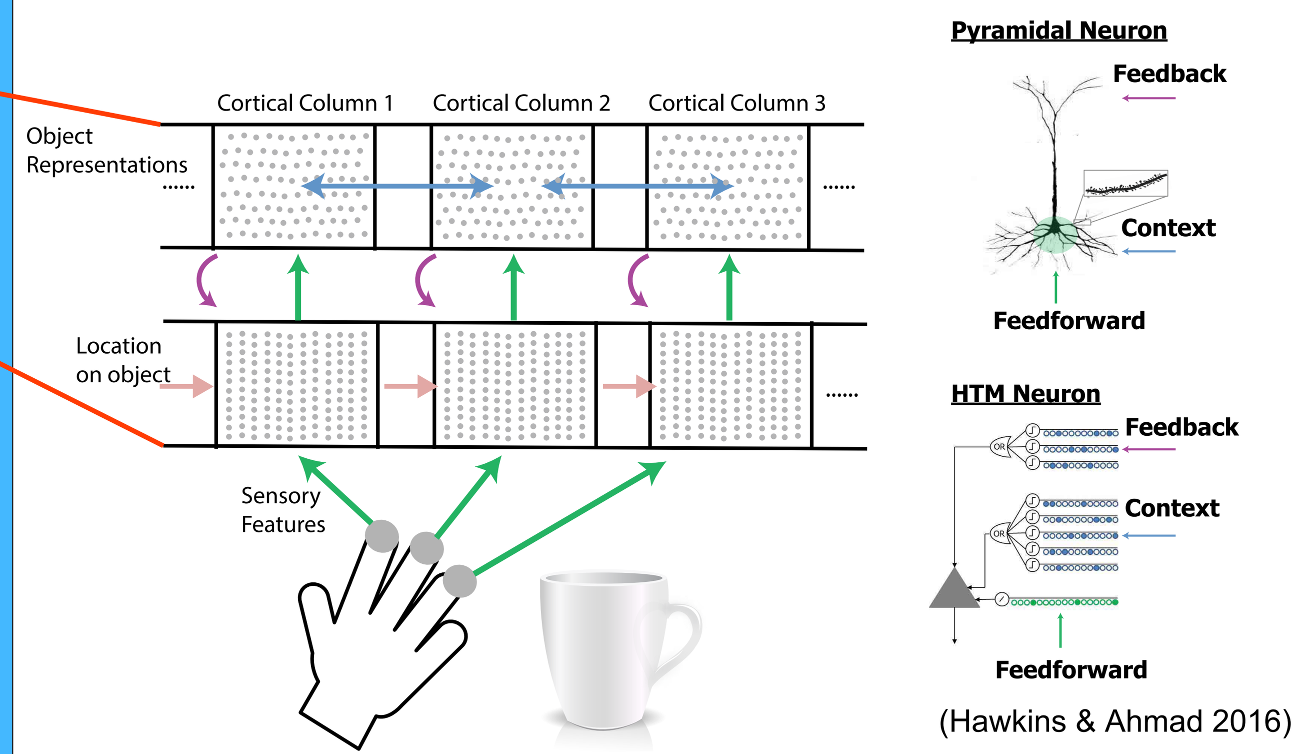
1. Cortex is organized into columns and layers



2. Cortical columns model 3D objects using sensory features and allocentric locations.



3. Detailed neural model



(Hawkins & Ahmad 2016)

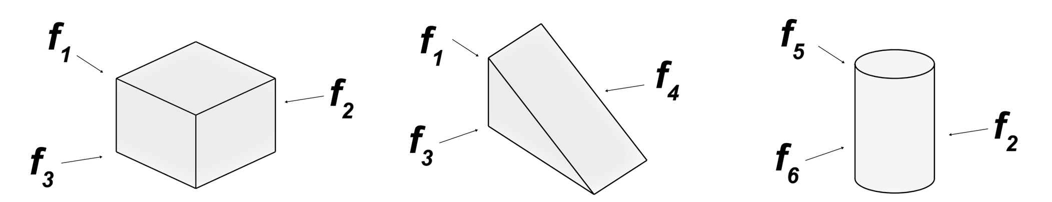
Network Model

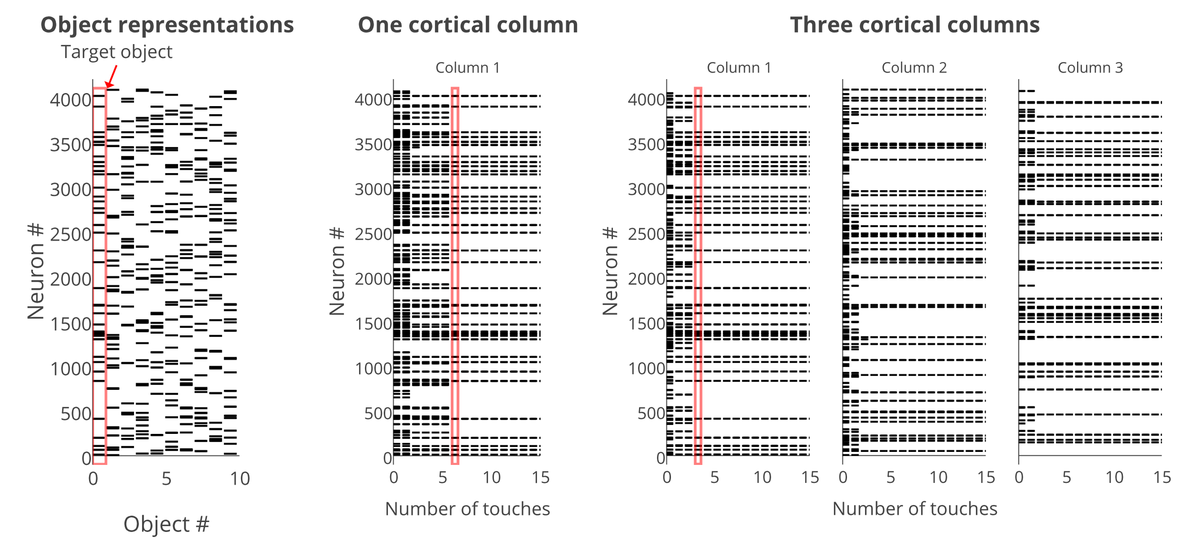
- The network models the two-layer motif that repeats twice in each cortical column.
- Input layer integrates features and location signals to form allocentric representations.
- Output layer learns stable representations of objects.
- Lateral connections across cortical columns integrates information across sensors.
- Feedback bias input layer towards representations that are consistent with recent inputs.

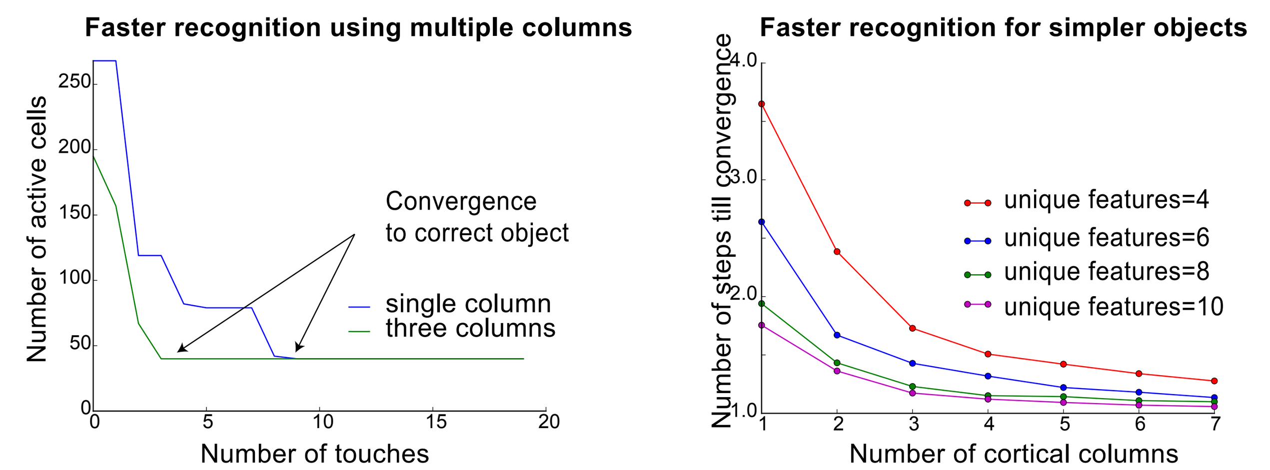
Neuron Model

- Proximal dendrites can recognize feedforward patterns and activate cells.
- Distal dendritic segments recognize lateral and feedback patterns and depolarize cells.

4. Simulation results

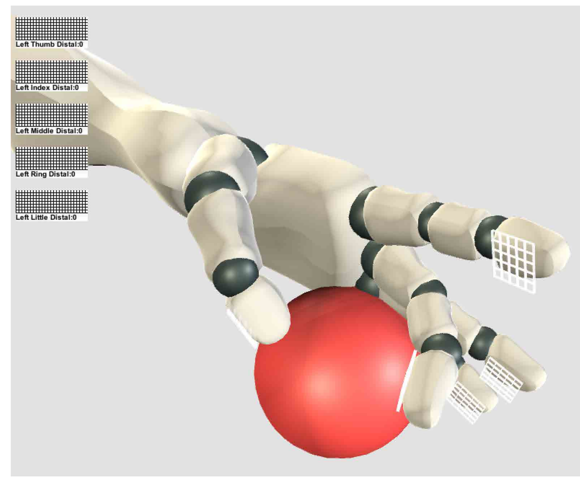






5. Object recognition with robotic hand

Simulated robot hand can grasp any object and recognize it.
Sensors on each fingertip sends touch information to corresponding column.



Dataset: Yale YCB Object Benchmark
Contains 77 objects

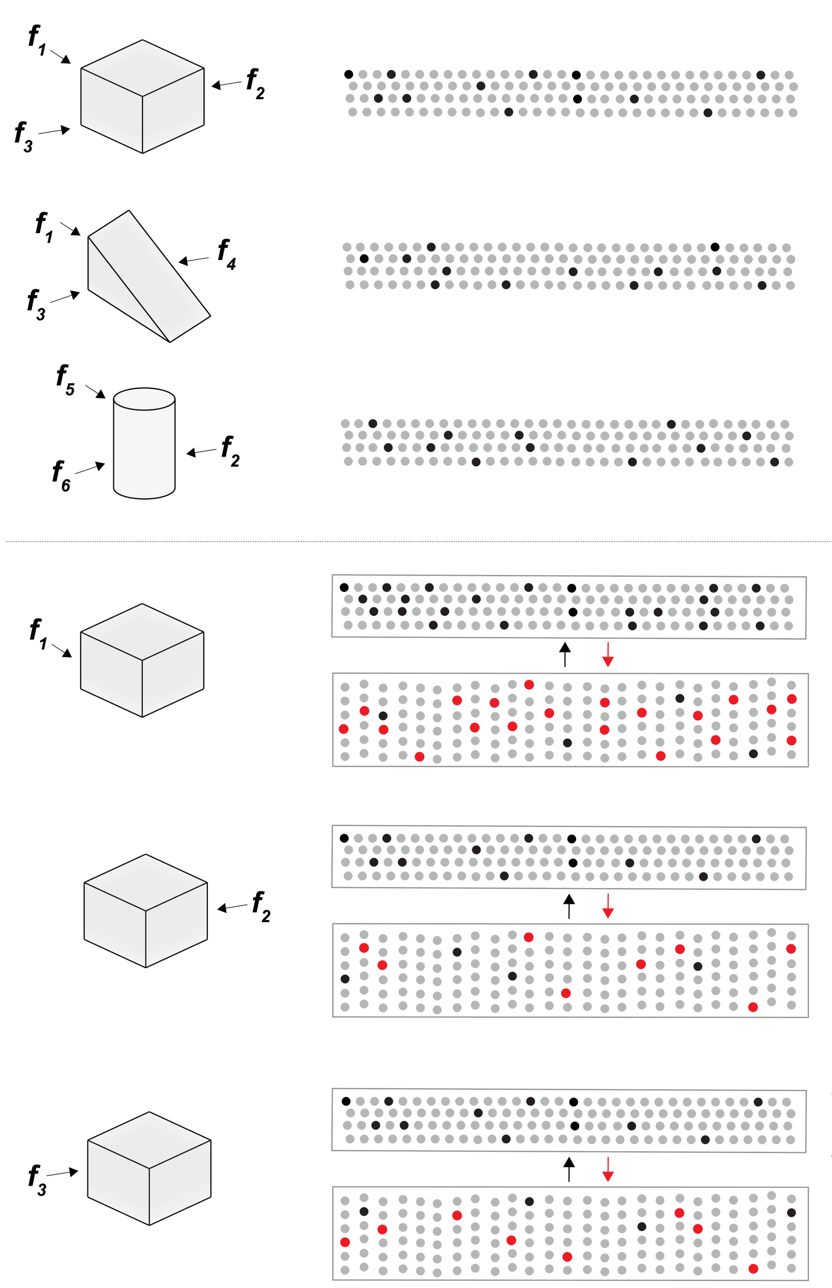
6. Summary

Cortical columns learn 3D sensorimotor models of the world by combining sensory inputs with allocentric location.

Cross-columnar connections allow faster inference.

We propose a detailed model consistent with anatomical and physiological evidence.

Activity in layers during inference



Representation for cube in output layer

Representation for wedge in output layer

Representation for cylinder in output layer

First touch is ambiguous and output layer contains union of two representations.

Second touch is ambiguous on its own, but unambiguous given first touch. Output layer contains only representation of cube.

Third touch is ambiguous on its own, but unambiguous given previous touches. Output layer contains only representation of cube.

Model details

Activation rules

Input Layer:

- If any cell in an active mini-column has lateral inputs, only those cells fire.
- If no cell in an active mini-column has lateral inputs, all cells in the mini-column fire.

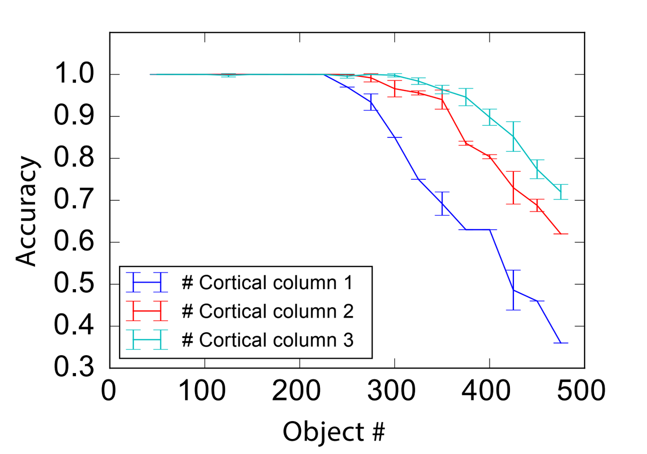
Output Layer:

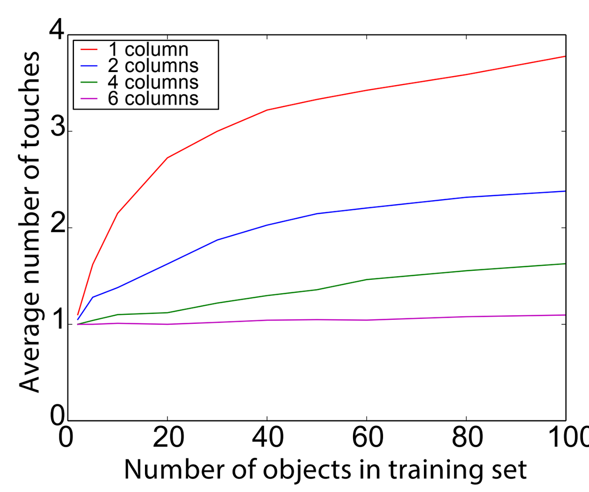
- Output cells with strong feedforward inputs and lateral inputs fire first.
- If no cell has lateral inputs, output cells with only feedforward inputs fire.
- Output cell activity persists if no feedforward inputs is provided.

Hebbian learning rules

- Whenever a cell is active, reinforce synaptic connections (LTP and LTD).
- The reinforcement for distal and apical segments is branch specific.

Capacity



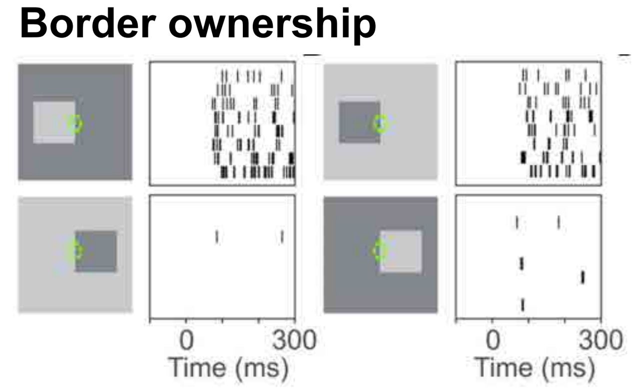


- A small network can store hundreds of complex objects.
- Network capacity increases with more cortical columns.
- Network capacity increases as a function of network size.

Experiment setup: Input layer:150 mini-columns, 16 cells/mini-column. Output layer: 4096 cells

Supporting experimental evidence

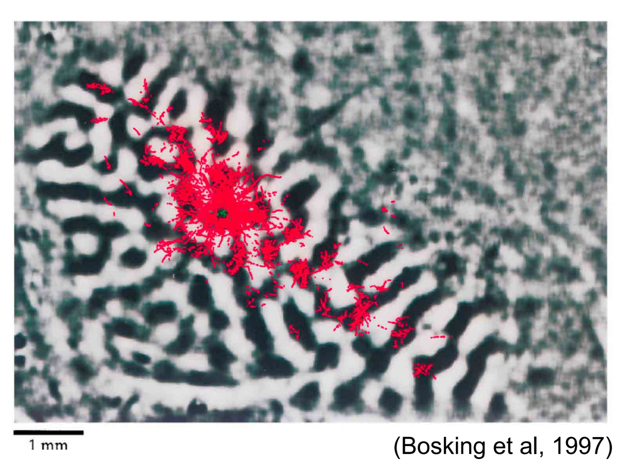
Border ownership



(von der Heydt, 2015):

- Some cells in V1 and V2 respond to location of specific features within an object's reference frame.
- Cells do not respond to same feature in different location.

Long range connections in layer 2

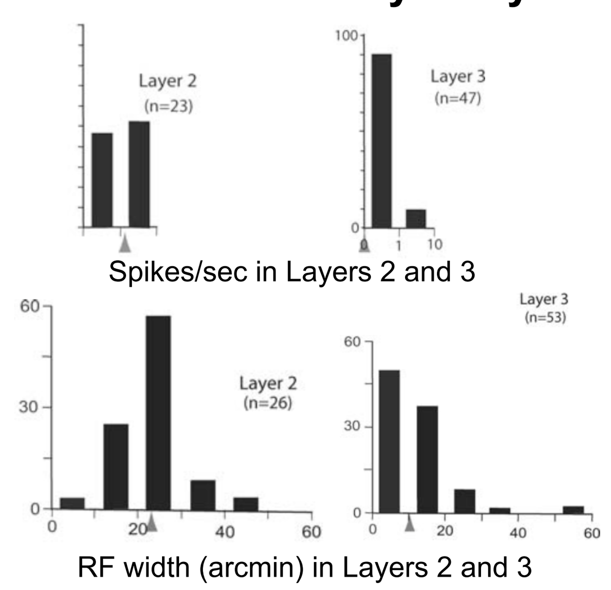


(Bosking et al, 1997)

(Bosking et al, 1997):

- Layer 2/3 cells have very long range lateral connections
- Connections are more dense locally

Increased stability in layer 2



(Gur and Snodderly, 2008):

- Layer 2 activity is more stable
- Layer 2 cells have wider RF's

Predictions of the theory

- Each region contains cells stable over movement of the sensor.
- The activity of these stable cells are specific to object identity.
- The output layers (those with long-range lateral connections) form these stable representations. Their activity will be more stable than input layers.
- Object representations within each column will converge on stable representation quicker with lateral connections.
- Object representations within each column will quickly become sparser as more evidence is accumulated for an object. Cell activity in output layer is denser for ambiguous objects.
- Each region contains cells tuned to location of features in object's reference frame (invariant to ego-position, e.g. border ownership).
- We expect to see these representations in the input layer.

References

J. Hawkins, S. Ahmad, Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex, Front. Neural Circuits. 10 (2016) 1–13.

A.M. Thomson, a P. Bannister, Interlaminar connections in the neocortex., Cereb. Cortex. 13 (2003) 5–14.

W.H. Bosking, Y. Zhang, B. Schofield, D. Fitzpatrick, Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex, J. Neurosci. 17 (1997) 2112–2127.

M. Gur, D.M. Snodderly, Physiological differences between neurons in layer 2 and layer 3 of primary visual cortex (V1) of alert macaque monkeys., J. Physiol. 586 (2008) 2293–306.

R. von der Heydt, Figure–ground organization and the emergence of proto-objects in the visual cortex, Front. Psychol. 6 (2015) 1695.