A thousand brains: toward biologically constrained AI
by: Kjell Hole and Subutai Ahmed

Presented by Eloy Geenjaar

Moving towards general AI

- Current deep learning approaches are narrow AI

- These approaches fundamentally can not reach general AI

- Using computational neuroscientific ideas from the neocortex may help

Moving towards general AI continued

Table 1 Problems with current AI and solutions proposed by neuroscience studies of the neocortex

Existing AI	Neuroscience
Offline, batch-oriented learning	Online, continuous learning
Supervised learning	Unsupervised learning
Massive training sets	Streaming of sensory data
Fragility to noise and faults	SDRs provide robustness to noise and faults
Manual process to create system	Automatic, general system
Inefficient power usage	Highly power efficient
Symbol manipulation	Embodied reasoning*

*More research is needed

Important aspects

- What is intelligence?
  "*Human intelligence* is the brain's ability to learn a model of the world and use it to understand new situations, handle abstract concepts, and create novel behaviors, including manipulating the environment"
- Sensorimotor integration
  - Integration of sensory processing and generation of motor commands may be essential to general intelligence
- Embodied reasoning
  - Sensorimotor mechanisms shape abstract reasoning abilities

Current AI

- “Current AI is almost independent of its environments”

- “Typical AI systems do not learn anything after training, no matter the type of learning algorithm they use”

- “Narrow AI performs one well-defined task in a single domain”

Limitations of narrow AI to reach more general AI

- They are greedy, brittle, and opaque

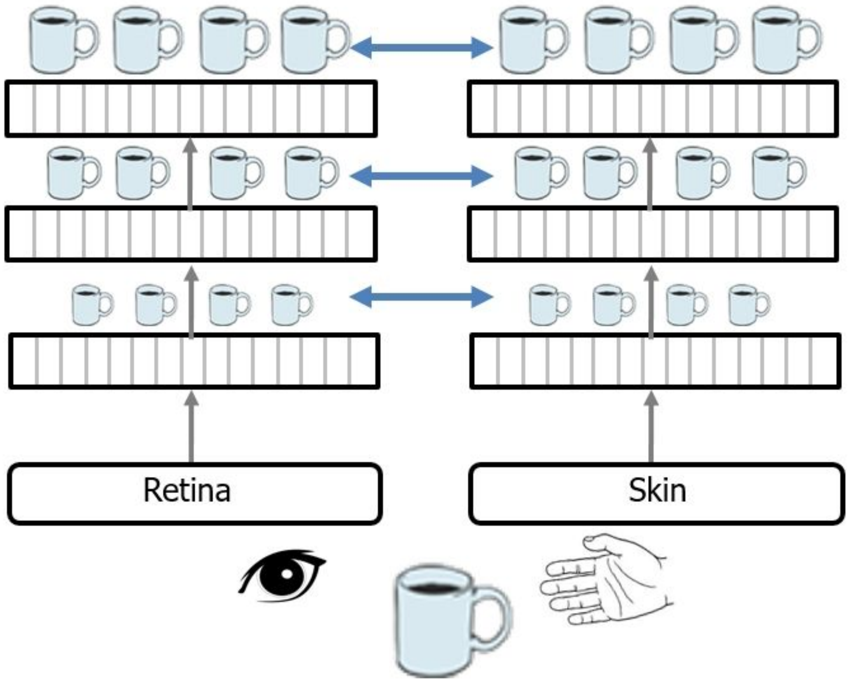
- They lack abstract reasoning abilities and possess common sense about the world

- Narrow AI is a proxy for the person that made it

The neocortex

- Neurons in the neocortex have about 1000-2000 synapses
- Each region operates according to the same principles
- Hierarchical, horizontal, and feedback integration

A thousand brains



Three data structure properties of the neocortex

- Sparse data representations

- Realistic neuron model

- Reference frames

Organization of neurons

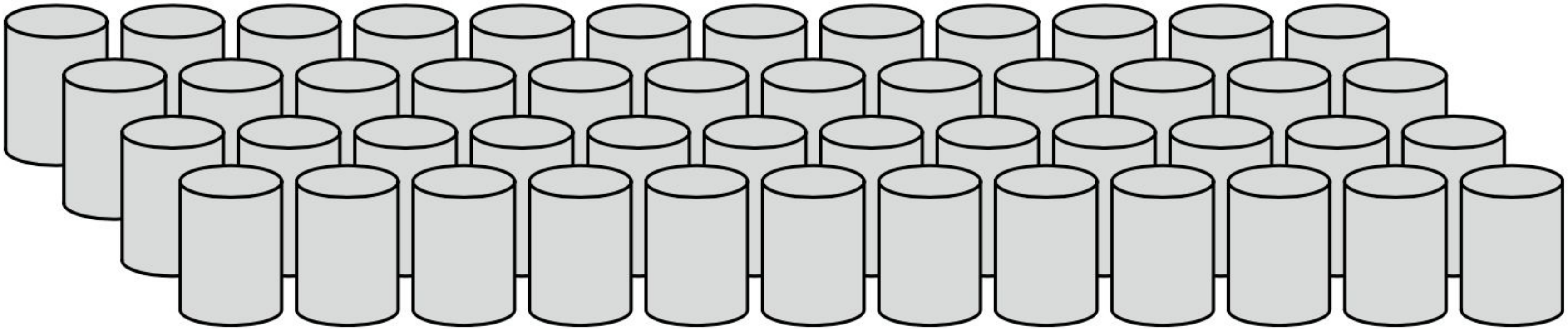


Fig. 2 Cortical columns in a region of the neocortex

Organization of neurons

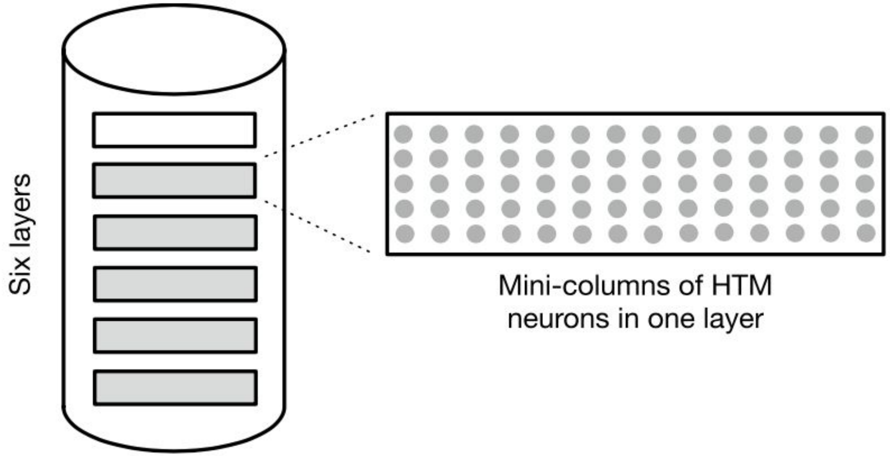


Fig. 3 Single cortical column in HTM with six layers, where five layers contain mini-columns of HTM neurons

Hierarchical temporal memory (HTM)

- Based on thousand brains theory

- Uses broad ideas in the computational neuroscience community for a general computation framework

- Mostly bases structure off of pyramidal neurons

Sparse distributed data representations

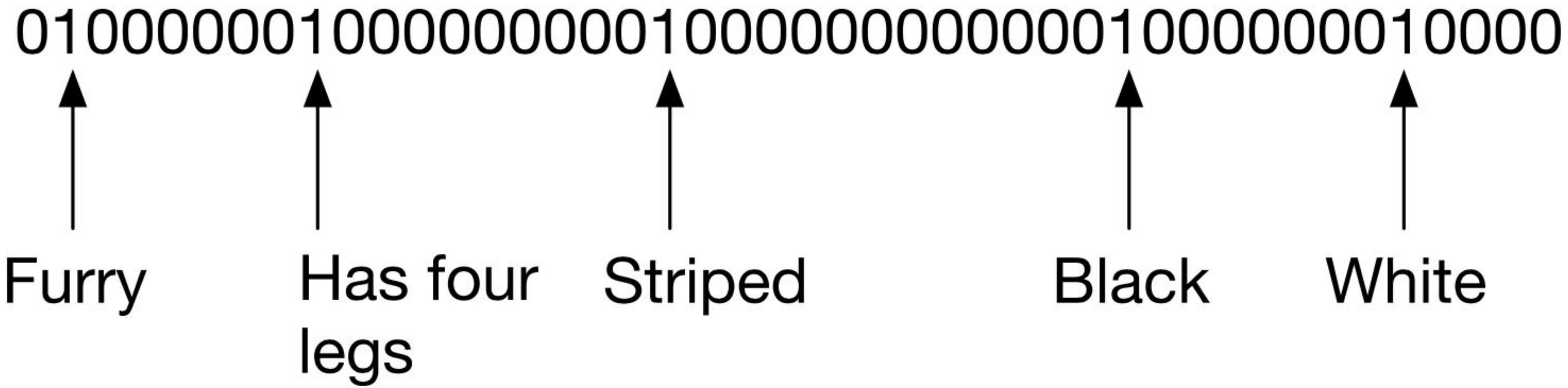
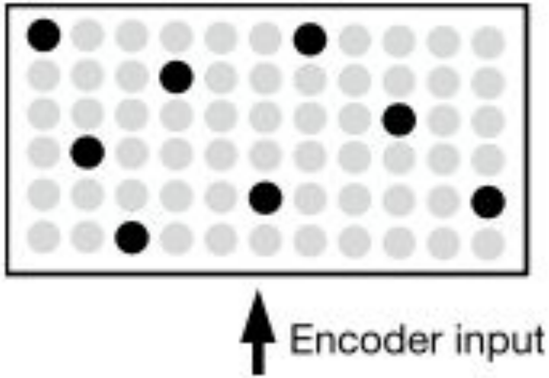


Fig. 5 Simplified SDR for a zebra. Adapted from [23]

Sparse distributed data representations

Fig. 7 A layer of HTM neurons organized into mini-columns. The black HTM neurons represent 1s and the rest represent 0s



Realistic neuron model

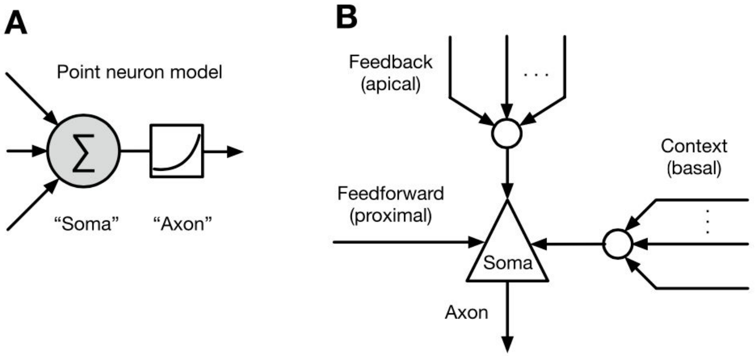


Fig. 6 **a** The point neuron is a simple model neuron that calculates a weighted sum of its inputs and passes the result through a non-linearity. **b** Sketch of the HTM neuron adapted from [37]. The feed-forward input determines whether the soma moves into the active state and fires a signal on the axon. The sets of feedback and context dendrites determine whether the soma moves into the predictive state

Reference frames

- Changes in movement are mapped to locations in the reference frame

- Every column maintains allocentric frame models of the objects it senses

- Locations are based on grid cells (entorhinal cortex)

HTM neurons: sparsity

- Semantic meaning depends on the input

- Nearby bit positions represent similar positions

HTM neurons

- Apical, basal, and proximal feedback

- Three states: active, predictive, inactive
  - Active: 1
  - Inactive, predictive: 0
  - Predictive: primed cell body

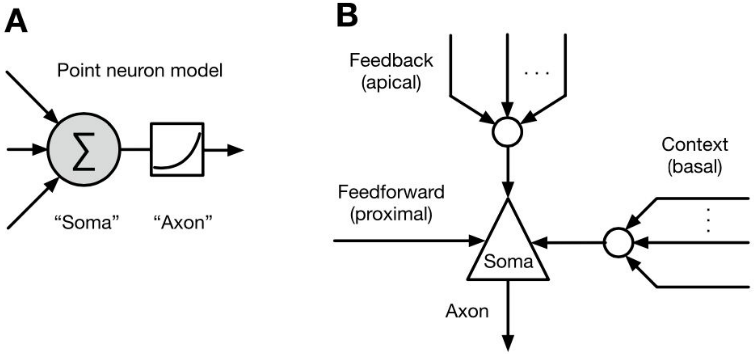


Fig. 6 **a** The point neuron is a simple model neuron that calculates a weighted sum of its inputs and passes the result through a non-linearity. **b** Sketch of the HTM neuron adapted from [37]. The feed-forward input determines whether the soma moves into the active state and fires a signal on the axon. The sets of feedback and context dendrites determine whether the soma moves into the predictive state

- Each neuron predicts its activations (moves into predictive state based on basal and apical dendrites)

HTM neuron predictions

- State at time t - 1 of network
- Some neurons move to their predictive state based on state at t - 1
- Predictions can be based on feedback projections from higher levels
- Learning is done by reinforcing connections consistent with the predictions

Learning of connections

- Connections are either 0 or 1
- They use a permanence value and how well two neurons align in terms of their activations

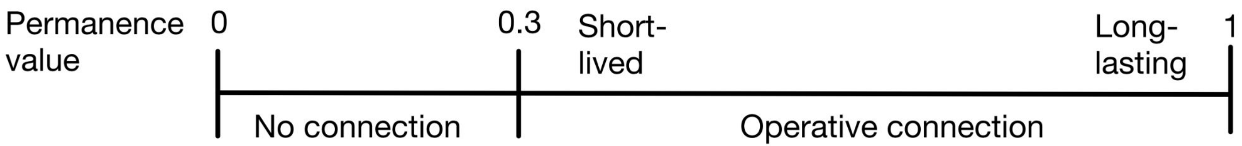


Fig. 9 The permanence value's effect on connections between HTM neurons

HTM sensorimotor integration

Two layers:

- First (input) layer combines sensory and input location
- Second (output) layer converges to stable representations
    - (1) Integrates information over time as sensors move relative to objects
    - (2) Lateral connections between columns communicate about information in different locations of the same object

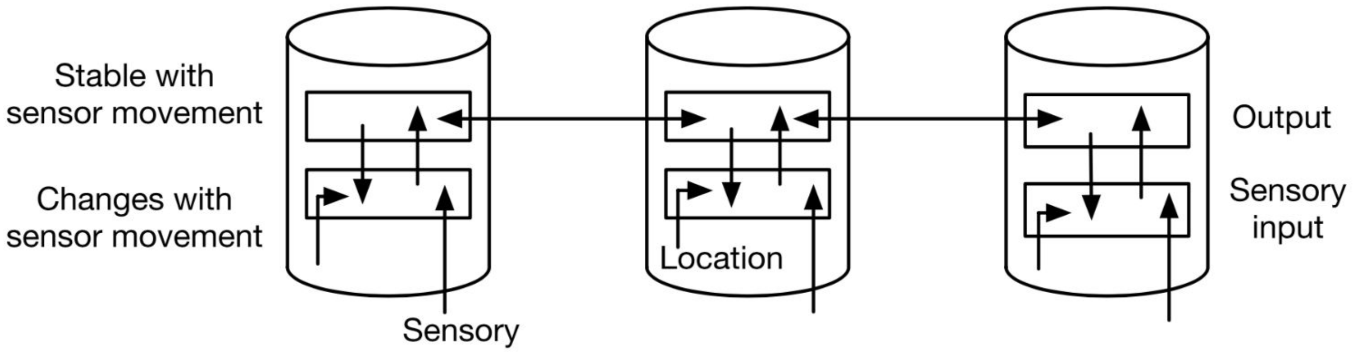
HTM reference frame

- If we observe a boat, each cortical column sees a different subset of the boat and are in different locations in the hierarchy

- The representations are different in each column and the columns communicate through traversal connections

HTM reference frame

Fig. 10 Three connected cortical columns in a region. The layer generating the location signal is not shown. Adapted from [42]



Simulation results

- HTM networks function even with 40% of neurons ‘off’

- Learning converges to a stable representation for different inputs

- Achieves comparable accuracy to other sequence learning algorithms, including statistical, feedforward, and recurrent networks

Simulation results

- Multiple columns work together to minimize the number of ‘sensations’ needed to recognize an object
- Future work: abstract reasoning
- HTM implemented on FPGA only uses 225W

Thalamus

- Necessary to include thalamus to abstract model of the neocortex because it is essential to its functioning

- “A cortical region sends information directly to another region, but also indirectly via the thalamus. These indirect routes between the regions strongly indicate that it is necessary to understand the operations of the thalamus to fully understand the neocortex.”

Thalamus

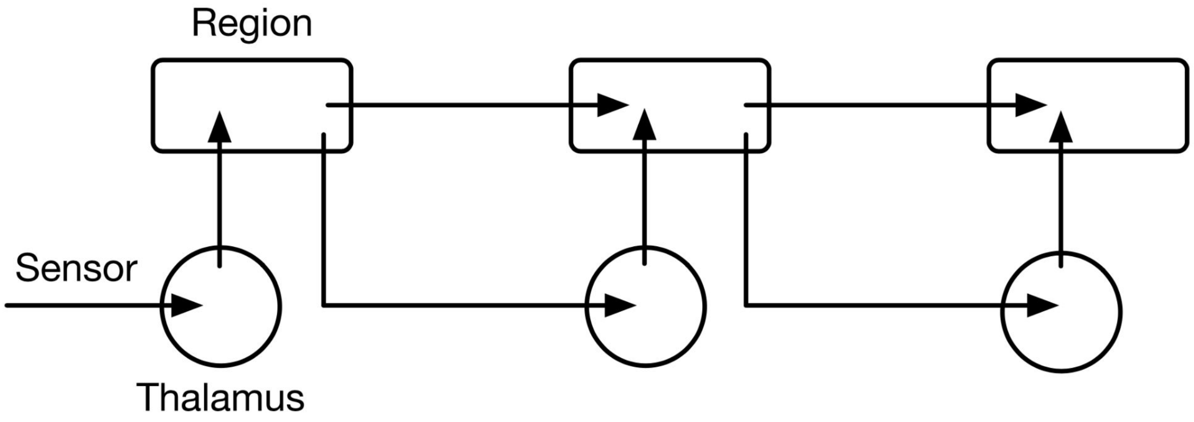


Fig. 11 The connections between the regions in the neocortex are both direct and via the thalamic nuclei. Adapted from [41, 47]

Conclusion

- New neuron models must allow prediction of future state
- Lack of sensorimotor integration for current AI major problem
- Collaboration between AI and neuroscience researchers essential
- HTM full potential not yet shown