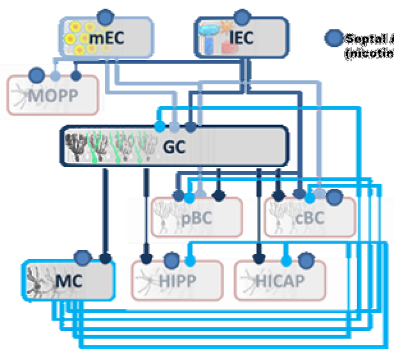
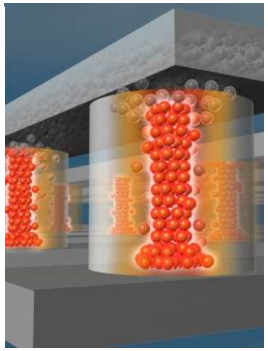
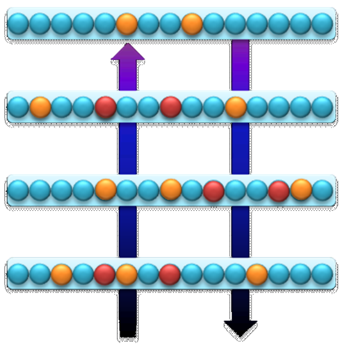
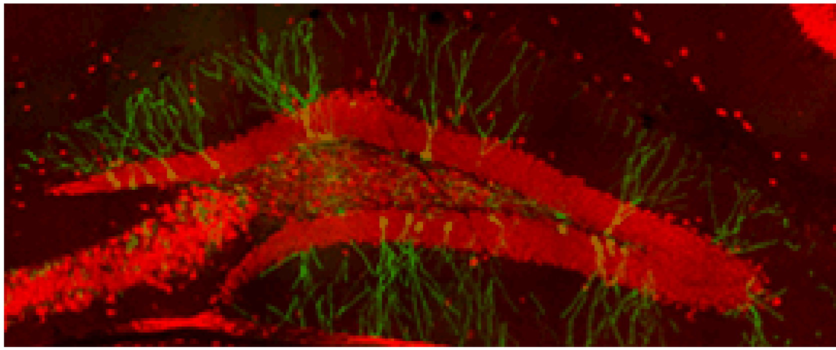
SAND2017-2181C









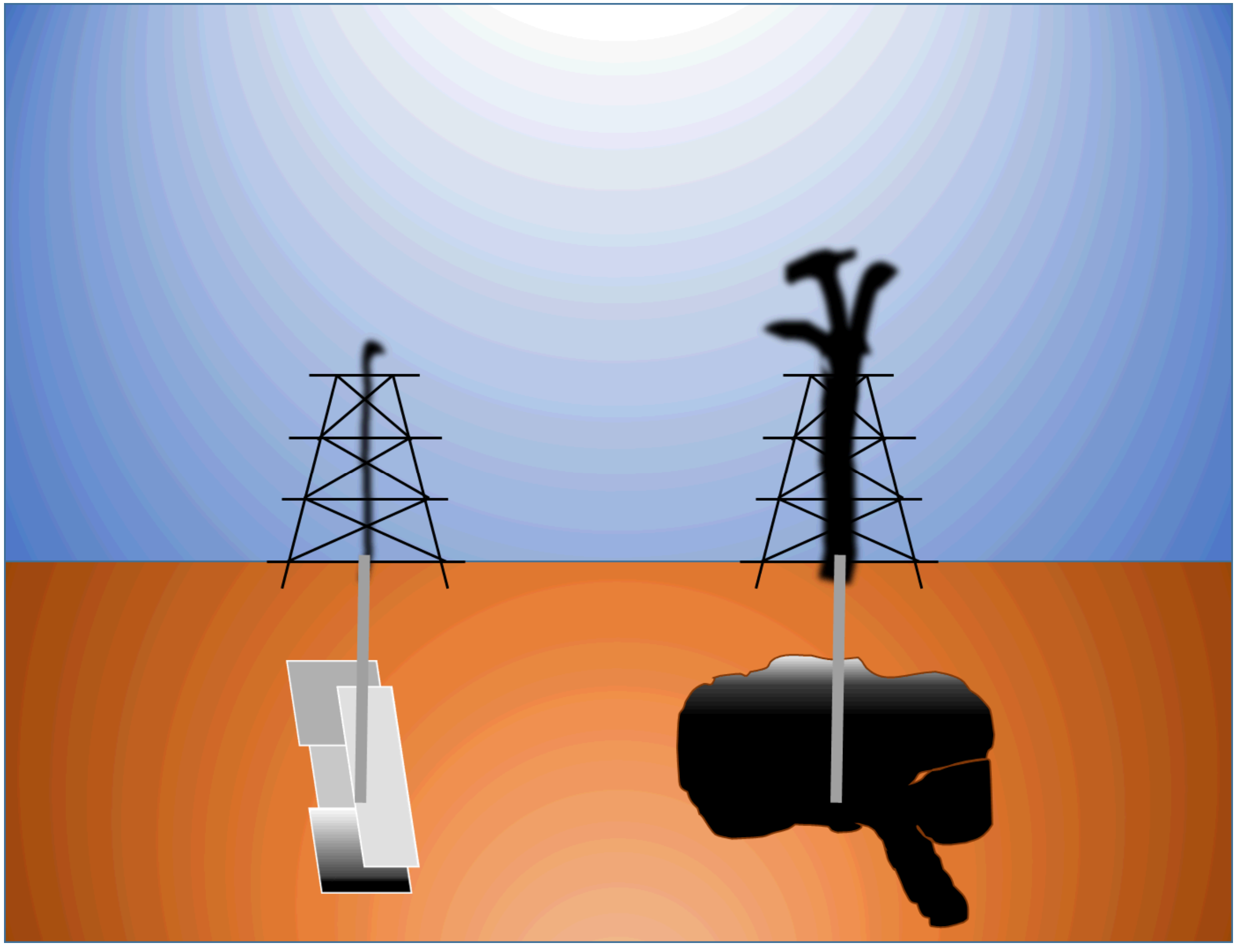
Hippocampus-inspired Adaptive Neural Algorithms

Brad Aimone

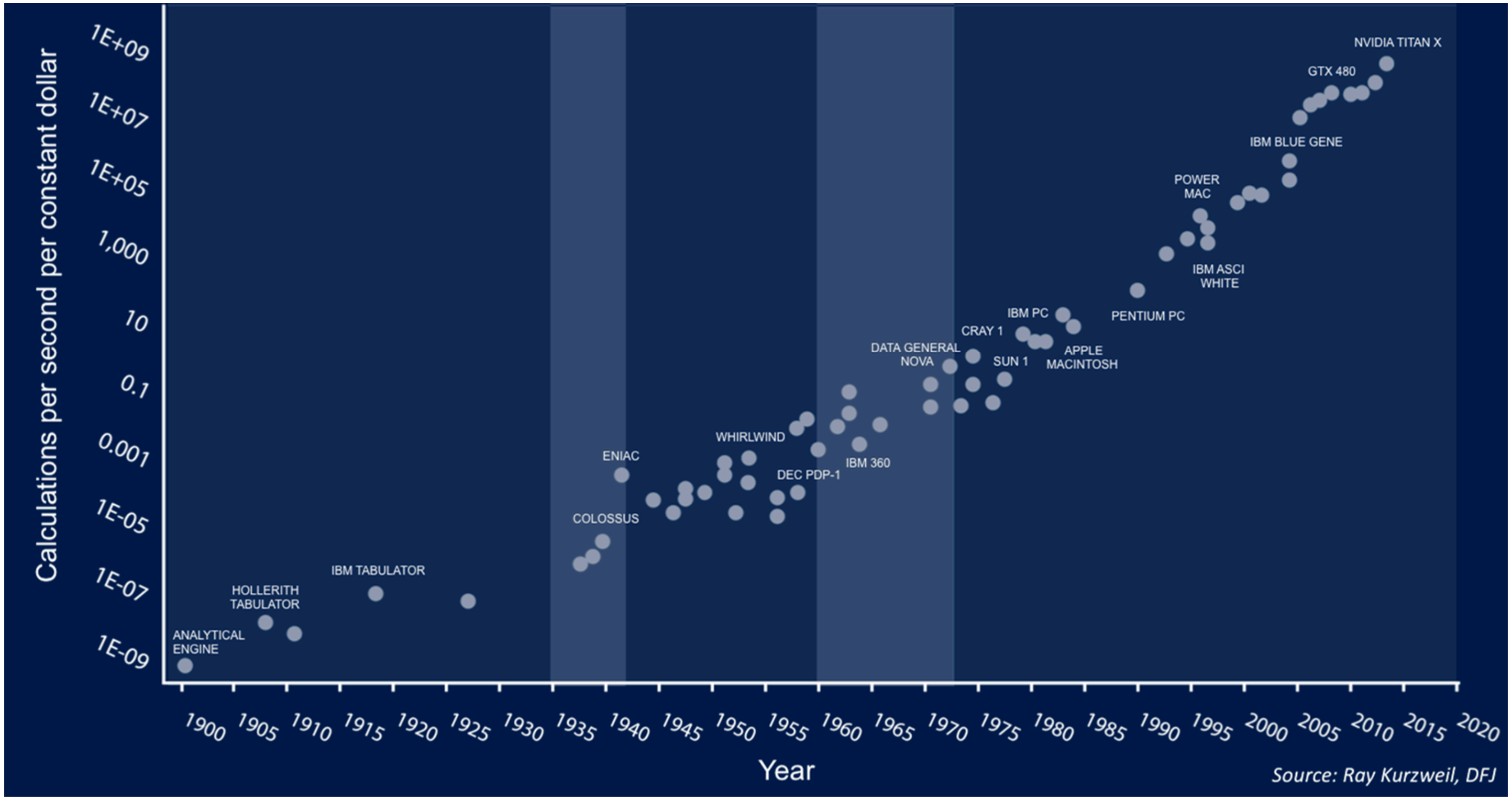
Center for Computing Research
Sandia National Laboratories; Albuquerque, NM

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Can neural computing provide the next Moore's Law?

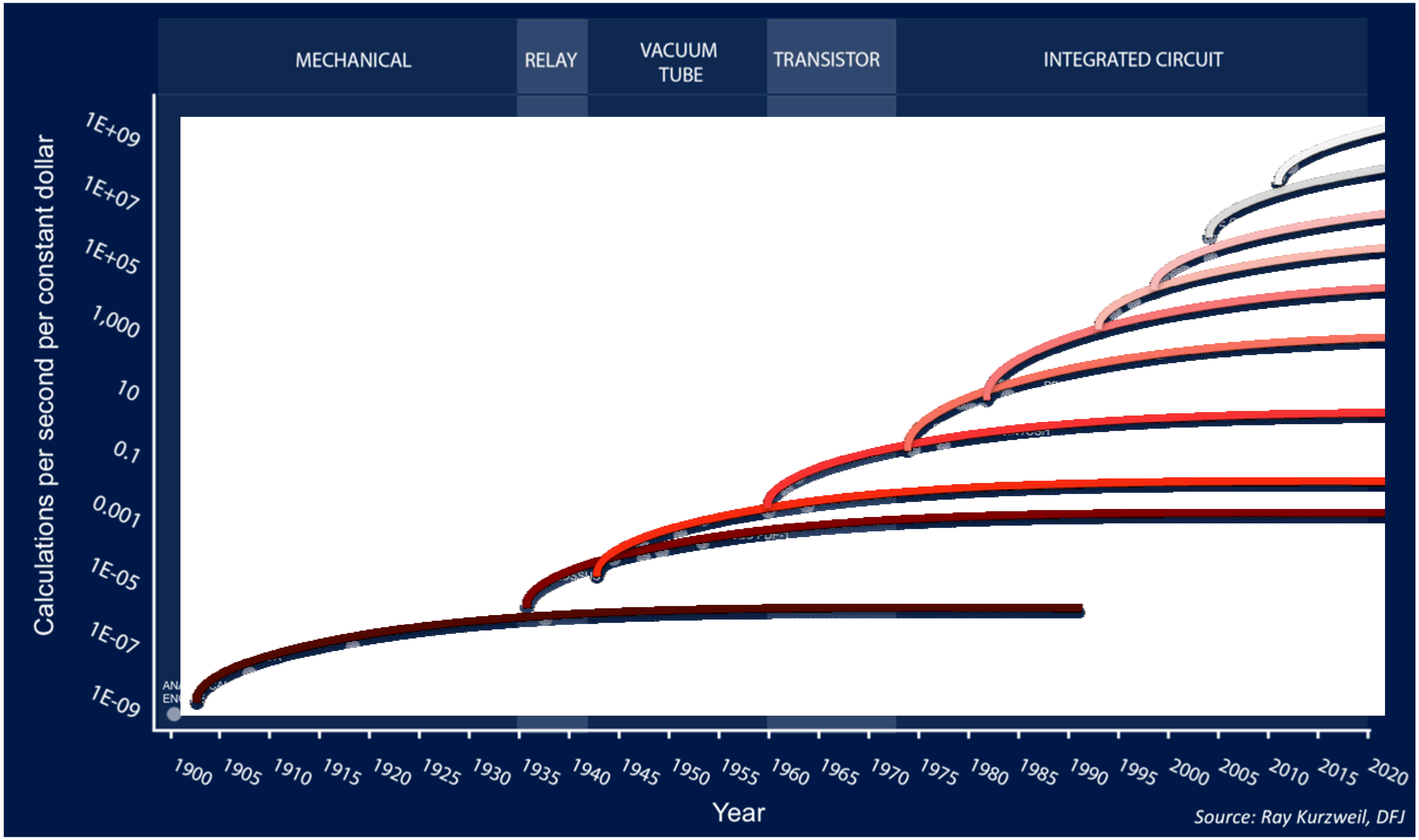


Moore's Law was based on scientific discovery and successive innovations



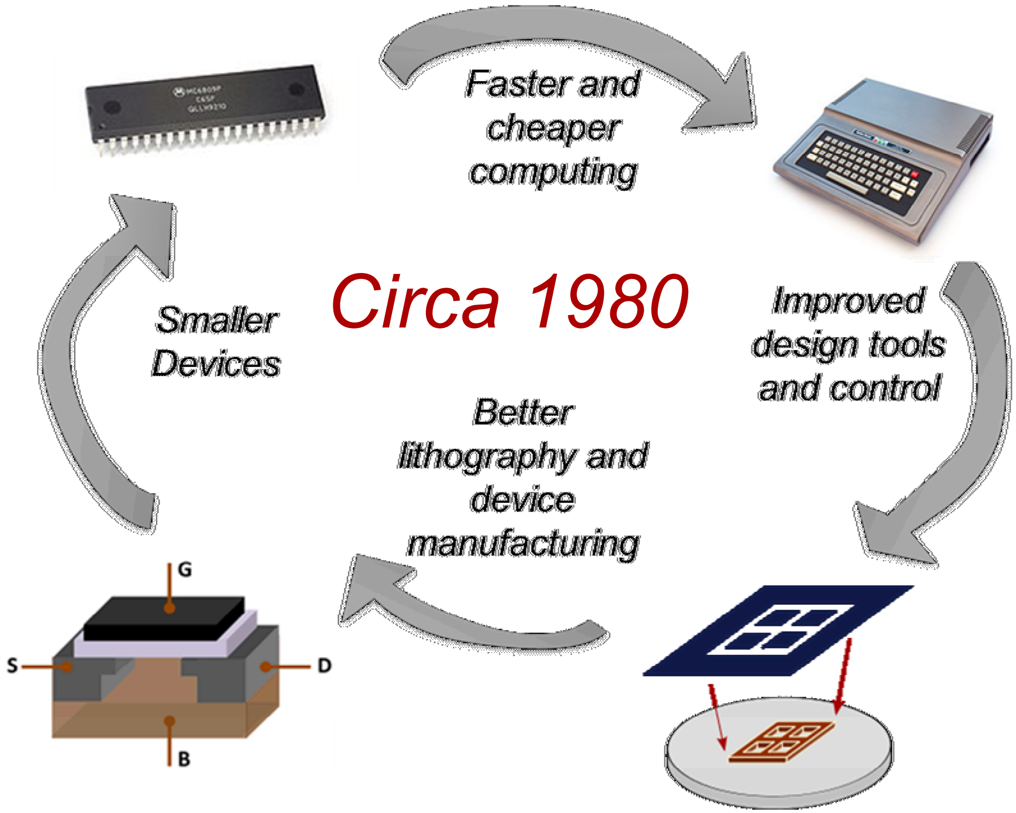
Adapted from Wikipedia

Each successive advance made more computing feasible



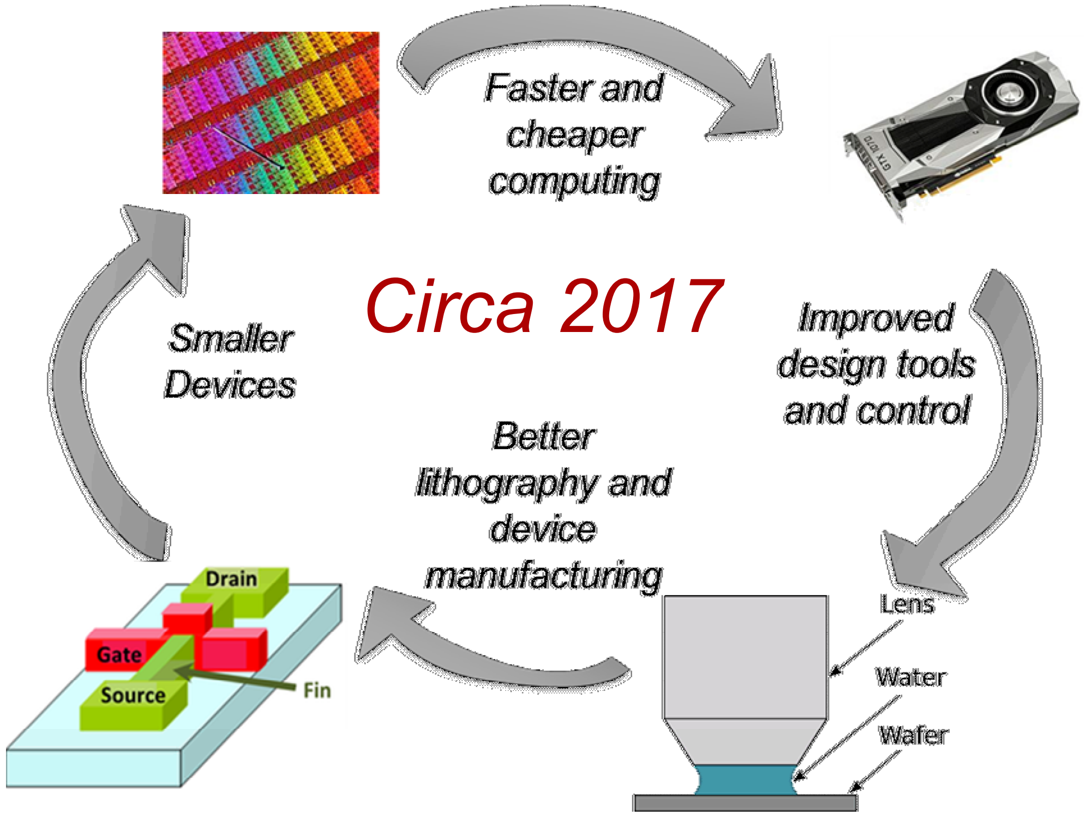
Adapted from Wikipedia

Better devices made better computers, which allowed engineering new devices…



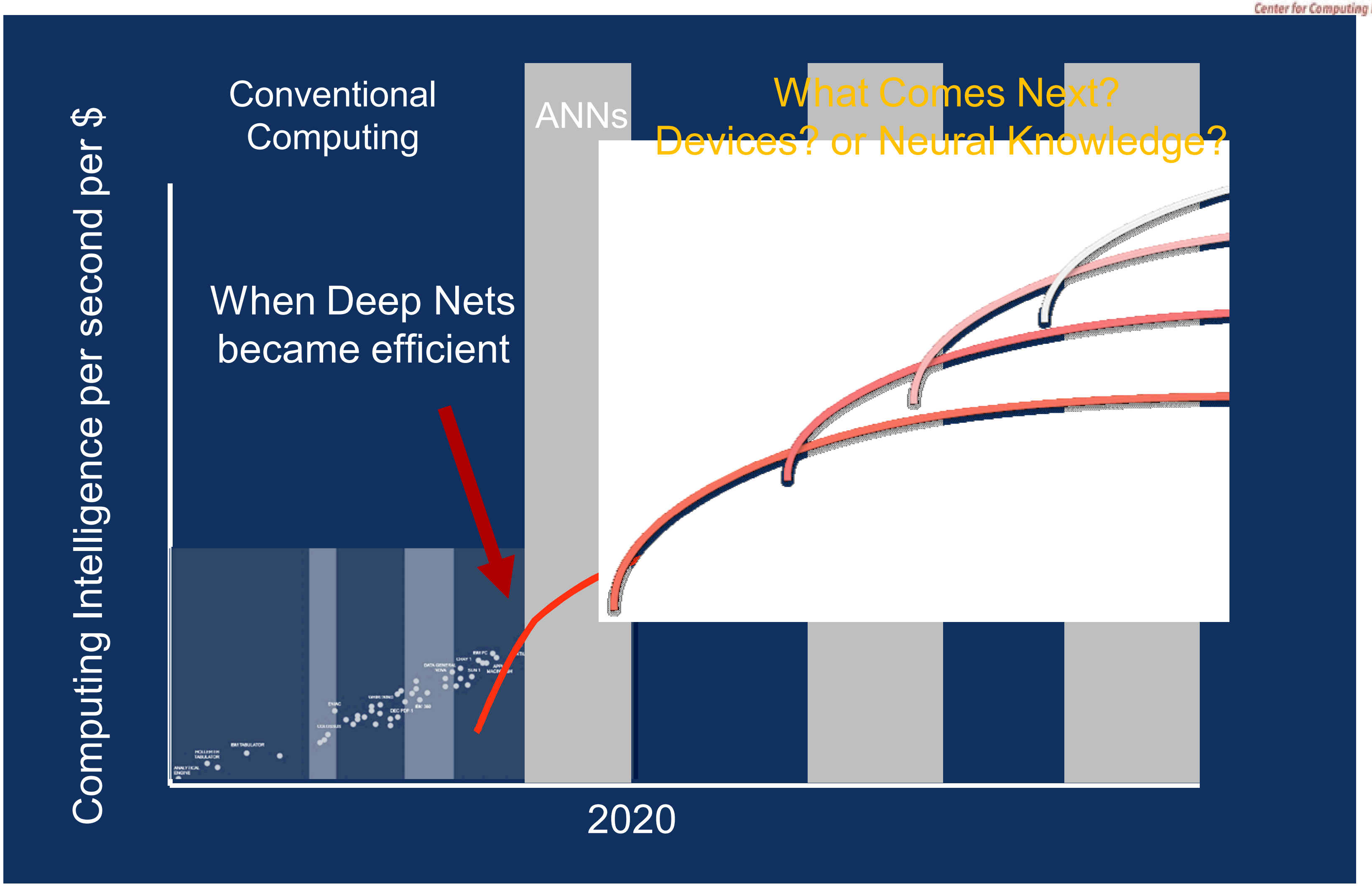
Images from Wikipedia

Better devices made better computers, which allowed engineering new devices…

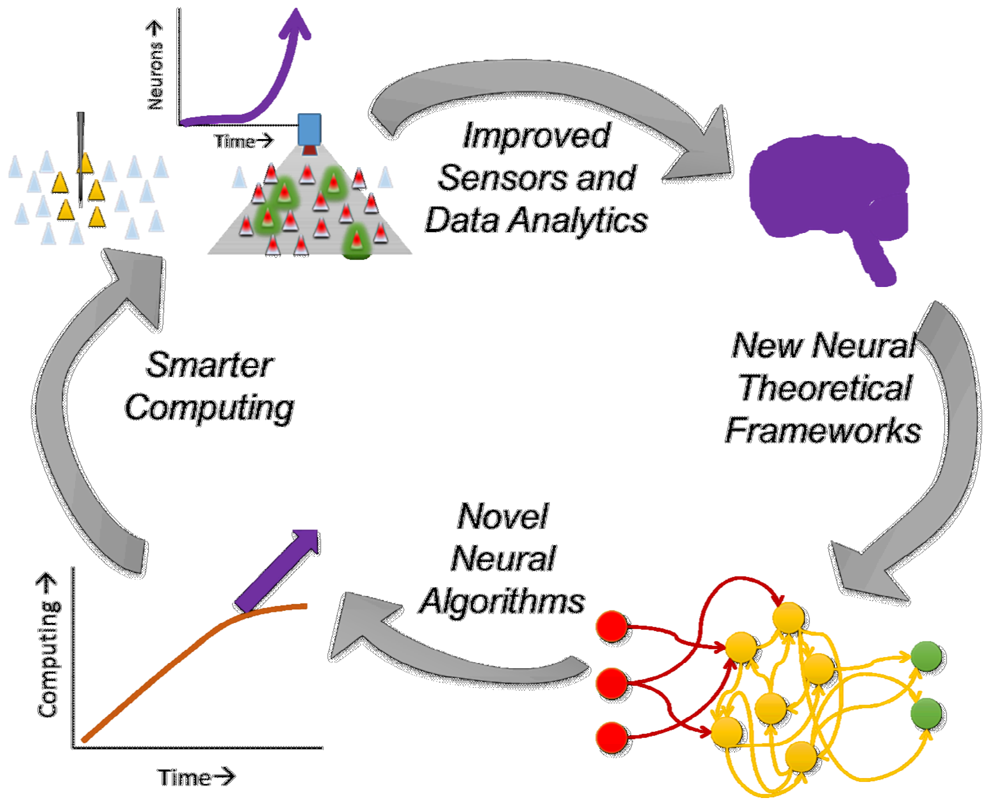


Images from Wikipedia

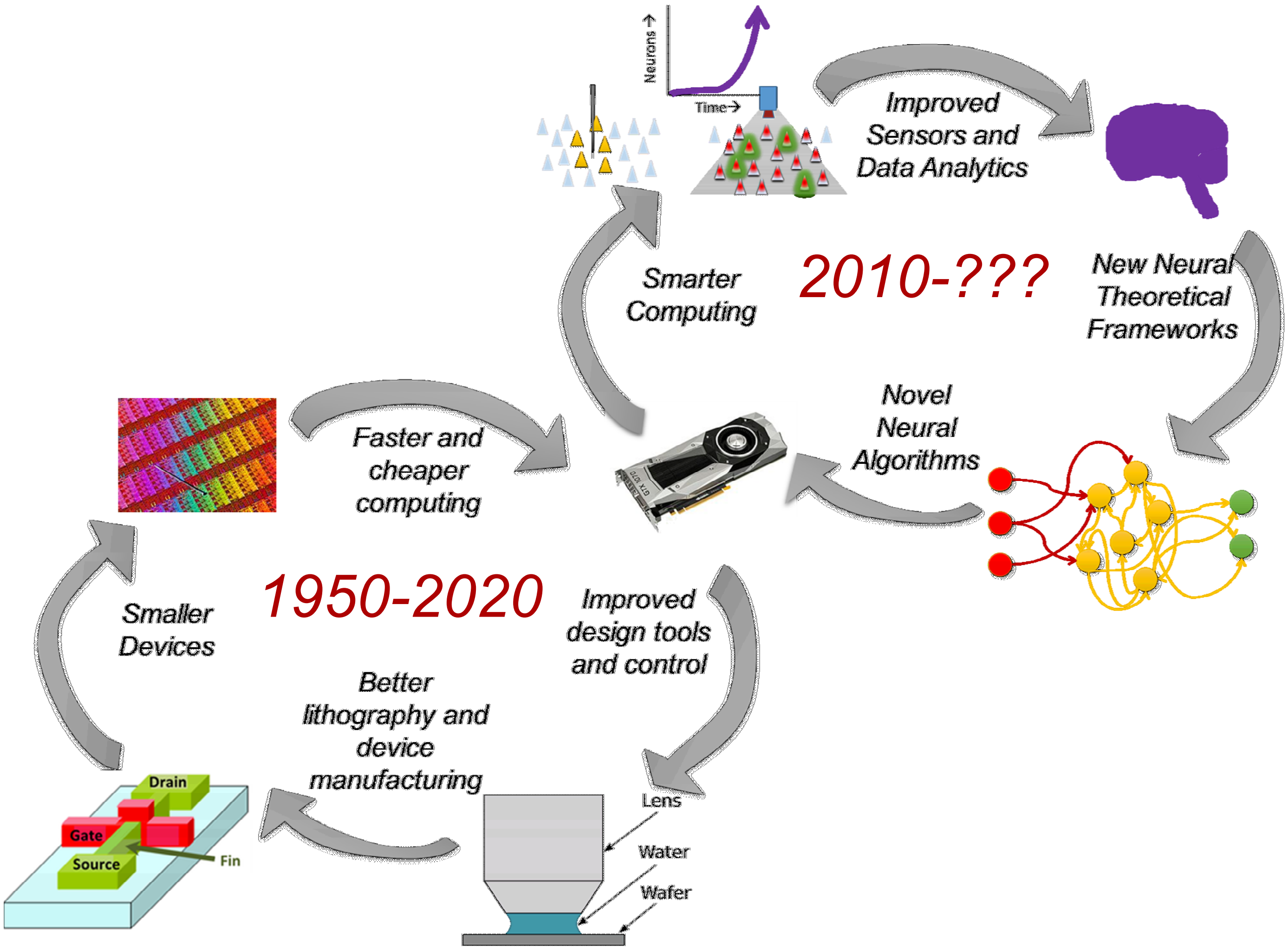
If we extrapolate *capabilities* out, it is not obvious better devices is the answer…



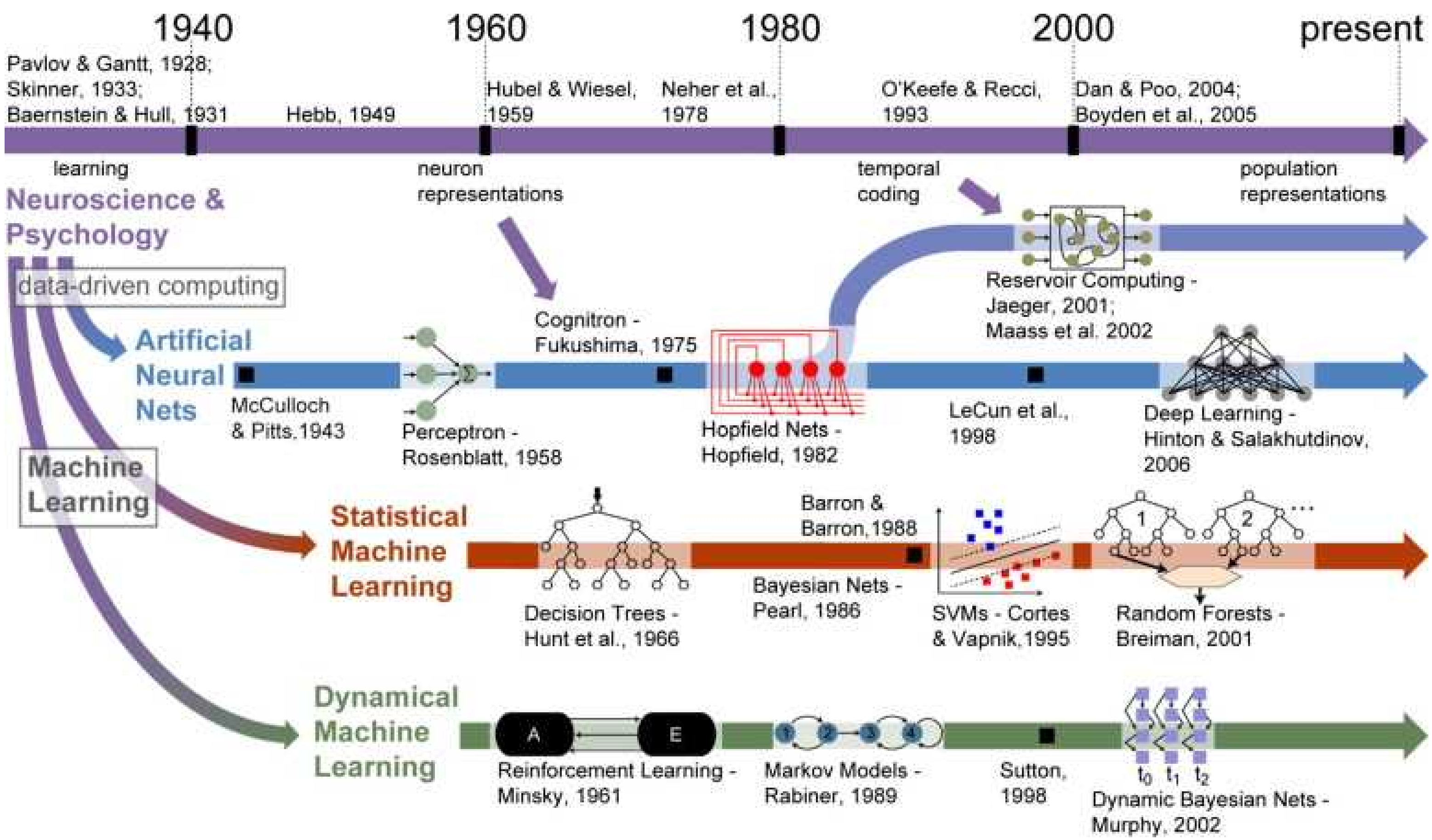
Cycle of computing scaling already has begun to influence neuroscience



Even if Moore's Law ends, computing ***will*** continue to scale to be smarter

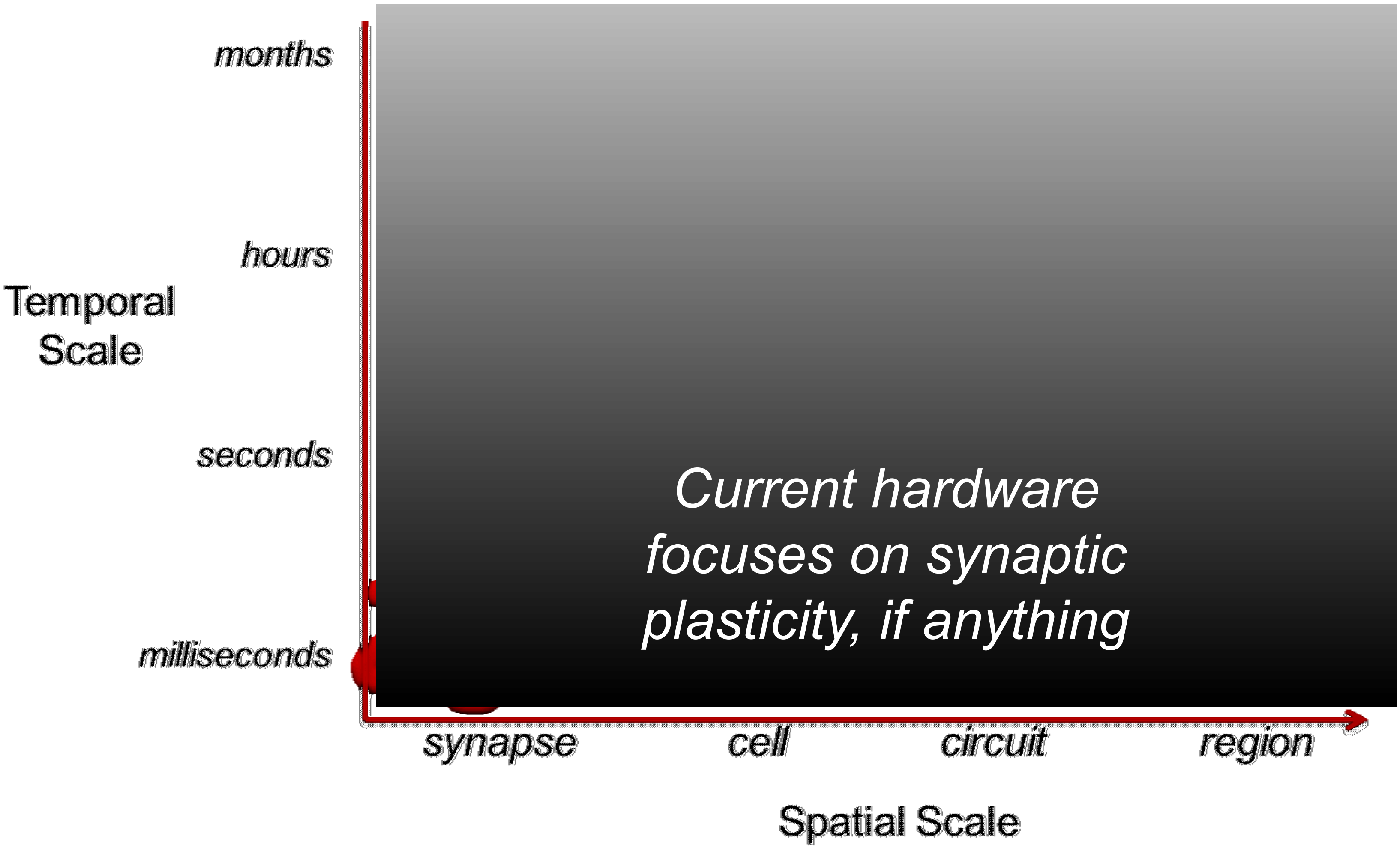


The reservoir of known neuroscience untapped for computing inspiration is *enormous*



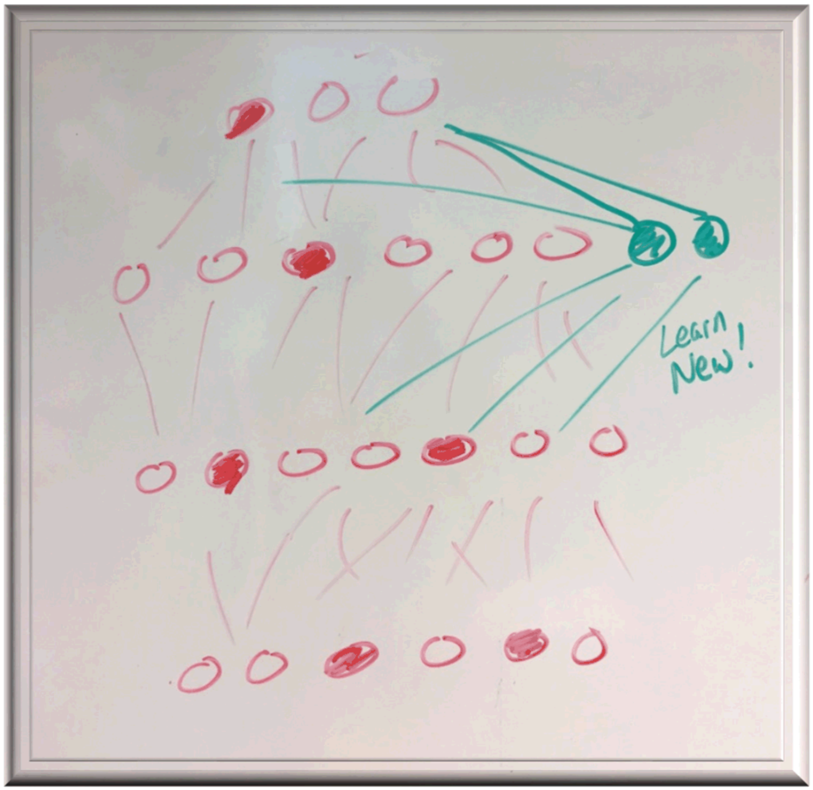
James, et al., BICA 2017

The brain has many mechanisms for adaptation; which are important for computing?



There are different algorithmic approaches to neural learning

- *In situ* adaptation
  - Incorporate "new" forms of known neural plasticity into existing algorithms
- *Ex situ* adaptation
  - Design entirely new algorithms or algorithmic modules to provide cognitive learning abilities



Neurogenesis Deep Learning

Neurogenesis Deep Learning

Extending deep networks to accommodate new classes

Timothy J. Draelos*, Nadine E. Miner*, Christopher C. Lamb*, Jonathan A. Cox*^, Craig M. Vineyard*, Kristofor D. Carlson*, William M. Severa*, Conrad D. James*, and James B. Aimone*

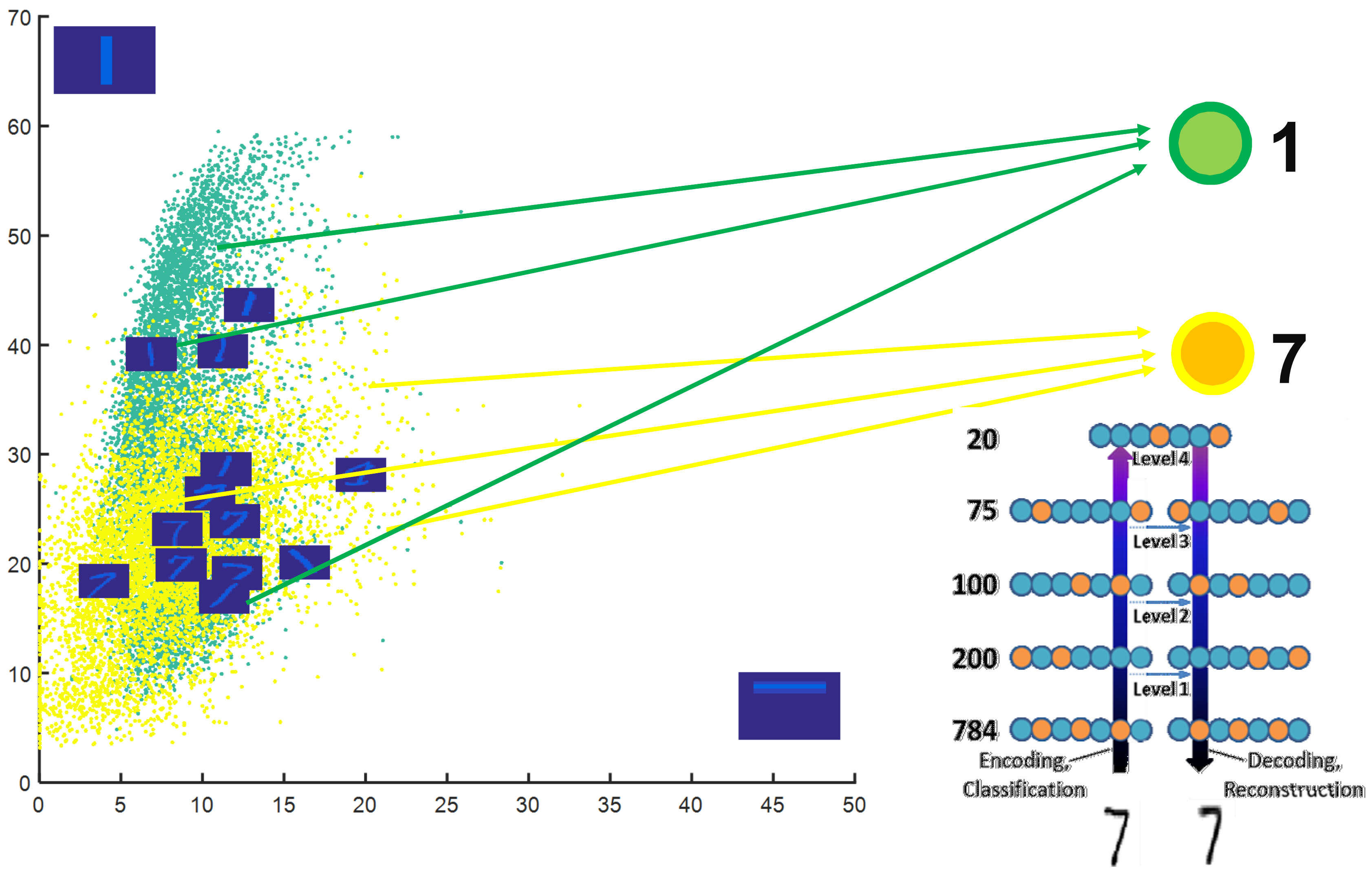
*Sandia National Laboratories, Albuquerque NM, 87185 USA

{tjdrael, nrminer, cclamb, cmviney, kdcarls, wmsever, cdjame, jbaimon}@sandia.gov

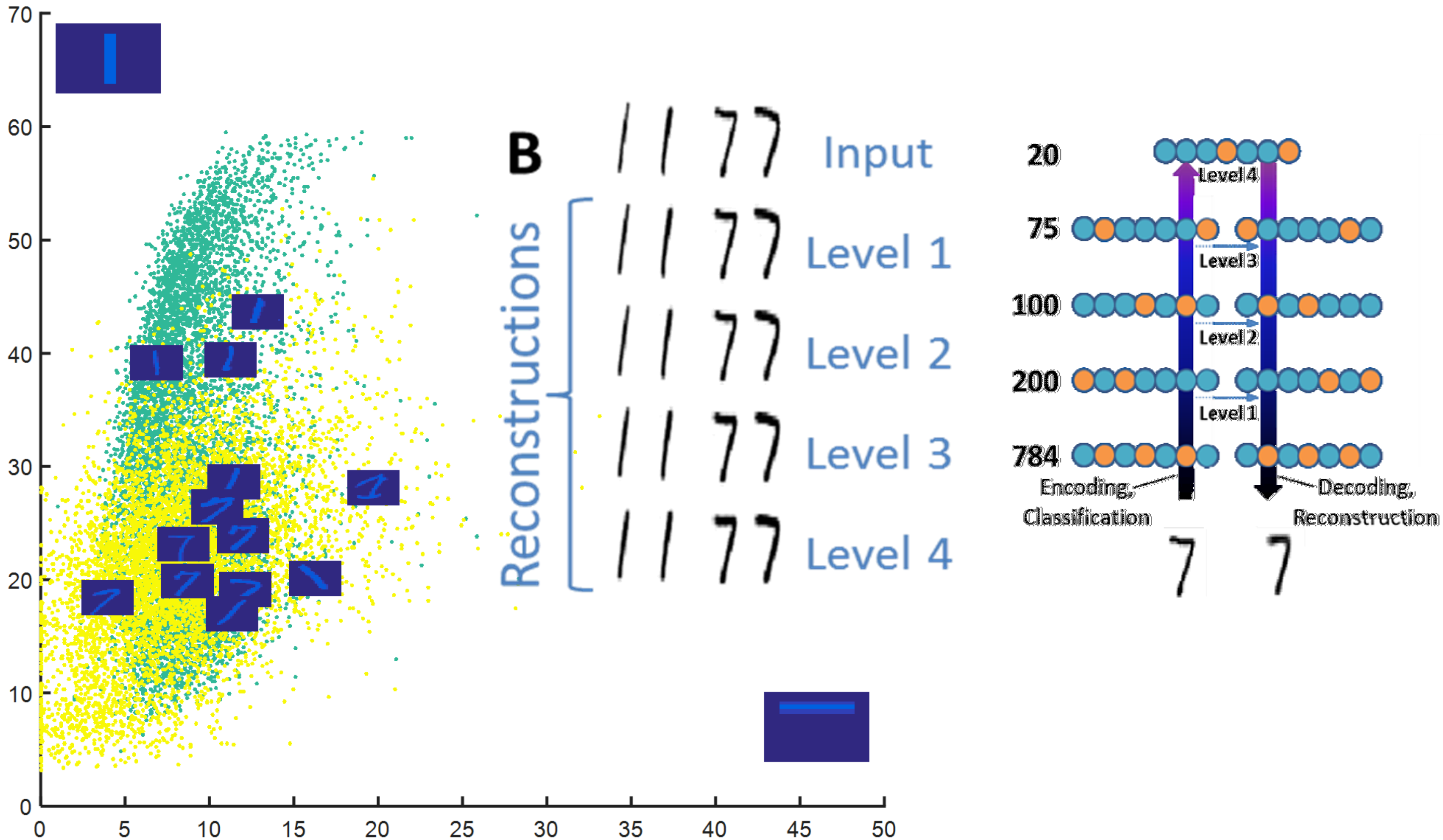
^ Present Address: Qualcomm Corporation, San Diego, CA USA

joncox@alum.mit.edu

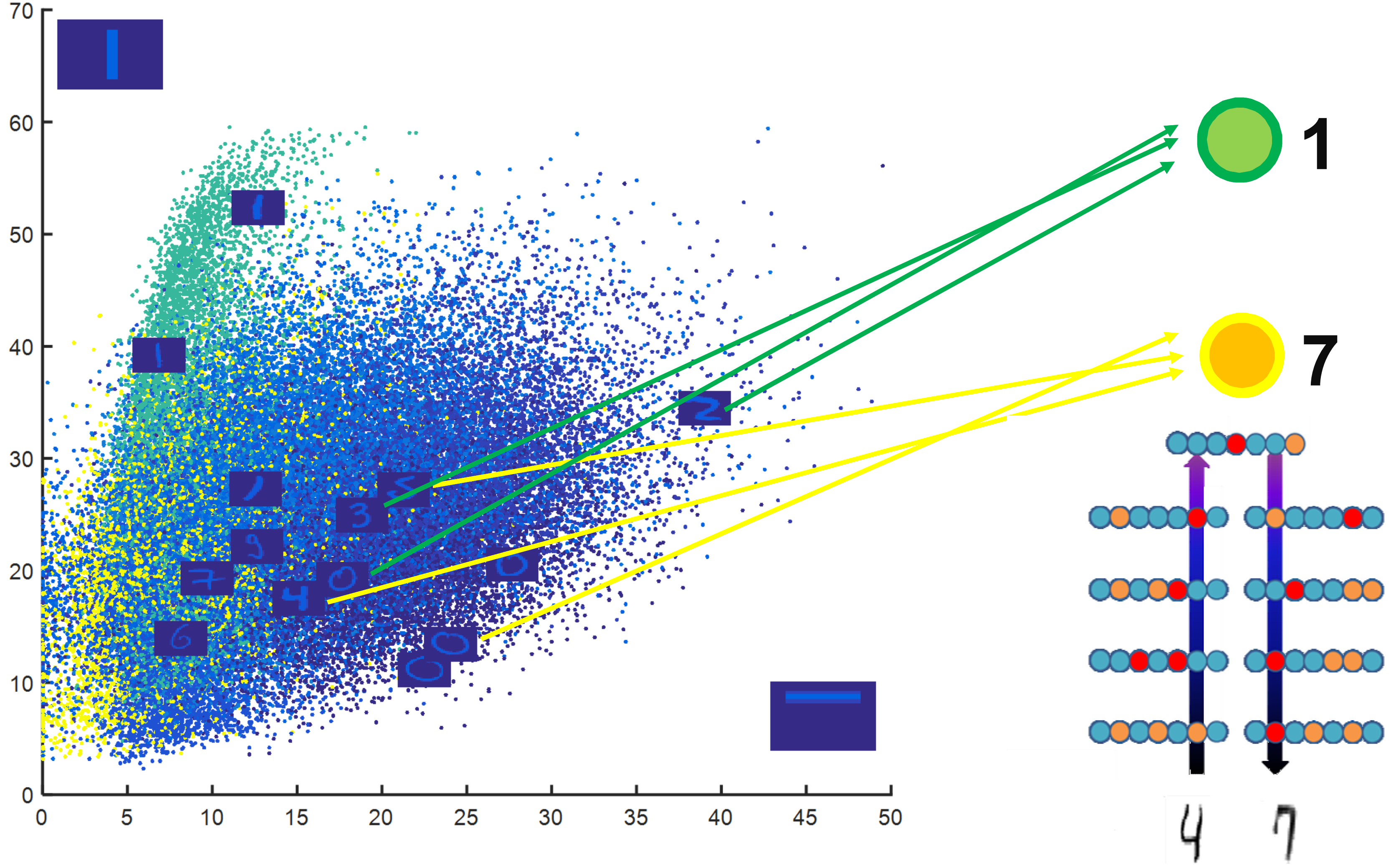
Deep Networks are a function of training sets



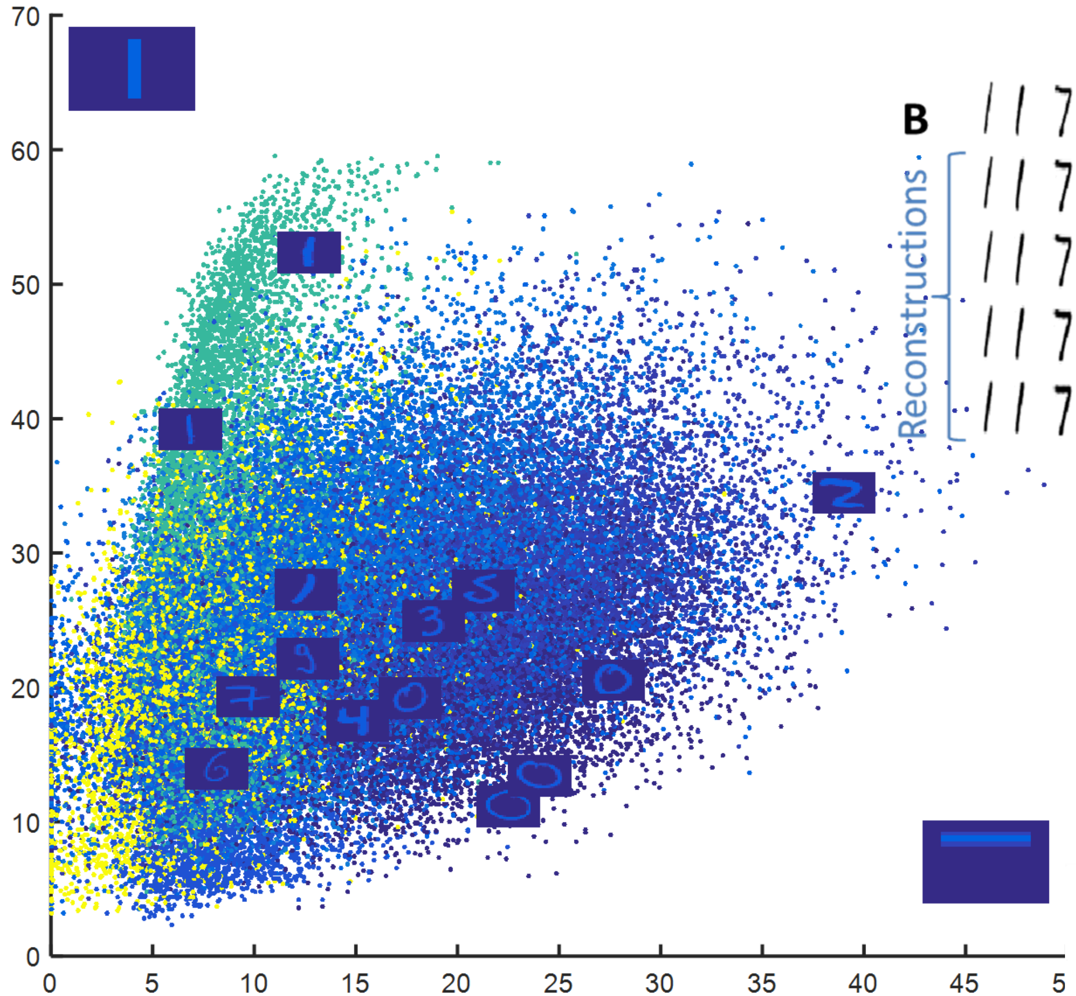
Deep Networks are a function of training sets

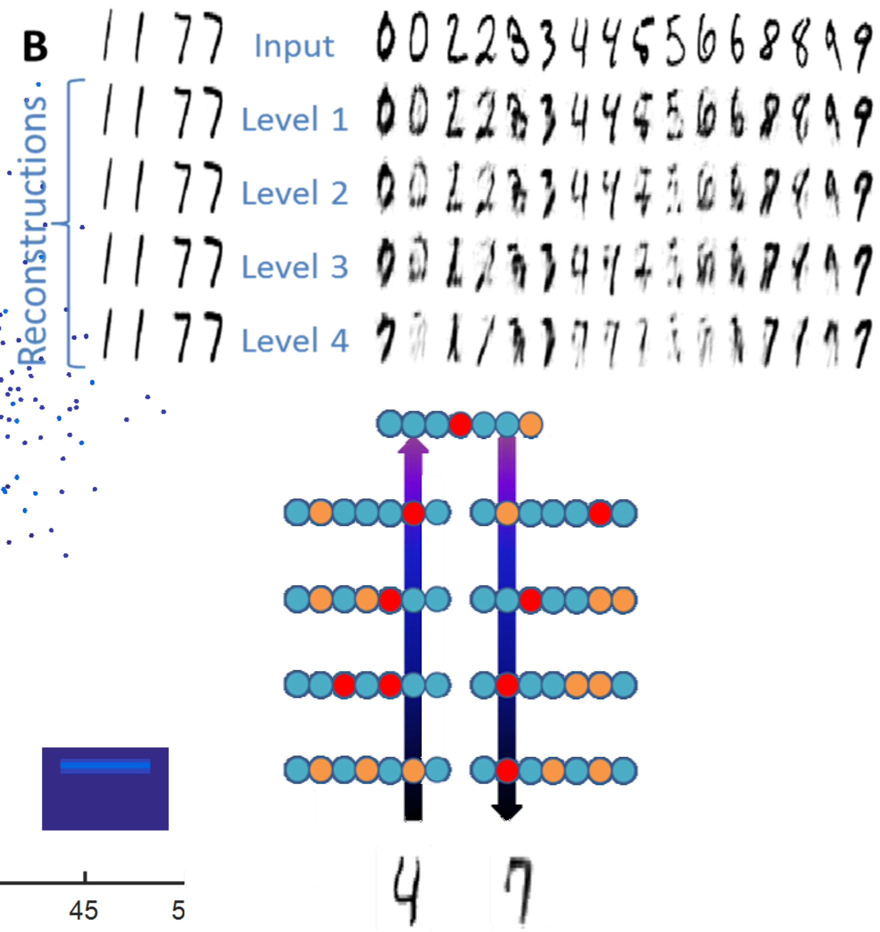


Deep networks often struggle to generalize outside of training domain

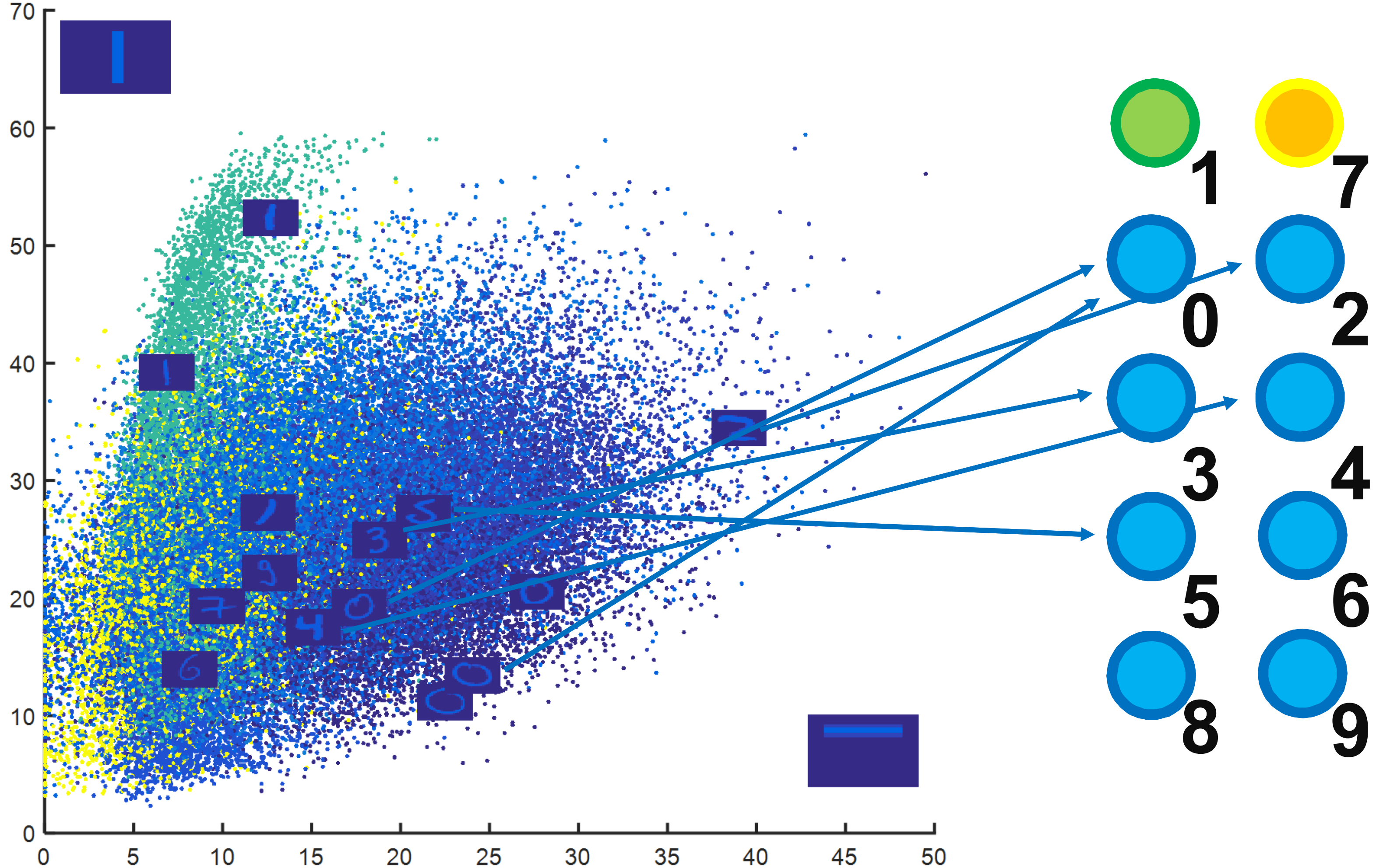


Deep networks often struggle to generalize outside of training domain

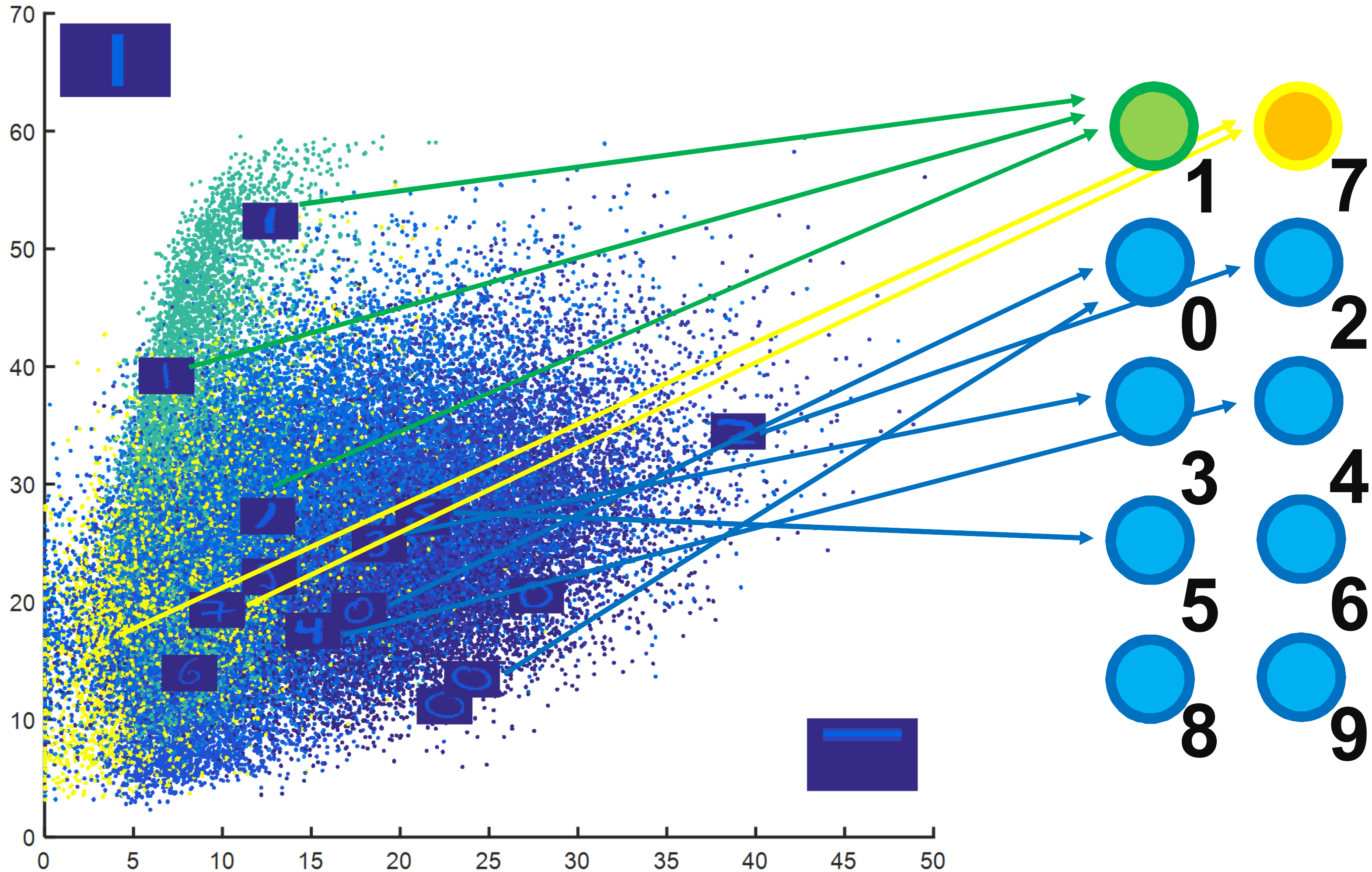




We want to adapt algorithms to adapt to new classes and information…

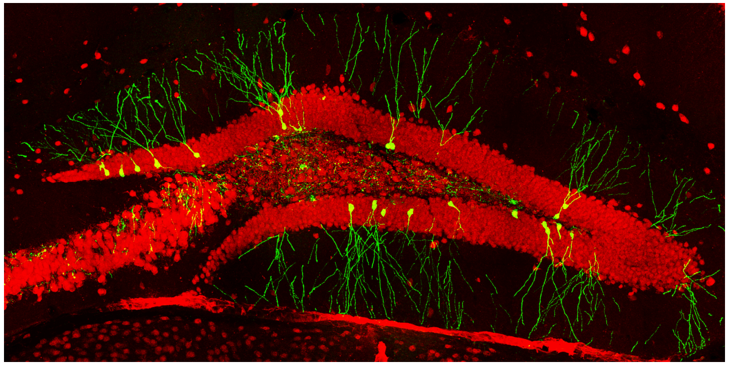


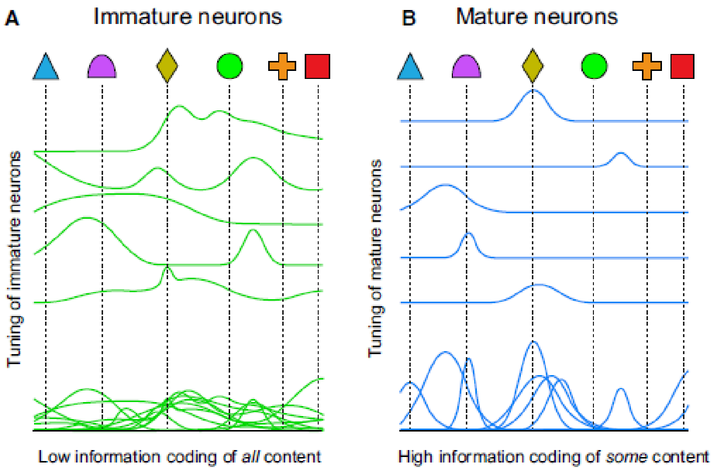
… while preserving old classes



Neurogenesis can be used to capture new information without disrupting old information

- Brain incorporates new neurons in a select number of regions
  - Particularly critical for novelty detection and encoding of new information
  - “Young” hippocampal neurons exhibit increased plasticity (learn more) and are dynamic in their representations
  - “Old” hippocampal neurons appear to have reduced learning and maintain their representations
  - Cortex does **not** have neurogenesis (or similar mechanisms) in adult-hood, but does during development



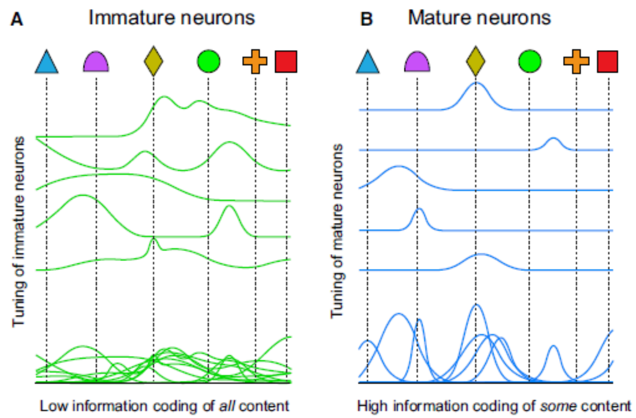


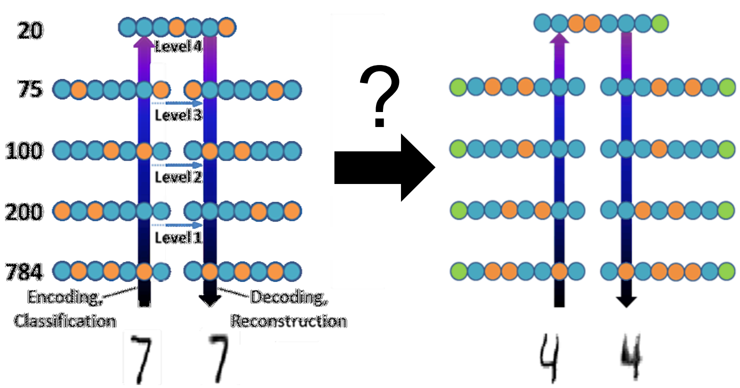
Aimone et al., Neuron 2011

Neurogenesis can be used to capture new information without disrupting old information

- Brain incorporates new neurons in a select number of regions

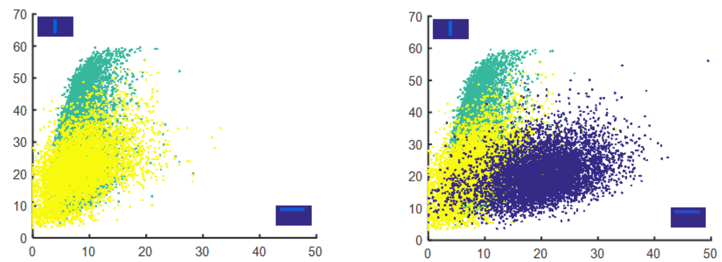
- Hypothesis: Can new neurons be used to facilitate adapting deep learning?

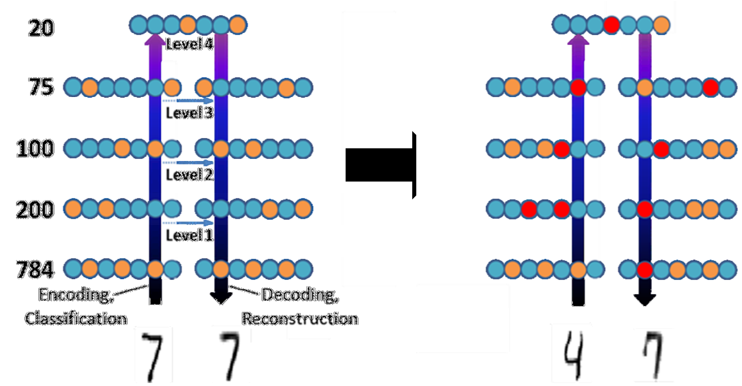




Neurogenesis can be used to capture new information without disrupting old information

- Brain incorporates new neurons in a select number of regions
- Hypothesis: Can new neurons be used to facilitate adapting deep learning?
- Neurogenesis Deep Learning Algorithm
  - Stage 1: Check autoencoder reconstruction to ensure appropriate representations



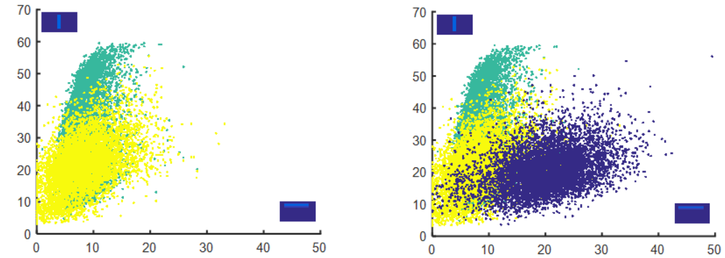


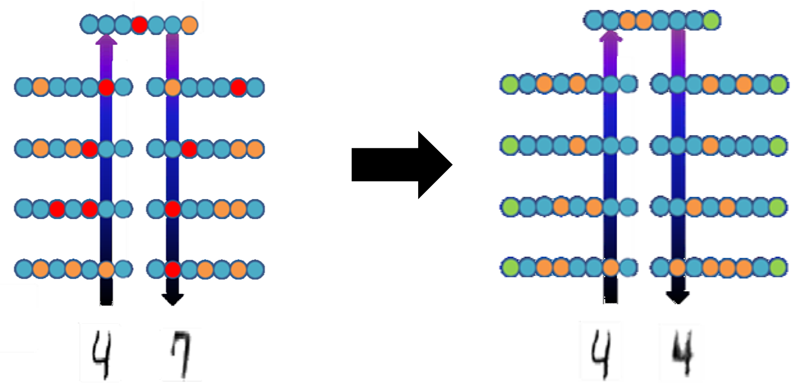
Neurogenesis can be used to capture new information without disrupting old information

Sandia National Laboratories

CCR Center for Computing Research

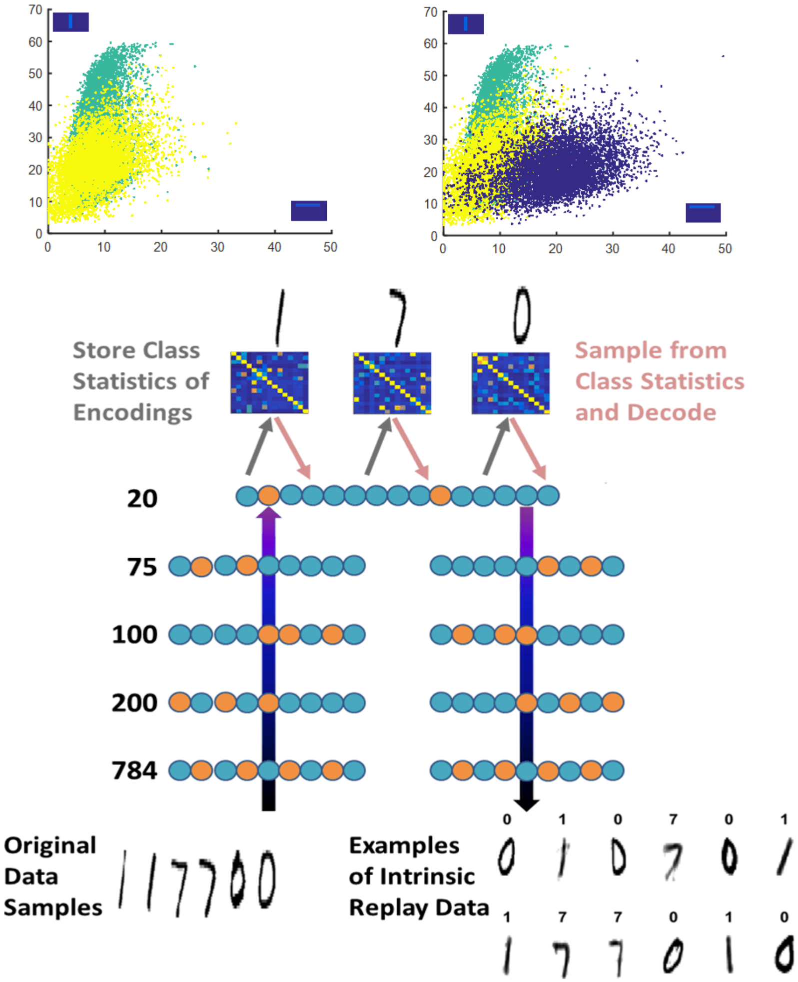
- Brain incorporates new neurons in a select number of regions
- Hypothesis: Can new neurons be used to facilitate adapting deep learning?
- Neurogenesis Deep Learning Algorithm
  - Stage 1: Check autoencoder reconstruction to ensure appropriate representations
  - Stage 2: If mismatch, add and train new neurons
    - Train ***new*** nodes with novel inputs coming in (reduced learning for existing nodes)



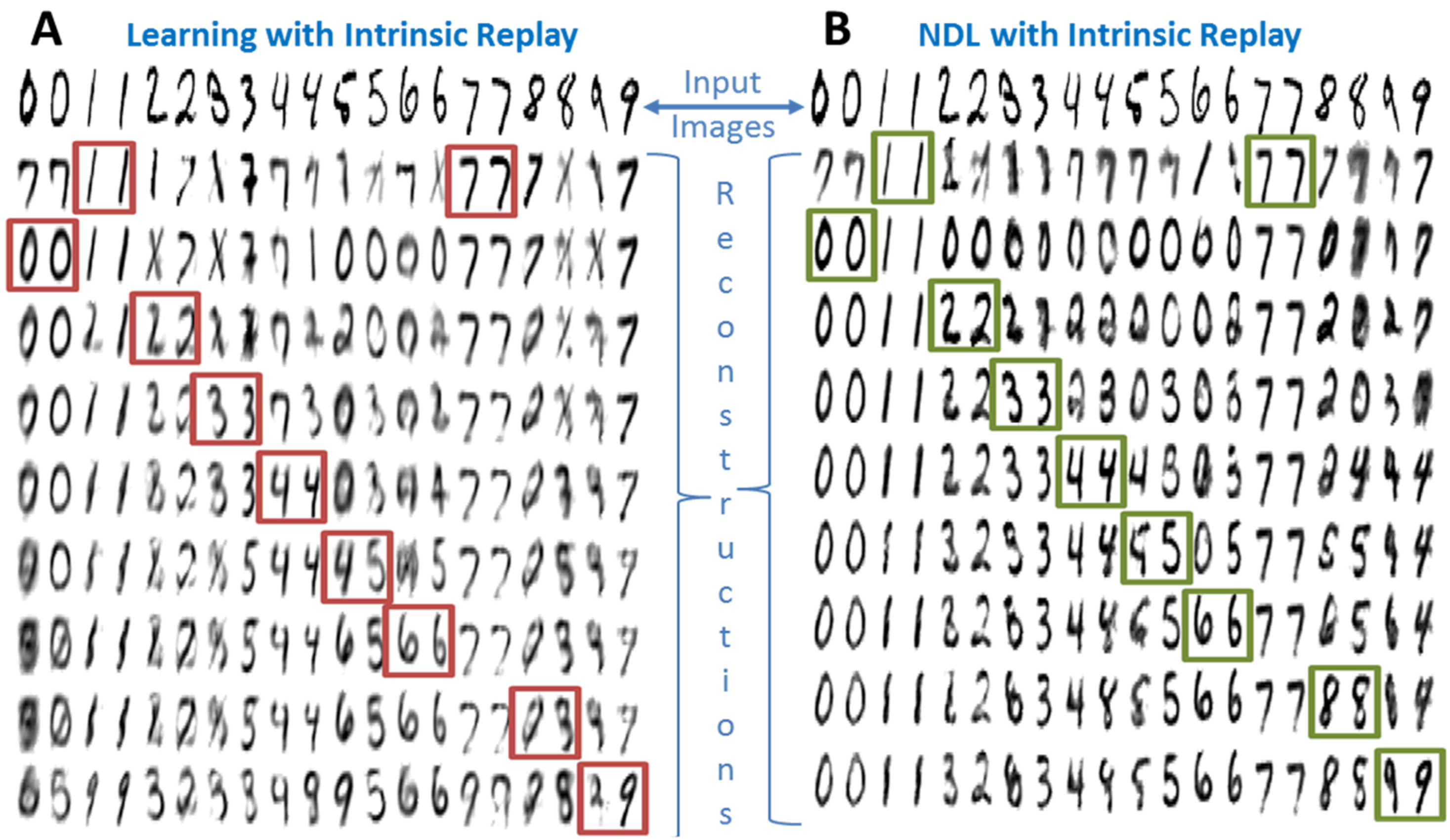


Neurogenesis can be used to capture new information without disrupting old information

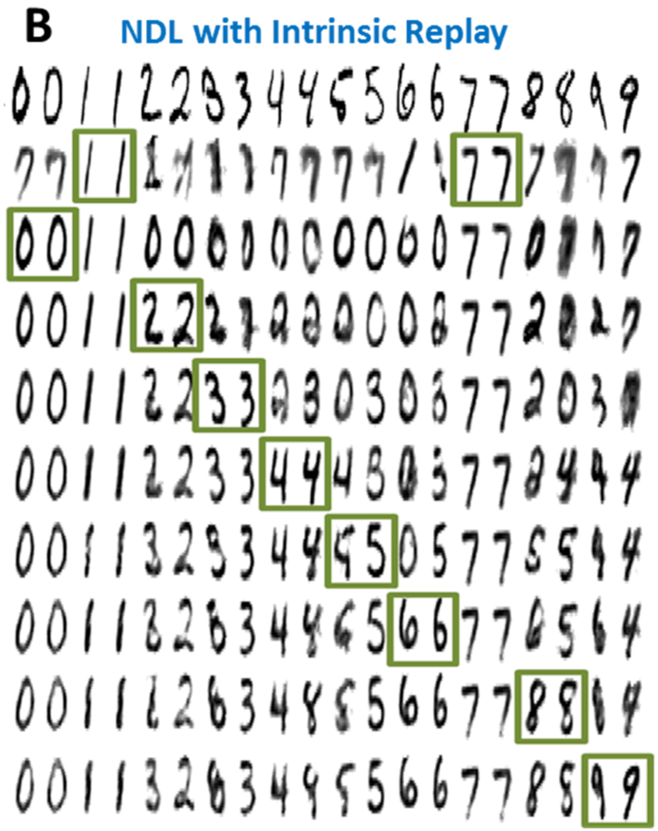
- Brain incorporates new neurons in a select number of regions
- Hypothesis: Can new neurons be used to facilitate adapting deep learning?
- Neurogenesis Deep Learning Algorithm
  - Stage 1: Check autoencoder reconstruction to ensure appropriate representations
  - Stage 2: If mismatch, add and train new neurons
    - Train ***new*** nodes with novel inputs coming in (reduced learning for existing nodes)
    - ***Intrinsically replay*** "imagined" training samples from top-level statistics to fine tune representations
  - Stage 3: Repeat neurogenesis until reconstructions drop below error thresholds

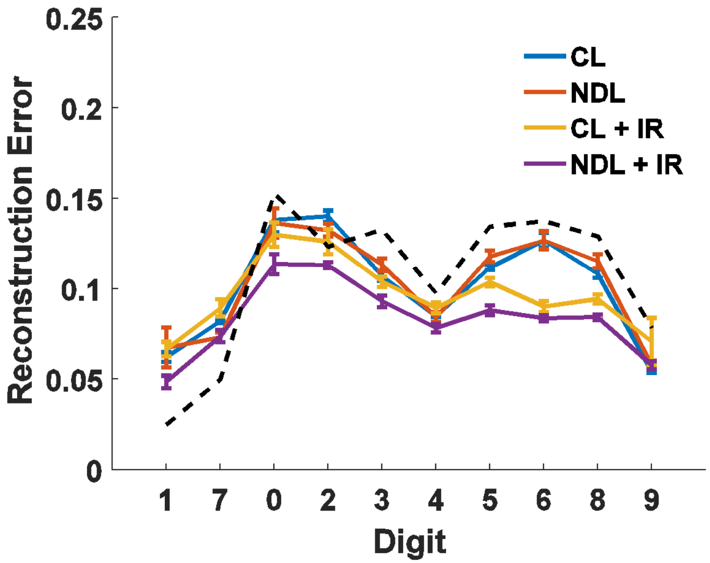


Neurogenesis algorithm effectively balances stability and plasticity



Neurogenesis algorithm effectively balances stability and plasticity

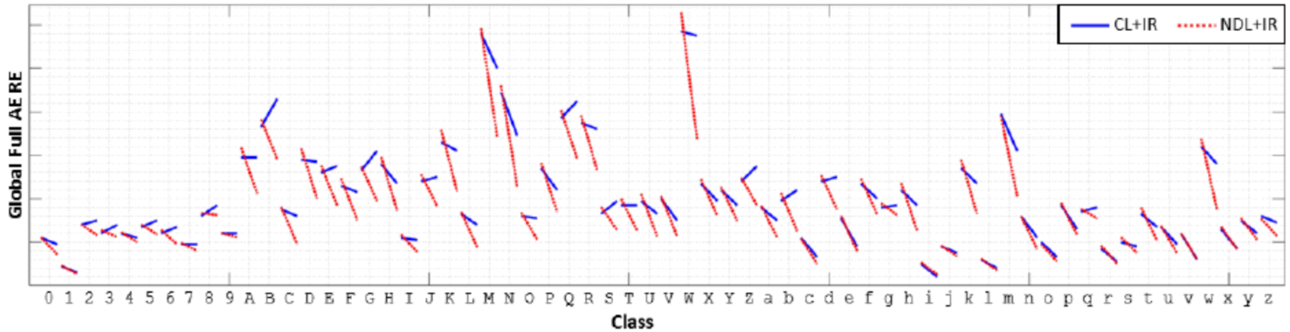


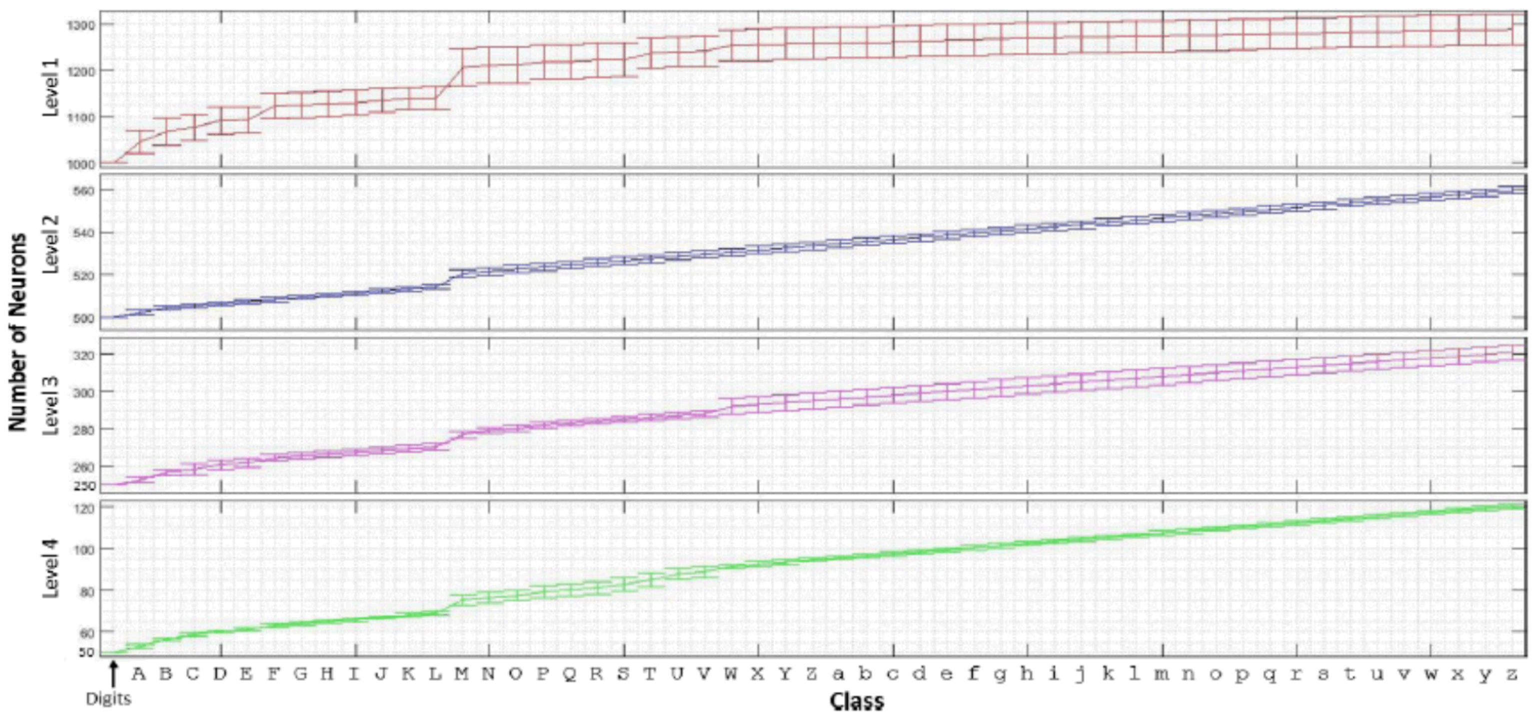


NDL applied to NIST data set

Sandia National Laboratories

CCR Center for Computing Research

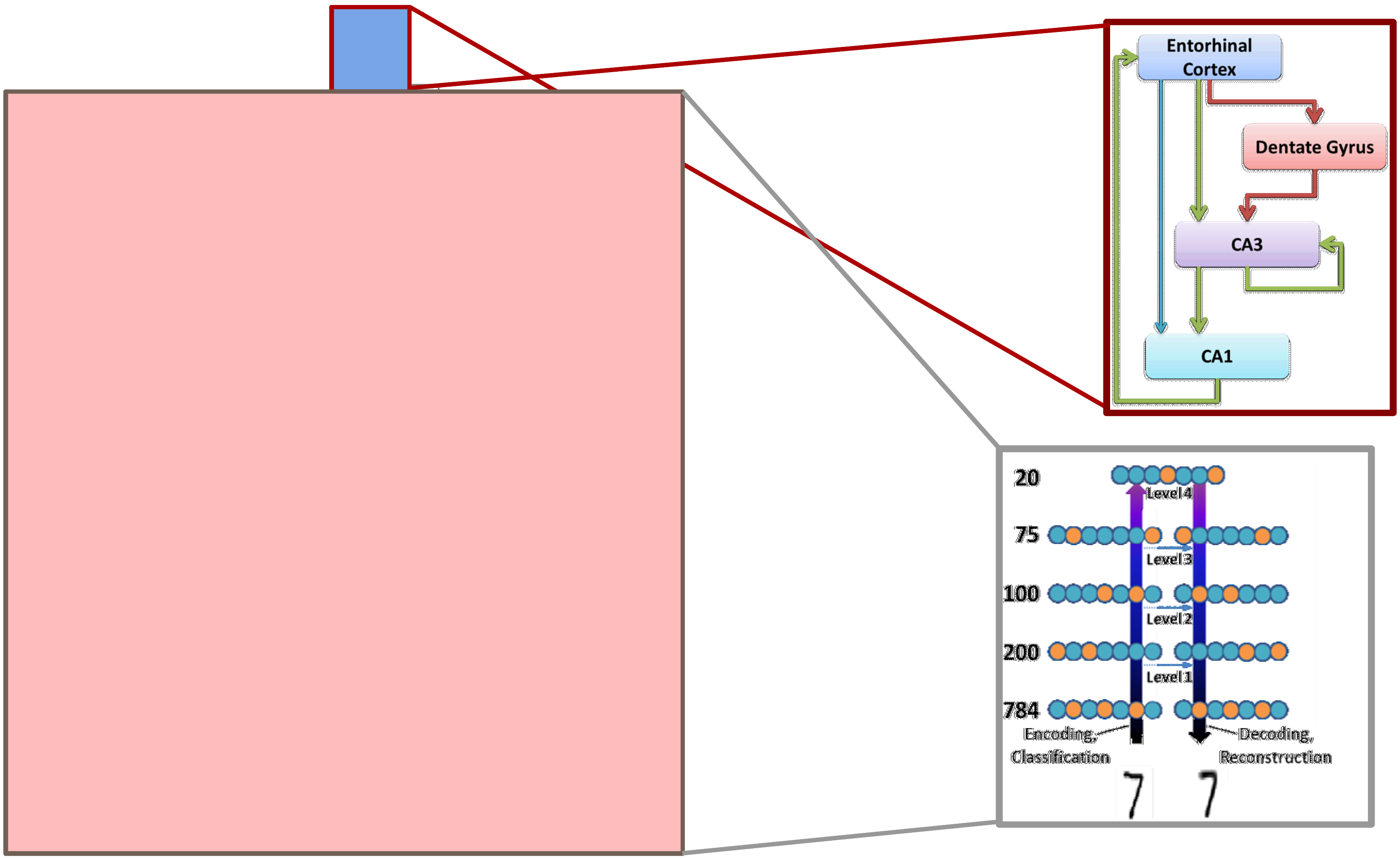




A New View of the Hippocampus

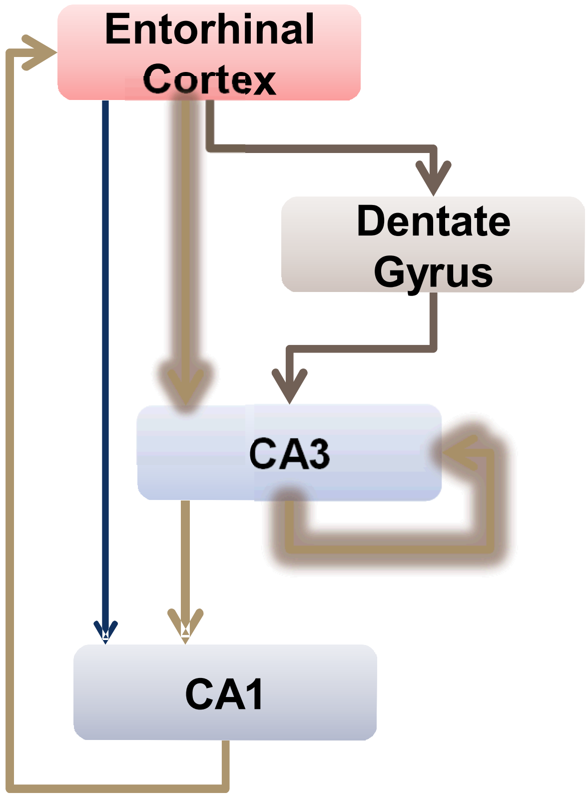
*William Severa, Kris Carlson,
Craig Vineyard, Frances Chance*

Deep learning ≈ Cortex
What ≈ Hippocampus?



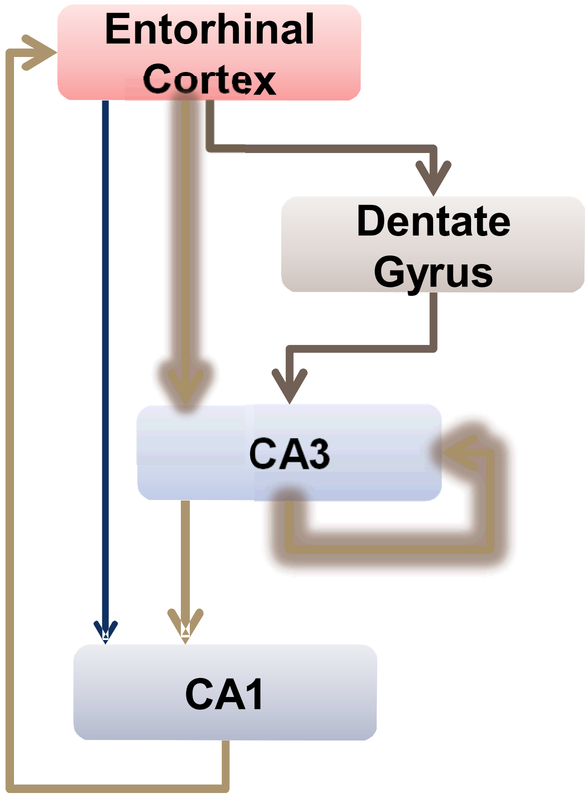
Canonical Hippocampus Algorithm

- EC has high level cortical representations
- **Learning**:
  - DG performs pattern separation
  - DG drives new attractor in CA3 network (CA3->CA3 recurrent conns) and EC->CA3 connections
  - CA1 does something…
- **Retrieval**:
  - DG bypassed
  - EC activates CA3 neurons
  - CA3 -> CA3 connections auto-retrieve learned attractor
  - CA1 reads out stable CA3 state



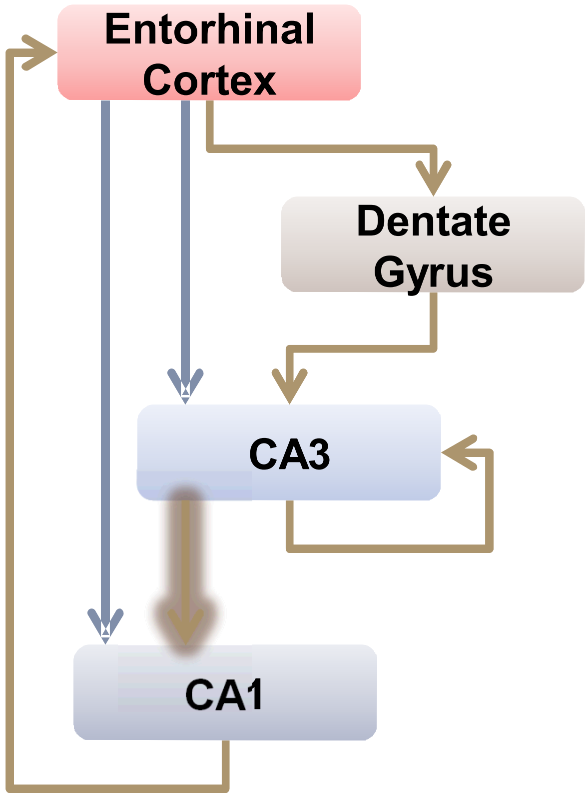
Canonical Hippocampus Algorithm

- Problems with canonical model
  - No clear function for CA1 (and CA3->CA1 plasticity is not obviously used)
  - EC -> CA1 pathway not involved
  - Bulk of learning is within CA3->CA3 autoencoder, which is computationally difficult
  - No good reason for DG plasticity
  - DG only involved in encoding, what is it doing during retrieval?



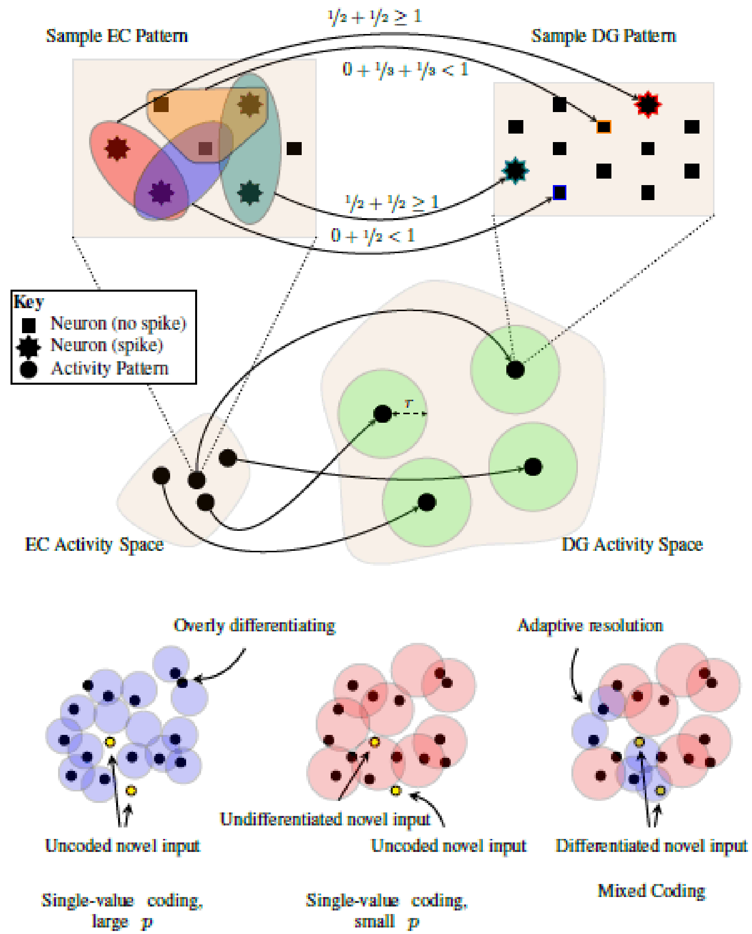
New View of Hippocampus

- EC has high level cortical representations
- DG provides decorrelated mapping of EC
- CA3 is “static” dynamical system
  - Limited recurrent plasticity
  - Recurrent dynamics provide a number of path attractors (*orbits*)
  - EC->CA3 inputs are weak and modulatory
    - shift dynamical manifold
    - CA3 Attractors are positioned in context-dependent locations
  - DG inputs “seed” CA3 network which propagates to attractor basin
- CA1 “learns” to read CA3 attractors
  - Context dependent readout



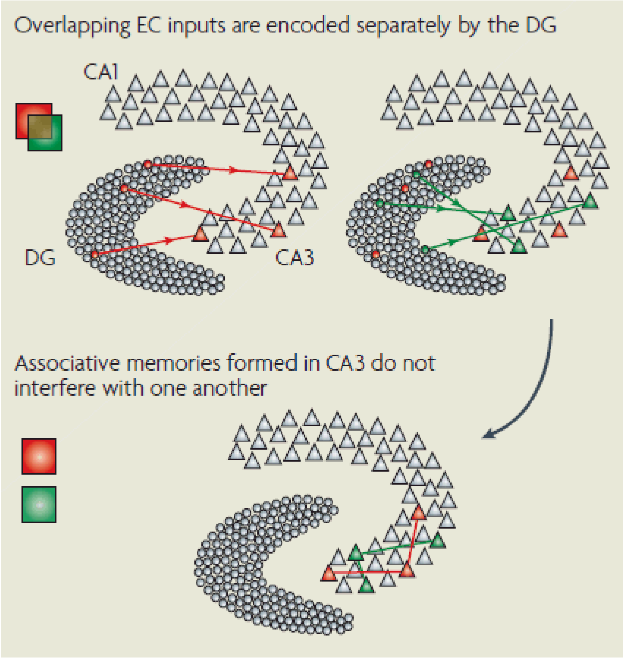
Combinatorial model of DG coding shows mixed coding advantages

- Formal model of DG designed for "pattern separation" with no loss of information
- Constraining EC inputs to have "grid cell" structure sets DG size to biological level of expansion (~10:1)
- Mixed code of broad-tuned (immature) neurons and narrow tuned (mature) neurons confirms predicted ability to encode novel information



William Severa, NICE 2016
Severa et al., Neural Computation, 2017

Classic model of CA3 uses Hopfield-like recurrent network attractors

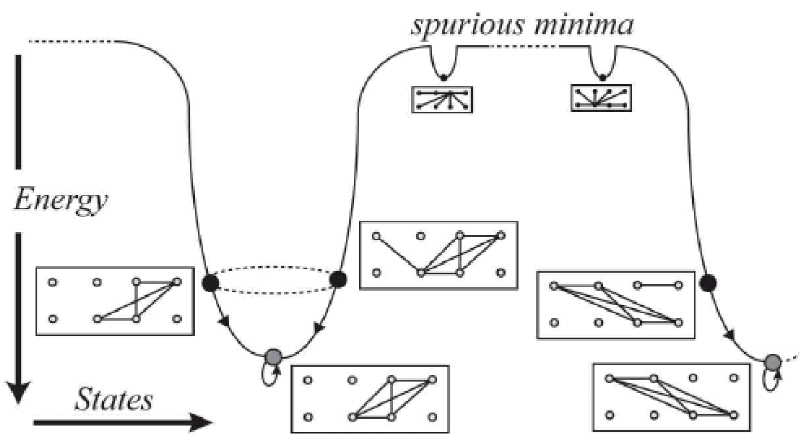


Deng et al., Nat Rev Neuro 2010

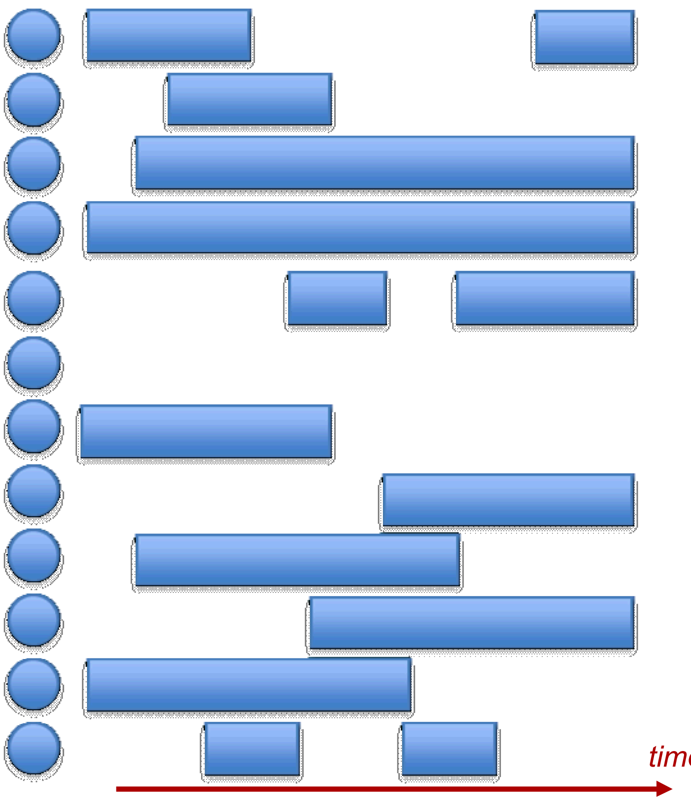
Problems

- "Auto-associative" attractors make more sense in frequency coding regime than in spiking networks
- Capacity of classic Hopfield networks is generally low
- Quite difficult to perform stable one-shot updates to recurrent networks

Moving away from the Hopfield "learned auto-association" function for CA3

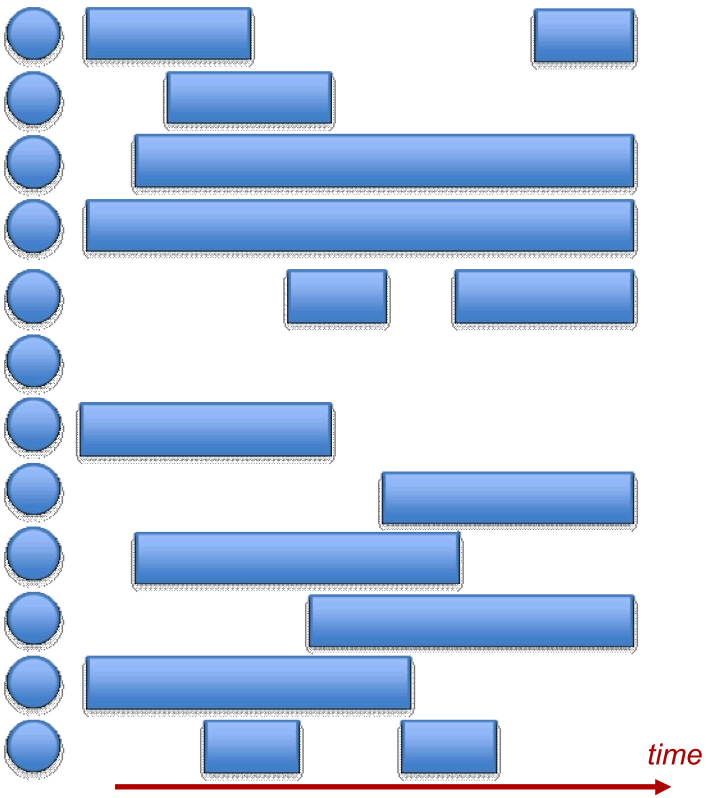


Hillar and Tran, 2014

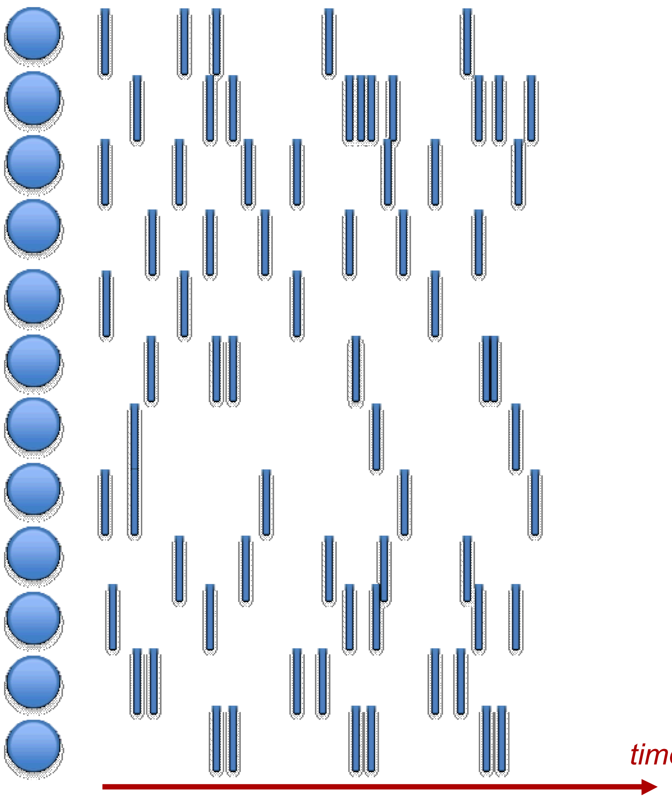


Hopfield dynamics are discrete state transitions

Spiking dynamics are inconsistent with fixed point attractors in associative models

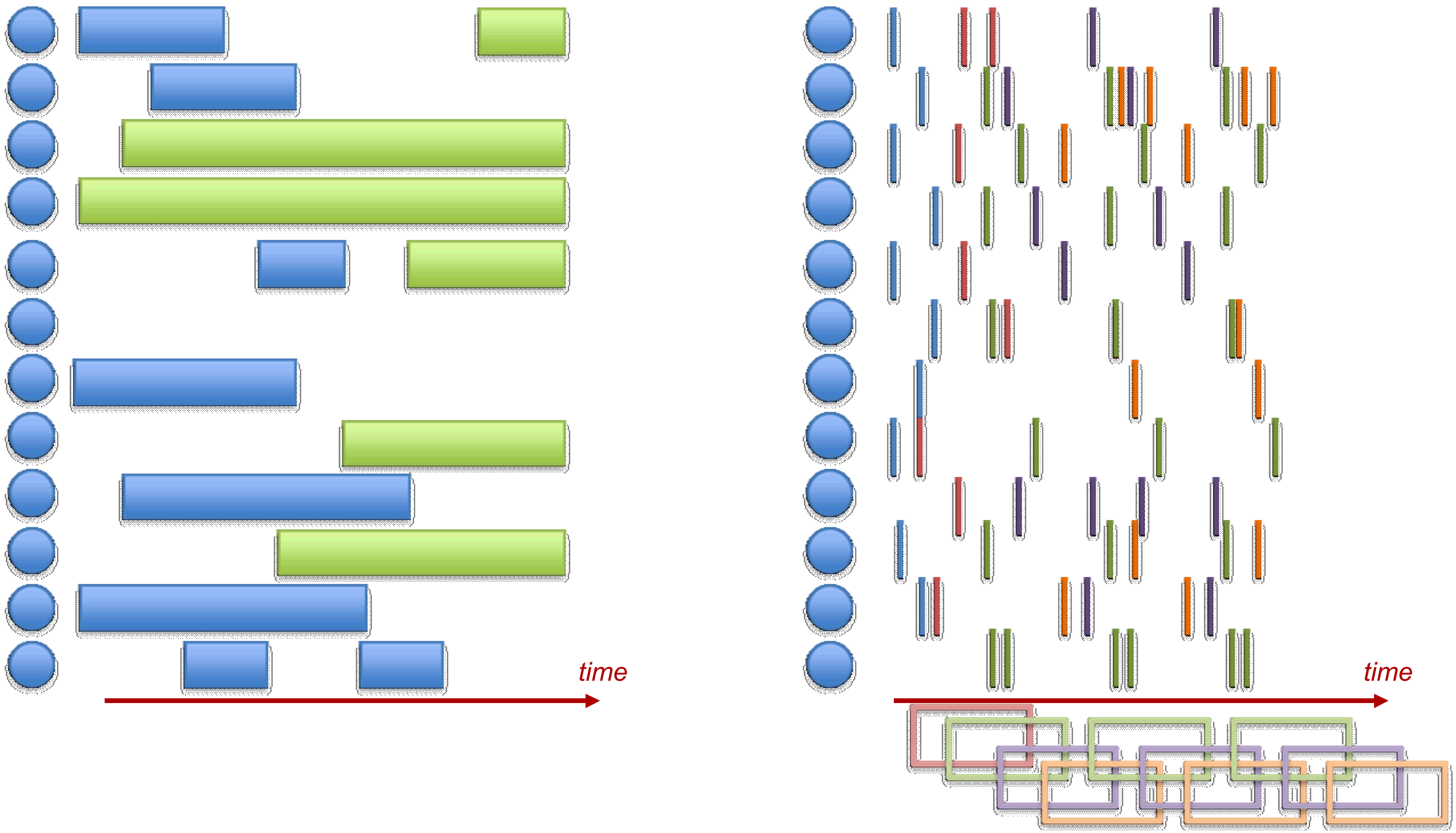


Hopfield dynamics are discrete state transitions



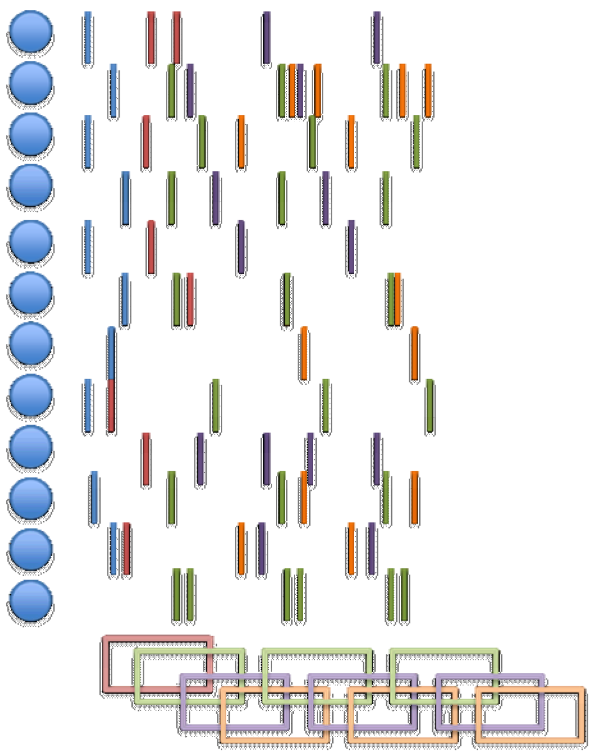
Biology uses sequence of spiking neurons?

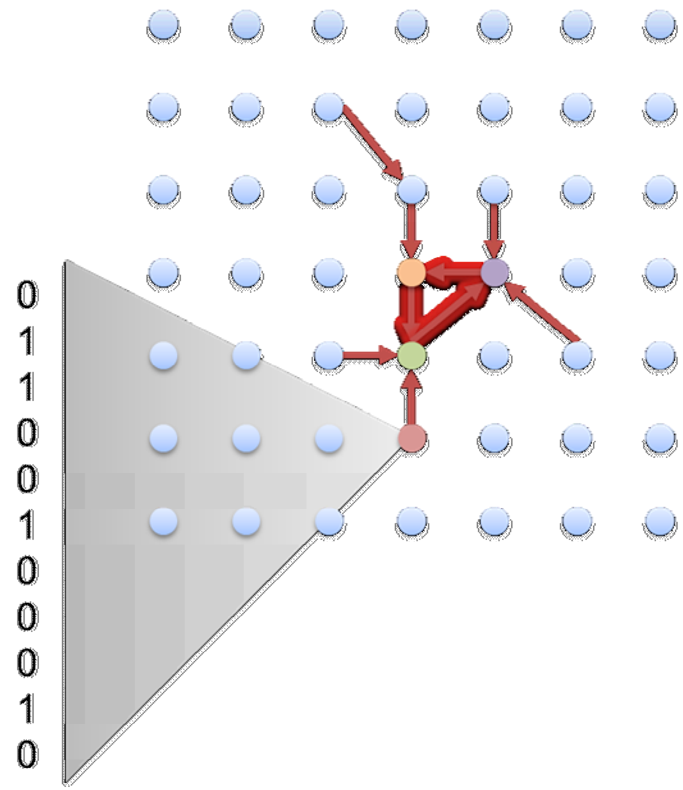
Spiking dynamics are inconsistent with fixed point attractors in associative models



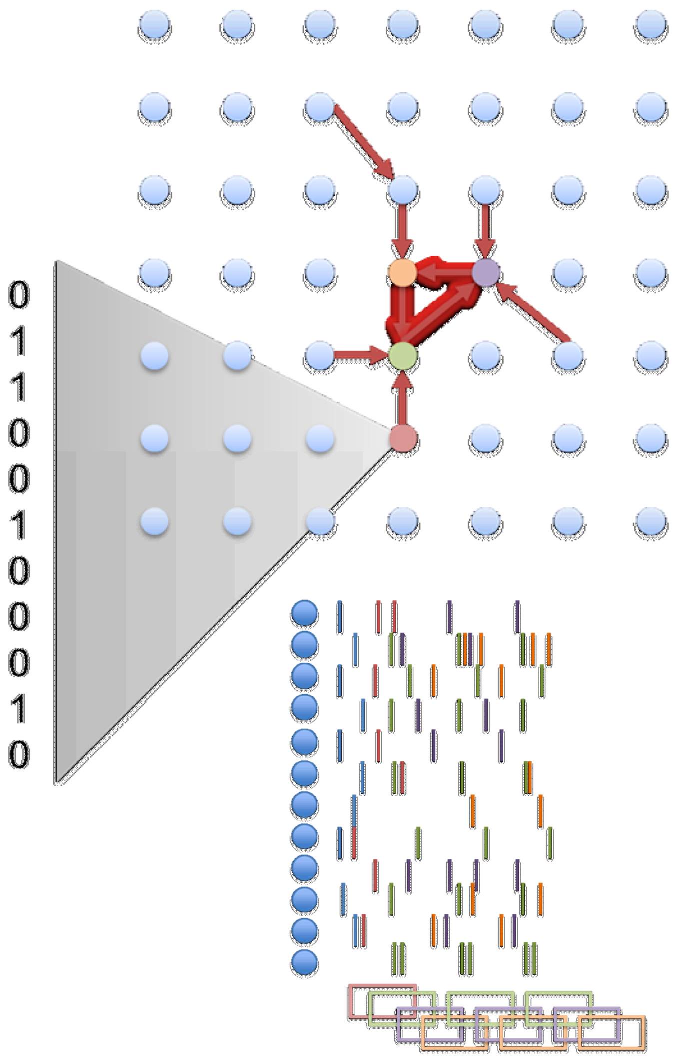
One can see how *sequences* can replace fixed populations

Path attractors, such as *orbits*, are consistent with spiking dynamics





A new dynamical model of CA3



Problems

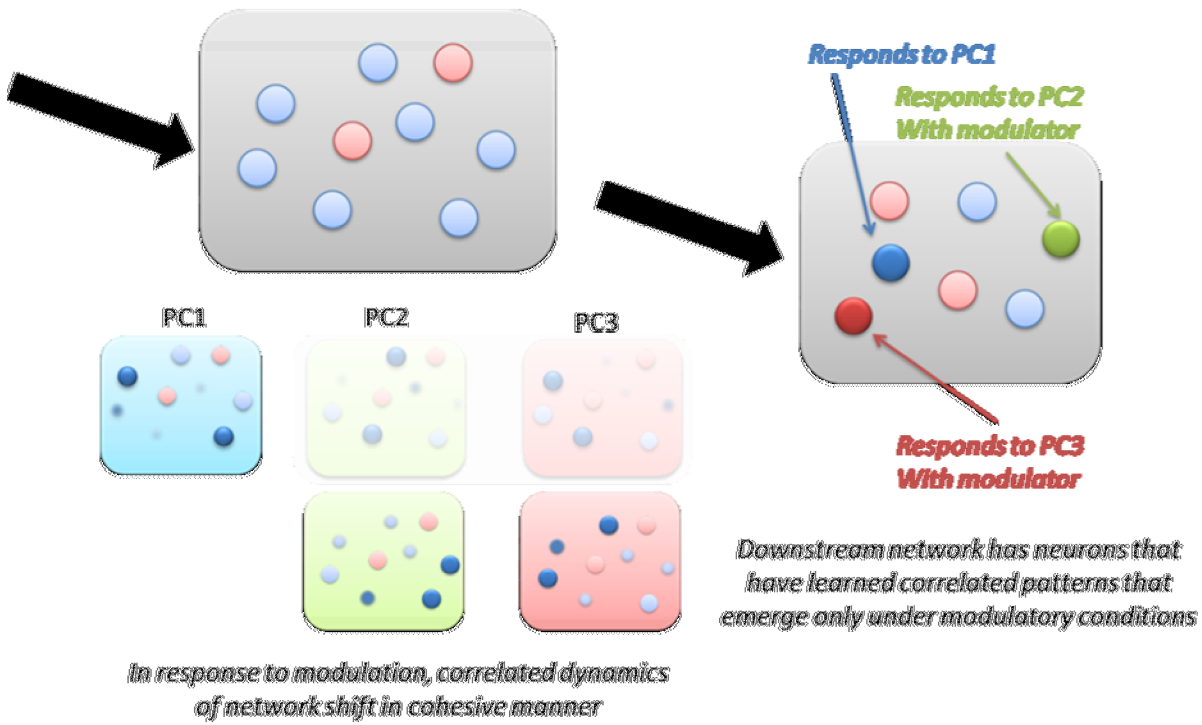
- "Auto-associative" attractors make more sense in frequency coding regime than in spiking networks

Orbits of Spiking Neurons

- Capacity of classic Hopfield networks is generally low

- Quite difficult to perform stable one-shot updates to recurrent networks

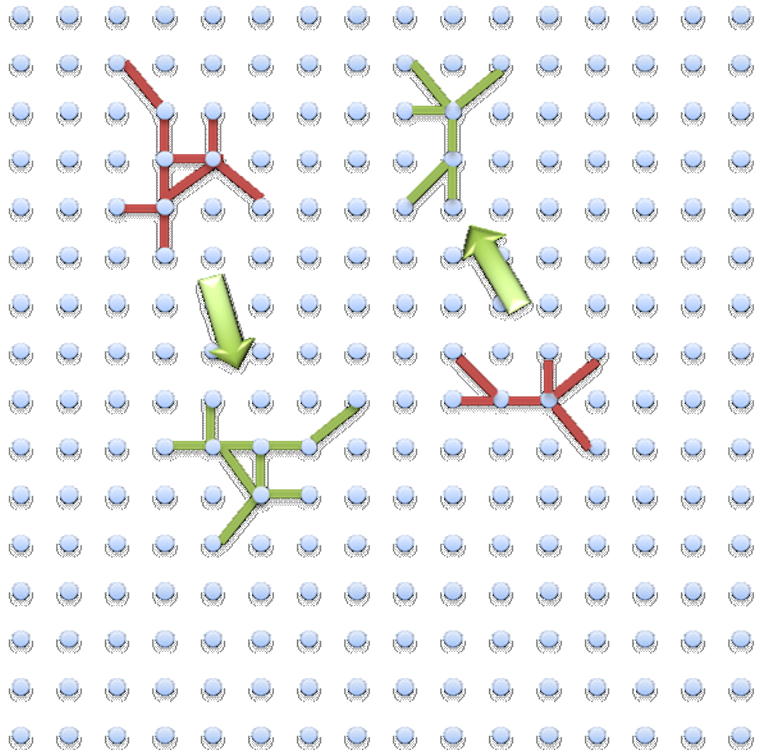
Neuromodulation can shift dynamics of recurrent networks



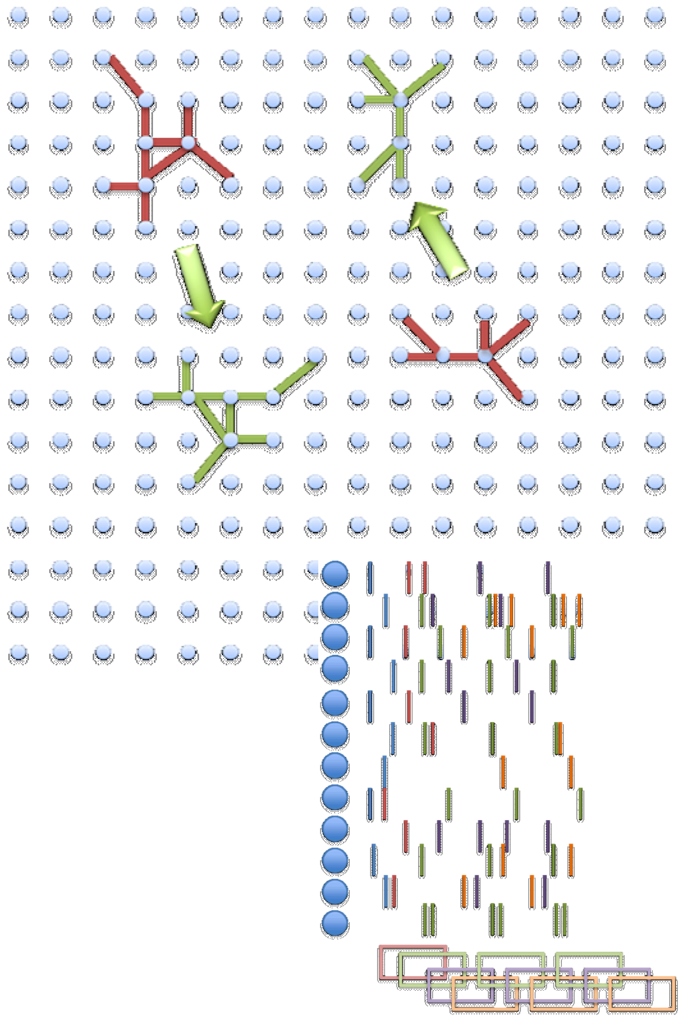
Carlson, Warrender, Severa and Aimone; in preparation

Cortex and subcortical inputs can modulate CA3 attractor access

- Modulation can be provided mechanistically by several sources
  - EC->CA3 inputs will bias some neurons more than others, thus shifting dynamical structure
  - Metabotrophic modulators (e.g., serotonin, acetylcholine) can bias neuronal timings and thresholds, which in turn shifts dynamics in a potentially reversible way
- Attractor network can thus have many “memories”, but only fraction are accessible within each context



A new modulated, dynamical model of CA3



Problems

- "Auto-associative" attractors make more sense in frequency coding regime than in spiking networks

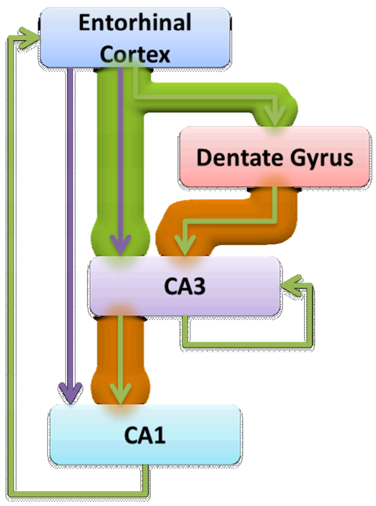
- Capacity of classic Hopfield networks is generally low

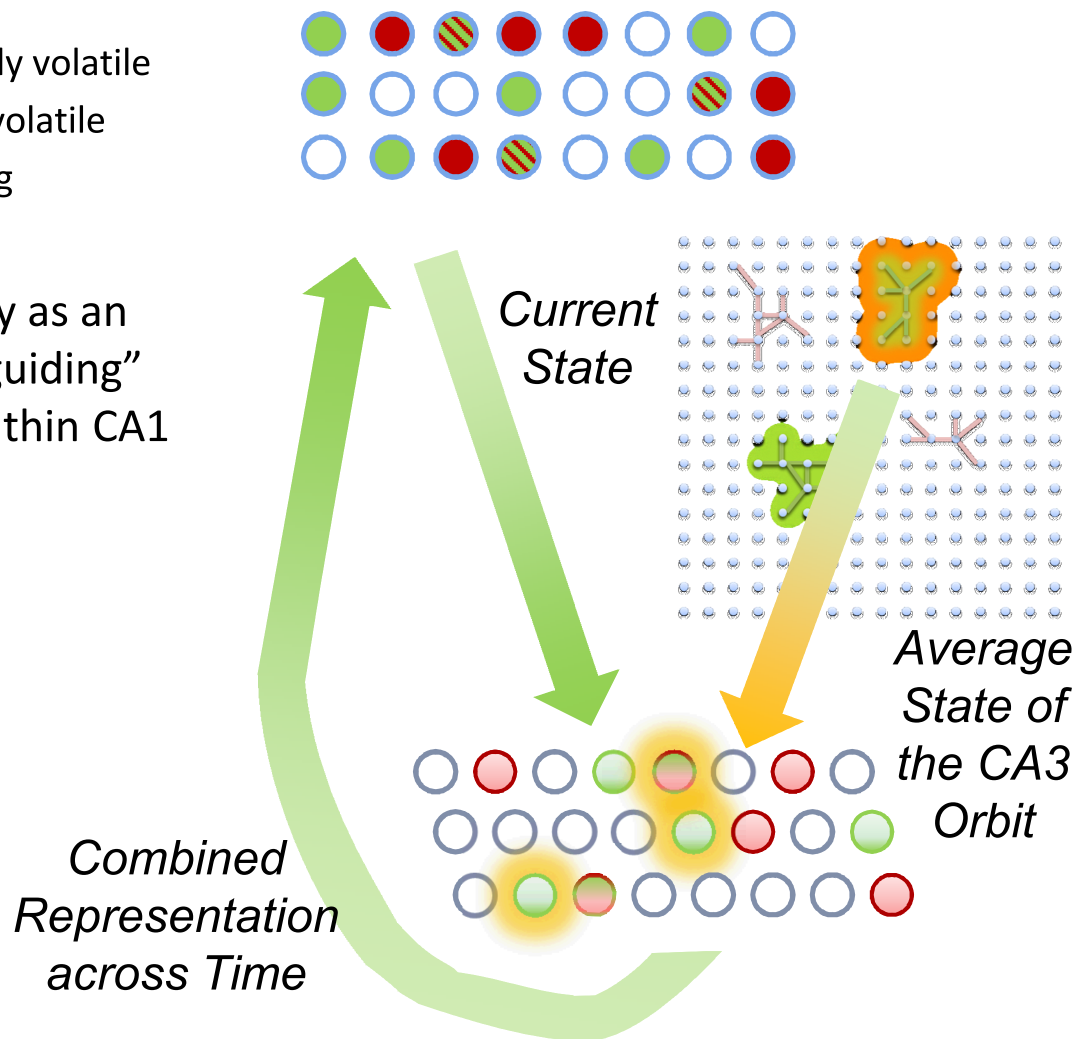
- Quite difficult to perform stable one-shot updates to recurrent networks

CA1 encoding can integrate cortical input with transformed DG/CA3 input

- CA1 plasticity is dramatic
  - Synapses appear to be structurally volatile
  - Representations are temporally volatile
  - Consistent with one-shot learning

- Can consider EC-CA1-EC loosely as an autoencoder, with DG / CA3 "guiding" what representation is used within CA1

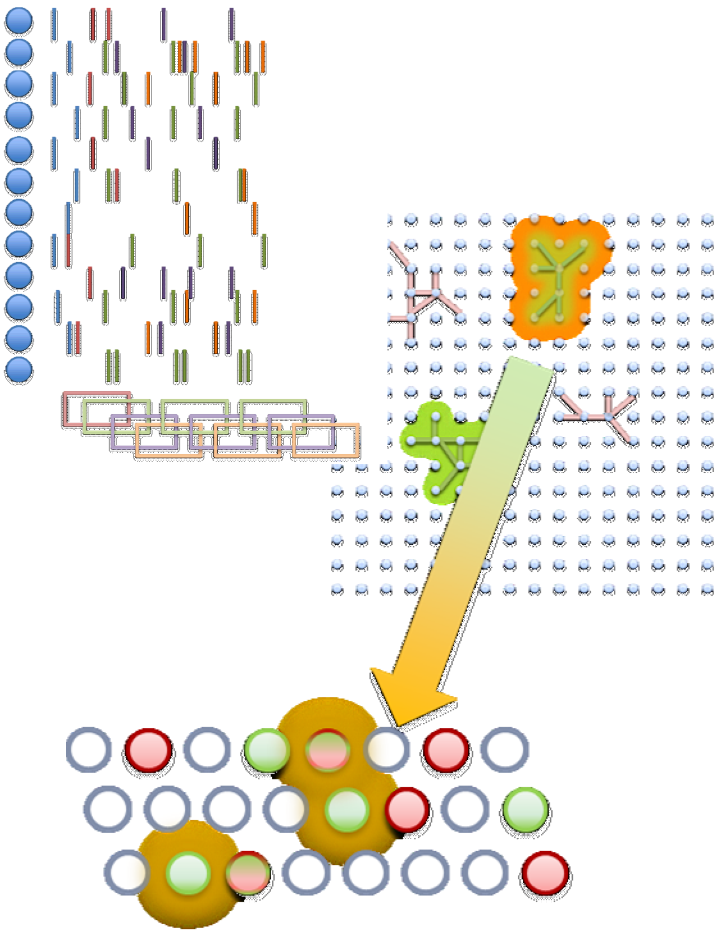




A new modulated, dynamical model of CA3

Sandia National Laboratories

CCR Center for Computing Research



Problems

- "Auto-associative" attractors make more sense in frequency coding regime than in spiking networks

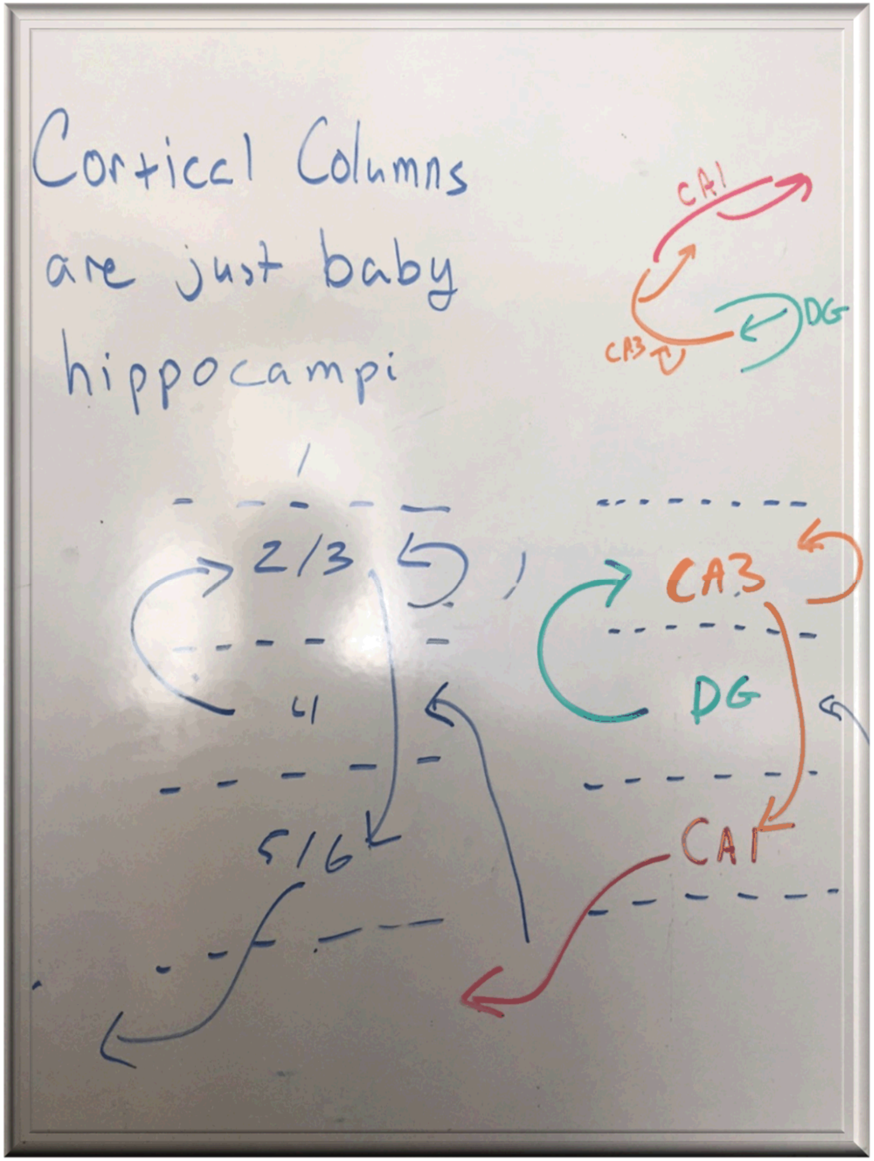
Orbits of Spiking Neurons

- Capacity of classic Hopfield networks is generally low

Context modulation

- Quite difficult to perform stable one-shot updates to recurrent networks

Schaffer Collateral (CA3-CA1) Learning



Thanks!

HAANA Grand Challenge LDRD
DOE NNSA Advanced Simulation and Computing Program

Neurogenesis Deep Learning:

Tim Draelos, Nadine Miner, Chris Lamb, Jonathan Cox, Craig Vineyard, Kris Carlson, William Severa, and Conrad James

Hippocampus Algorithm:

Kris Carlson, William Severa, Ojas Parekh, Frances Chance, and Craig Vineyard