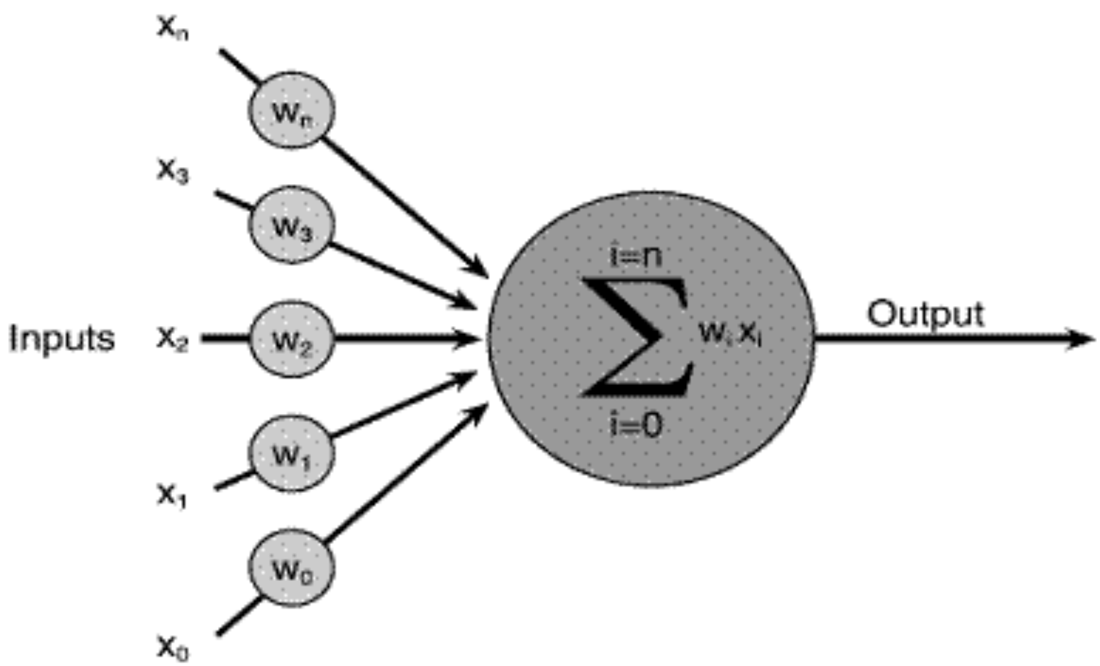
Physical Neural Networks



Jonathan Lamont

November 16, 2015

Contents

- Artificial Neural Network
- History of Development
- AHaH Computation
- Future Goals of AHaH System

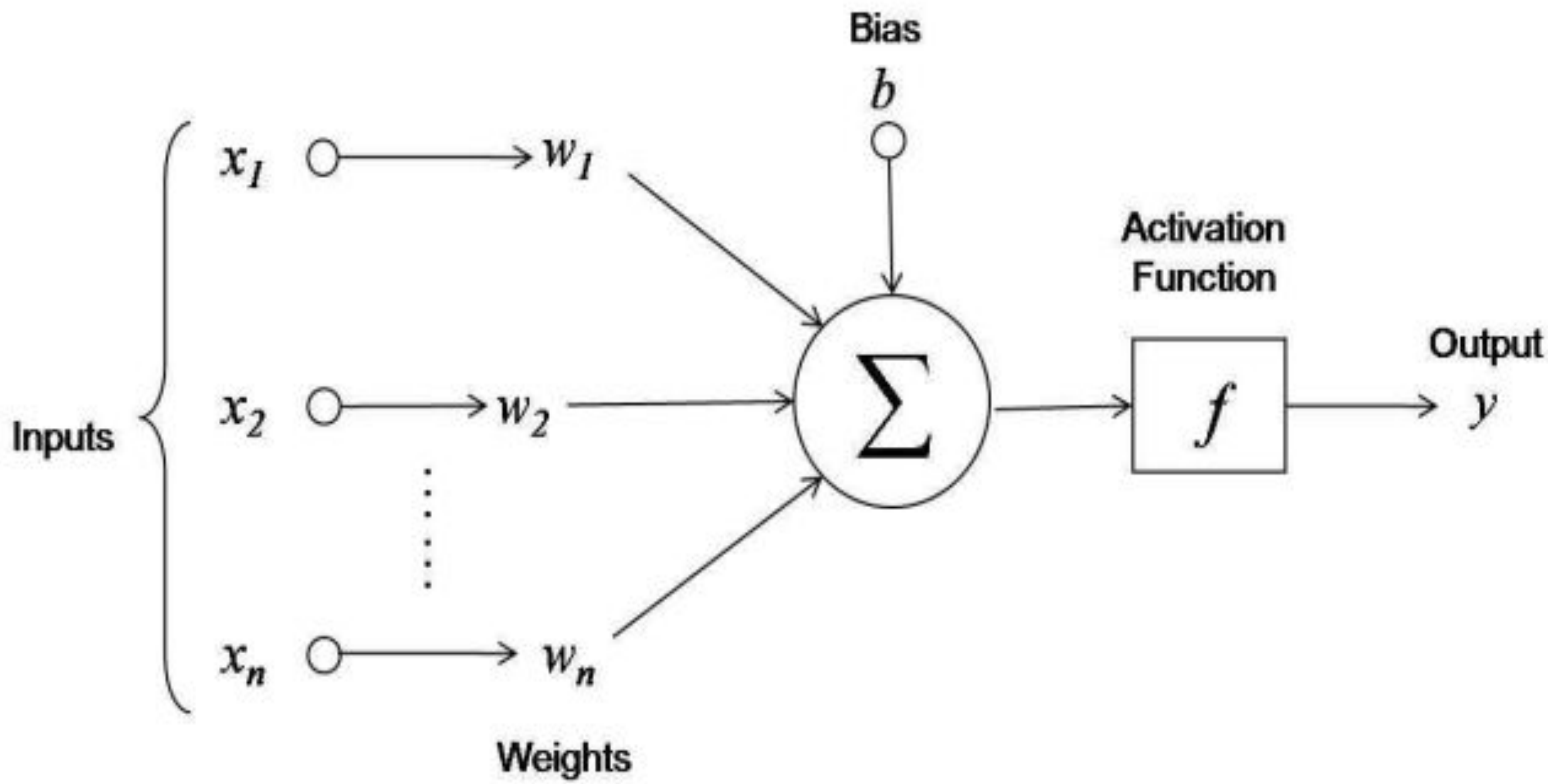
Neural Computation

- Understanding biological neural systems falls in a spectrum between mimicry and algorithmic solutions
- To mimic the brain, shift question from “how do brains compute?” to “how do brains build and repair themselves as dissipative attractor-based structures?”

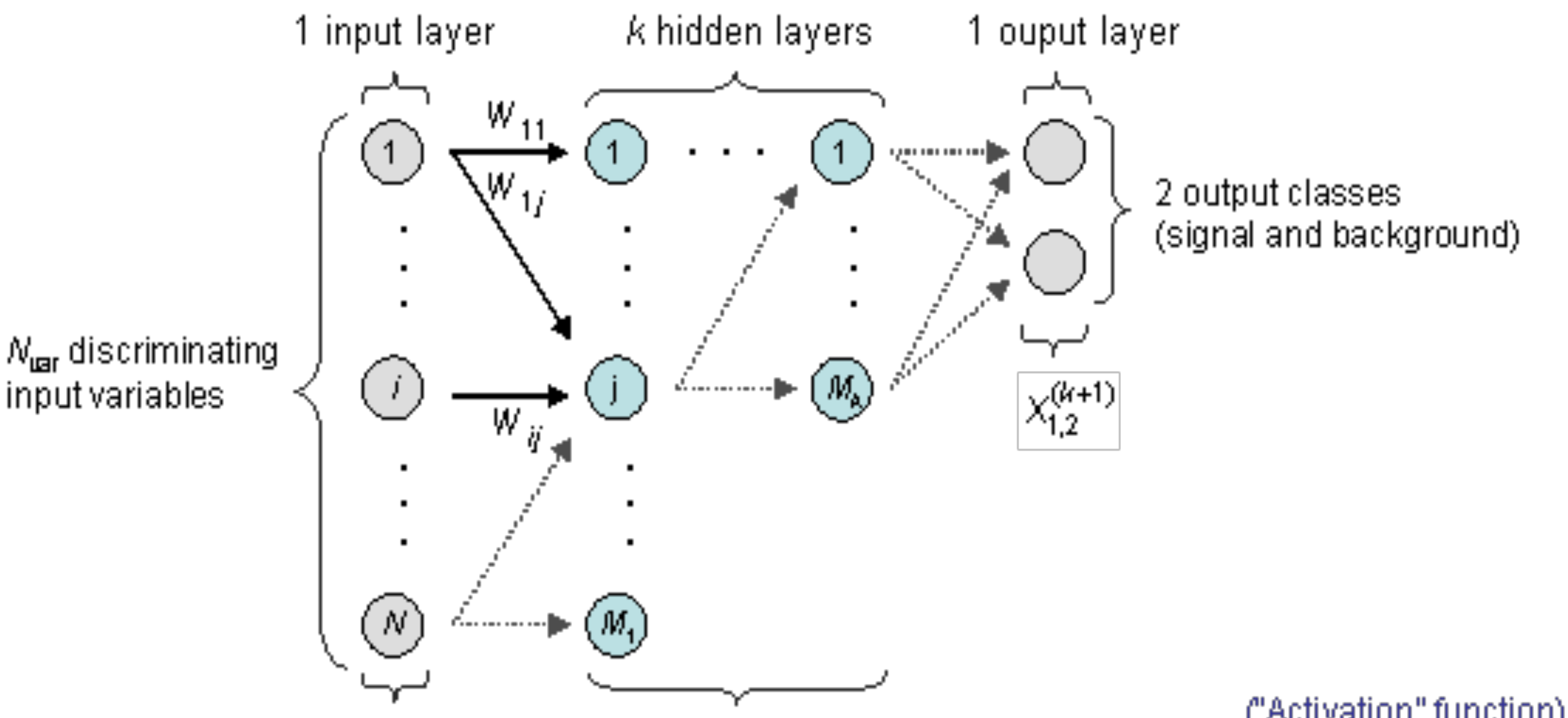
What is an Artificial Neural Network?

- ANNs are generally densely connected groups of individual cells, called neurons, that produce desired outputs for trained inputs (trained rather than programmed)
- Inspired by biological neural networks
- Used in machine learning and cognitive science to estimate functions

Neuron Diagram



ANN Diagram



$$x_j^{(k)} = A\left(w_{0j}^{(k)} + \sum_{i=1}^{M_{k-1}} w_{ij}^{(k)} \cdot x_i^{(k-1)}\right) \quad \text{with:} \quad A(x) = \left(1+e^{-x}\right)^{-1}$$

Inspiration for Physical Implementation

- IBM's Simulation with 1 Billion Neurons and 10 Trillion Synapses required 147,456 CPUs, 144 TB of memory running, and requires 2.9 MWs of power at 1/83 real-time.
  - Would require 7 GW of power to simulate human cortex based on these estimates, using von Neumann Architecture
  - Average human neocortex contains 150,000 billion connections
- Core Adaptive Power Problem: Energy wasted during memory-processor communication.

Adaptive Power Problem

- Constant dissipation of free energy allows living systems to adapt at all scales
- Each adaptation must reduce to memory-processor communication as state variables are modified
  - Energy consumed in moving this information grows linearly with number of state variables that must be continuously modified

History of Development

- Alan Turing's 'Intelligent Machinery' (1948)
  - Describes machine consisting of artificial neurons connected in any pattern with modifier devices
  - Modifier devices could pass or destroy a signal
- Hebb proposed synaptic plasticity (1949)
  - Describes the ability of synapses to strength or weaken over time in response to increases or decreases in activity
- ADALINE or "Adaptive Linear Neuron" (1960)
  - Physical device using electrochemical plating of carbon rods to emulate synaptic elements called memistors.
  - Memistor – Resistor with memory able to perform logic operations and store information.

History of Development (Continued)

- Memristor Theory (1971) – regulates flow of electrical current in circuit and remembers charge that previously flowed; retain memory without power (non-volatile)
- Very-large-scale integration (VLSI) (1980)–integrated circuit created by combining thousands of transistors into single chip
  - Helped Neural Networks (Hopfield), Neuromorphic Engineering (Mead), and Physics of Computation (Feynman)

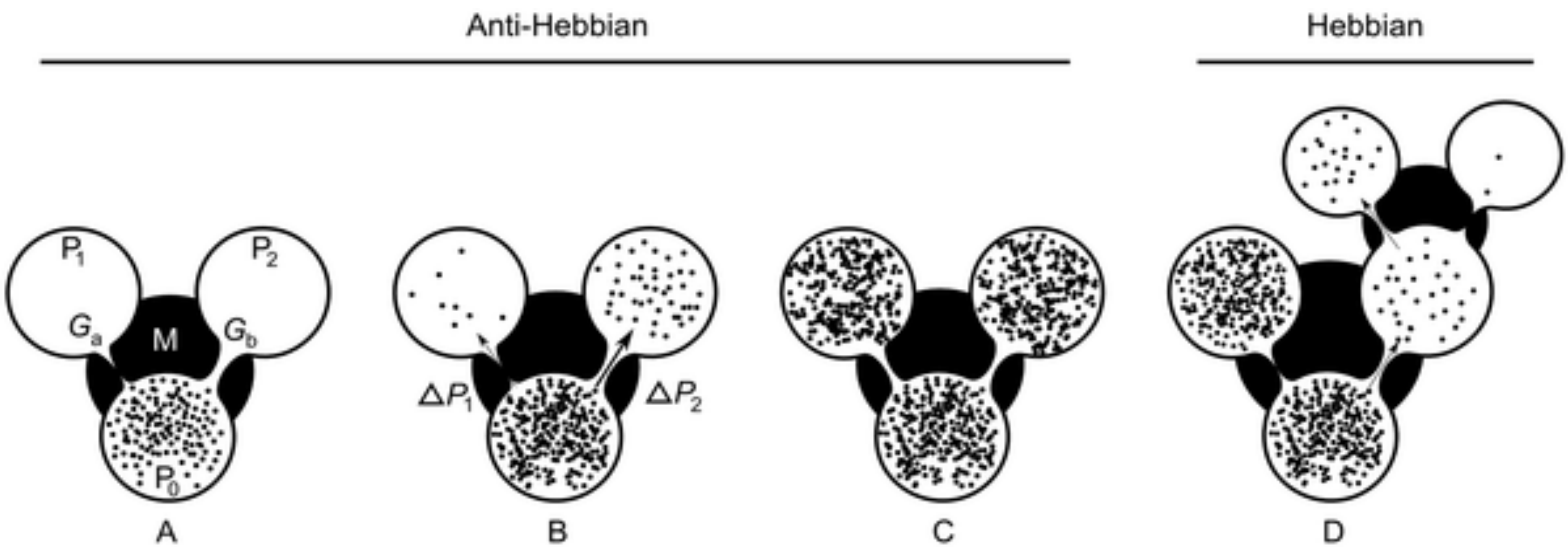
AHaH Computation

- Exploit controversial “Fourth Law of Thermodynamics” to retrain weights from Swenson (1989):
  - “A system will select the path or assembly of paths out of available paths that minimizes the potential or maximizes the entropy at the fastest rate given the constraints.”
- This thermodynamic process can be understood:

1) Particles spread out along available pathways and erode differentials that favor one path or the other

2) Pathways that lead to dissipation are stabilized

Figure 1. AHaH process.



Nugent MA, Molter TW (2014) AHaH Computing–From Metastable Switches to Attractors to Machine Learning. PLoS ONE 9(2): e85175. doi:10.1371/journal.pone.0085175
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0085175

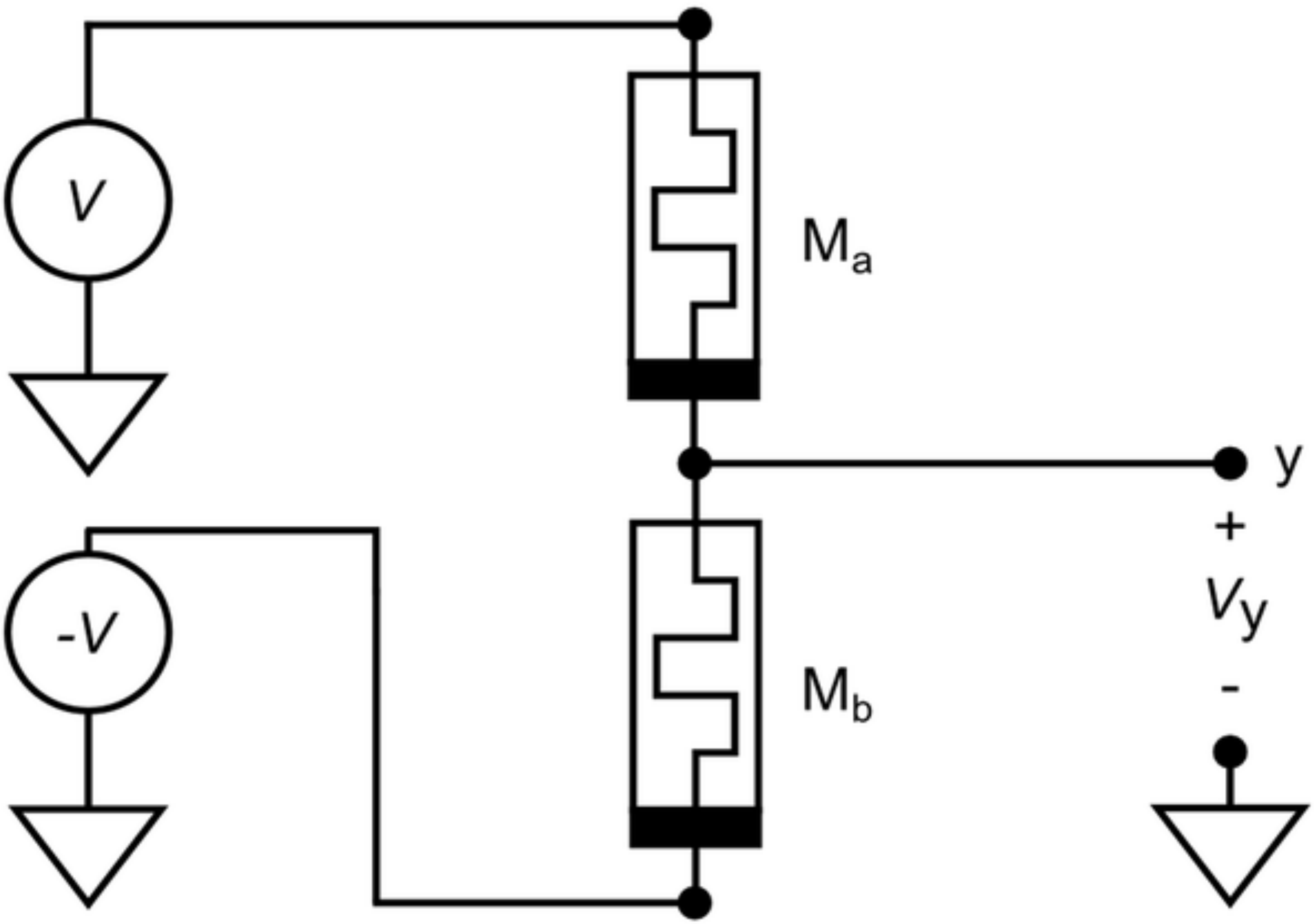
AHaH Plasticity

- Define synaptic weight:

  $w = G_a - G_b$, where G_a and G_b are conductance variables

- Anti-Hebbian (erase path)
  - Modification to synaptic weight that reduces probability that the synaptic state will remain the same upon subsequent measurement
- Hebbian (select path)
  - Modification to synaptic weight that increases probability that synaptic state will remain the same upon subsequent measurement

Figure 4. A differential pair of memristors forms a synapse.



Nugent MA, Molter TW (2014) AHaH Computing–From Metastable Switches to Attractors to Machine Learning. PLoS ONE 9(2): e85175.
doi:10.1371/journal.pone.0085175
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0085175

Linear Neuron Model

- $y = b + \Sigma^{N}_{i=0} x_i w_i$
  - y is the output of the selected neuron, b is the bias to the function, x_i is the output of neuron i, and w_i is the weight between the selected neuron and neuron x_i
- Weights and biases change based on the AHaH rule

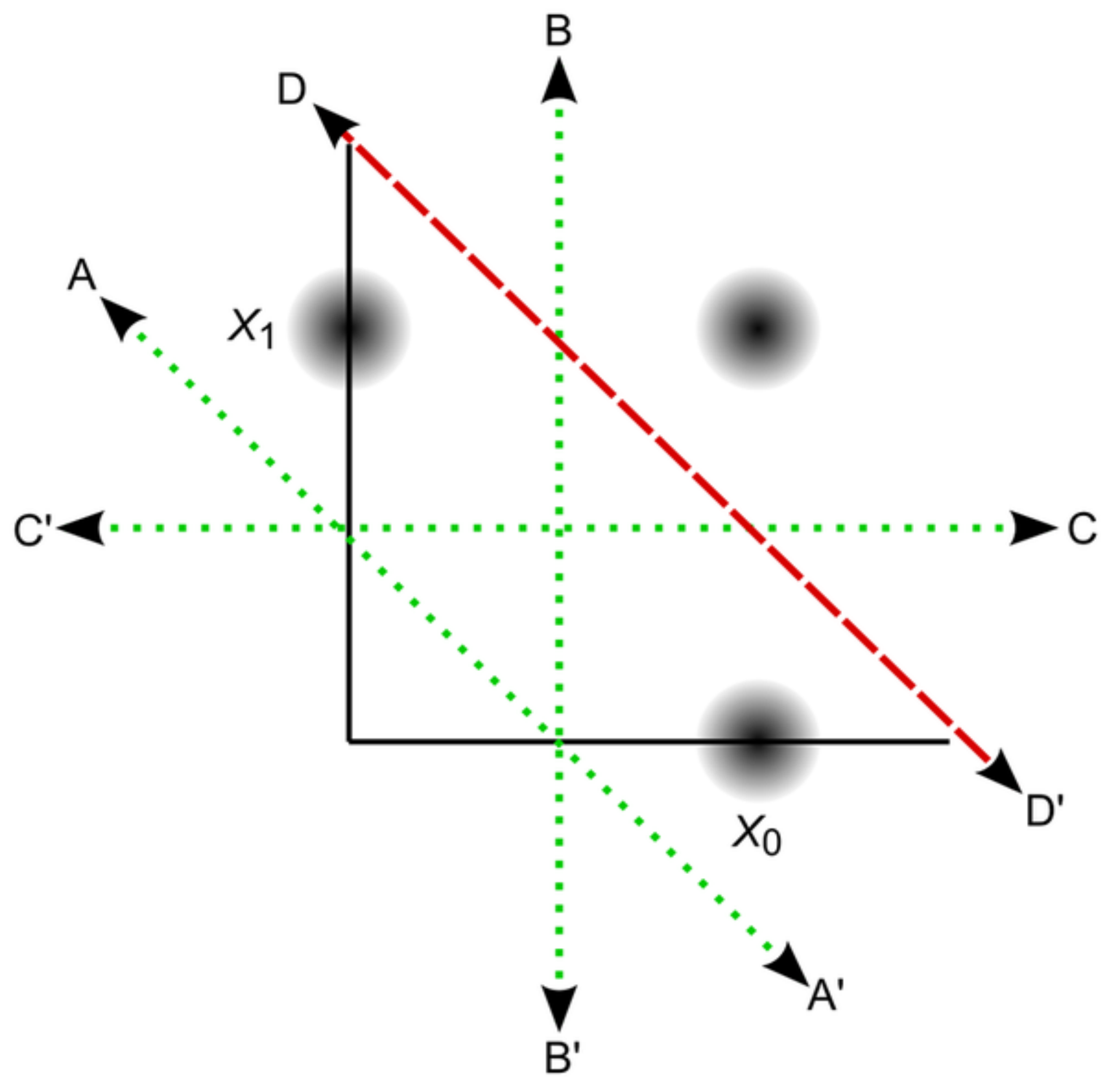
$$\Delta w_i = x_i \left(\alpha y - \beta y^3 \right)$$

  - α and β are constants that control the relative contribution of Anti-Hebbian and Hebbian plasticity

AHaH for Machine Learning

- AHaH node: uses AHaH rule to bisect input space using a hyperplane to make a binary decision
- Since there are many ways the AHaH node can bisect input, the hyperplane chosen to bisect the input space is "the one that maximizes the separation of the two classes" of inputs.

Figure 2. Attractor states of a two-input AHaH node.



Nugent MA, Molter TW (2014) AHaH Computing–From Metastable Switches to Attractors to Machine Learning. PLoS ONE 9(2): e85175.
doi:10.1371/journal.pone.0085175
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0085175

AHaH Power Consumption

- Major motivation is to eliminate von Neumann bottleneck regarding memory-processor communication
- Energy output is defined:

$$E = E_r + E_w = \Delta T V^2 \left(\frac{G_a G_b}{G_a + G_b} + (G_a | G_b | 0) \right),$$

- In practice, the energy consumption per synapse per update is around 12 pJ, although scaling trends project energy consumption to be lowered to 2 pJ in the future.

Future Goals

- Provide a physical adaptive learning hardware resource (AHaH circuit) that provides memory similar to how RAM provides memory to current computing systems.
- Use AHaH node as a building block for higher-order adaptive algorithms not yet conceived.

Conclusion

- Artificial Neural Networks are great function approximations, but there is a major bottleneck from memory-processor communication using von Neumann architecture
- Reaching back to the late 1940s, this type of computation has been theorized and studied
- AHaH computation exploits physical phenomenon to implement an artificial neural network, lessening the memory-processor communication bottleneck for power while allowing massive parallelism.

Sources

- AHaH Computing-From Metastable Switches to Attractors to Machine Learning
  - http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085175
- "Neural Network Learning"
  - http://web.eecs.utk.edu/~mclennan/Classes/420-527-S14/handouts/Part-2A-1pp.pdf
- "Neural network models for pattern recognition and associative memory"
  - http://www.sciencedirect.com/science/article/pii/089360808990035X
- "Perceptrons, Adalines, and Backpropagation"
  - http://www-isl.stanford.edu/~widrow/papers/bc1995perceptronsadalines.pdf