Neuro-inspired Computational Engines: Beyond von Neumann/Turing Architecture and Moore’s Law Limits

2015 NICE Workshop

Murat Okandan

Sandia National Laboratories

February 24, 2015

Sandia National Laboratories is a multi-program 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.

NiCE – Neuro-inspired Computational Engines

GOALS:

1) Analyze, predict and control natural and engineered systems, in ways not possible or feasible with conventional computing.

2) Understand the physical mechanisms underlying thought, consciousness, …

OUTCOME:

"predict the future in the most efficient manner possible"

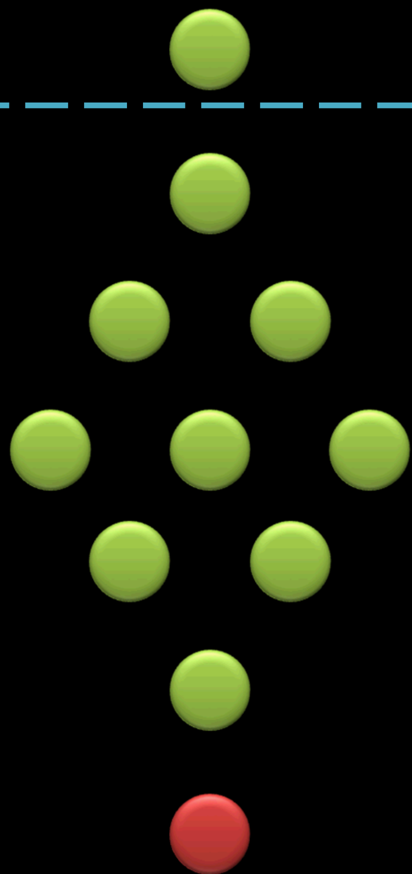
M. Okandan

Why? - How? - What?

- Why should we consider neuro-inspired systems?
  - Why do we need neuro-inspired systems?

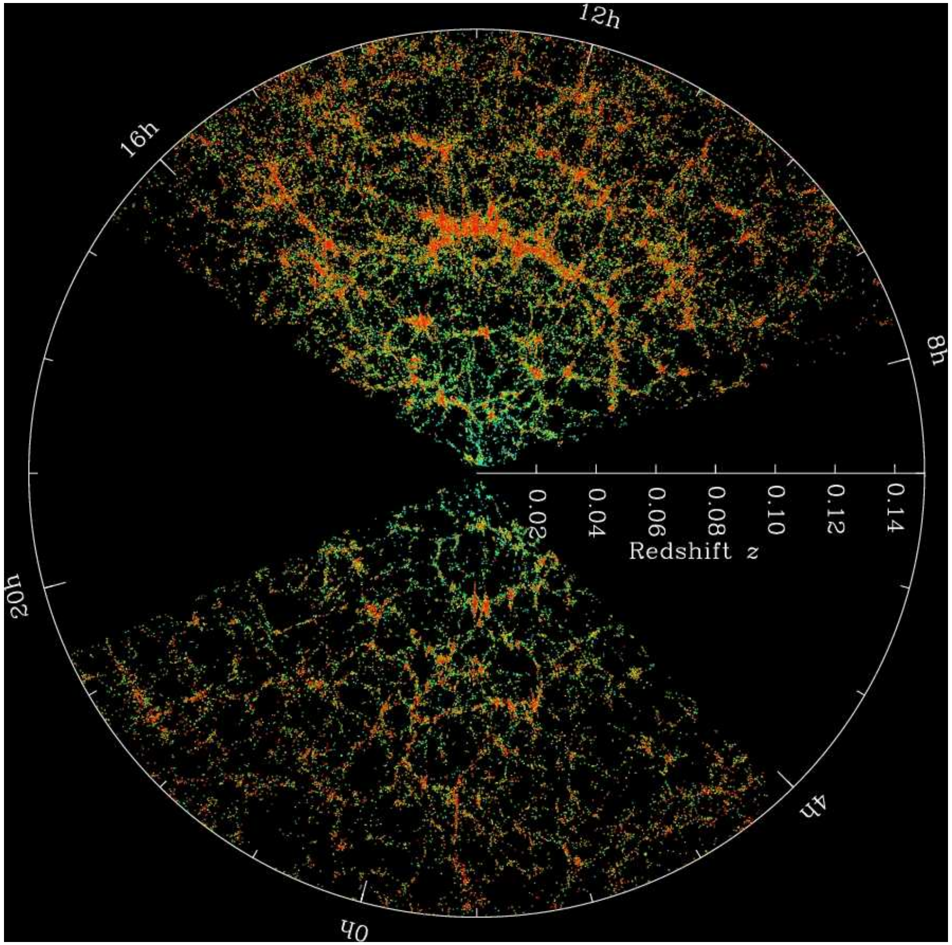
- How are we going to build these neuro-inspired systems?

- What are we going to do with them?



- 1 – physics
- 1 – existence proof
- 2 – key functions
- 3 – ways to build it
- 2 – things to do with it
- 1 – evaluation project
- 1 – ?

Biggest Picture View



Sloan Digital Sky Survey

ENERGY

1 – physics

1 – existence proof

2 – key functions

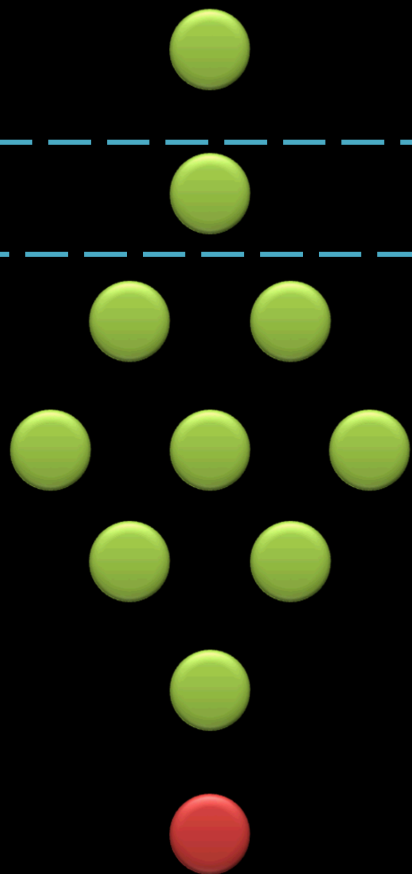
3 – ways to build it

2 – things to do with it

1 – evaluation project

1 – ?

What lets us control our bodies to capture, store and use energy?



1 – physics

1 – existence proof

2 – key functions

3 – ways to build it

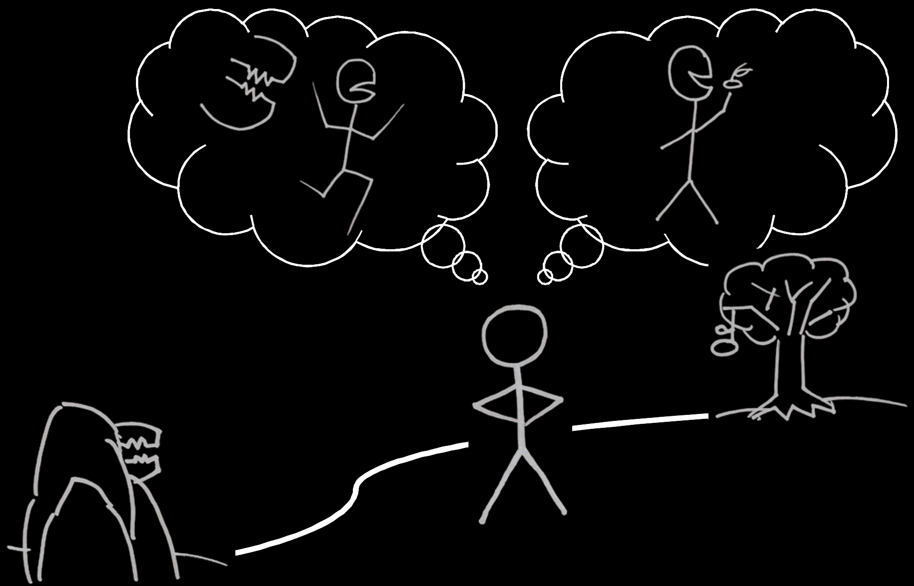
2 – things to do with it

1 – evaluation project

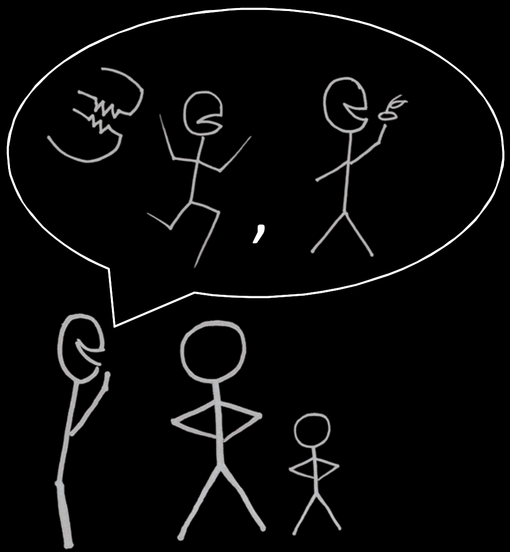
1 – ?

What are the two key functions of a neural system?

1) Learn and Predict the Future



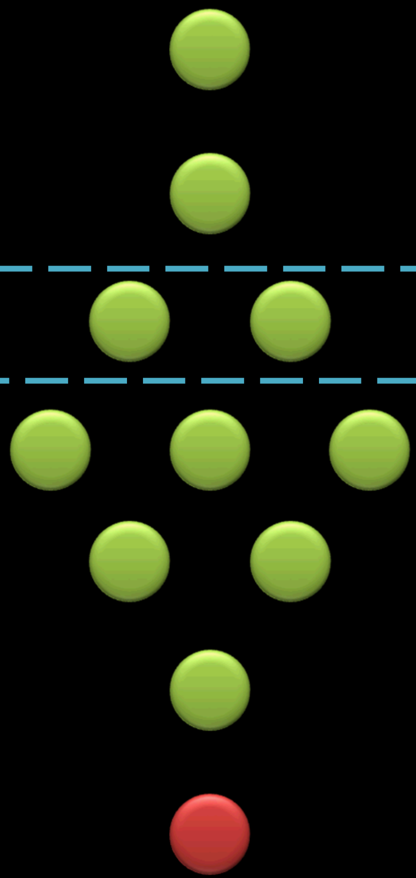
2) Communicate



how does a brain, a neural structure accomplish these functions?

sparse, hierarchical, spatio-temporally encoded
representation, processing, storage and recall

not by symbolic manipulation, mathematical calculations, or algorithmically
solving ODE/PDE problems
(avoiding a rock or throwing a spear)



1 – physics

1 – existence proof

2 – key functions

3 – ways to build it

2 – things to do with it

1 – evaluation project

1 – ?

How could you build a neuro-inspired/neuromorphic system?

1) software - use clever algorithms on the fastest machines, simulate brain activity
   (deep learning, Google, Facebook, HBP, …)

2) tweaked (digital+analog) hardware – add new devices to accelerate specific functions
   (specialized GPU/FPGA/ASIC, novel devices on CMOS, neuromorphic wafers)

3) novel architecture that natively implements
   sparse, hierarchical, spatio-temporal encoded representation
   (liquid state machines, reservoir computing)

energy efficiency and speed argument for (3)
sparse, spatio-temporal encoding:

neural system example – 10,000 inputs – 10,000 outputs per neuron,
how do you compute 10,000 in,
generate a spike and
route that to 10,000 outputs?

only way to do this in a conventional electronic system is packet switching – assume 16-bit address, how much energy does it take to encode, route and decode to deliver 1-bit payload (a spike, 0 to 1 transition) to one output –

~ **10pJ** and time delay of **100x** (1ms of activity takes 100ms to simulate/calculate)

* SpiNNaker system - real-time run capability.

It would be possible to do the same in a substrate that is specifically designed to implement these functions, embedded at the lowest device physics level (micro-opto-electronic devices), doing local computation, driving network reconfiguration with local rules –

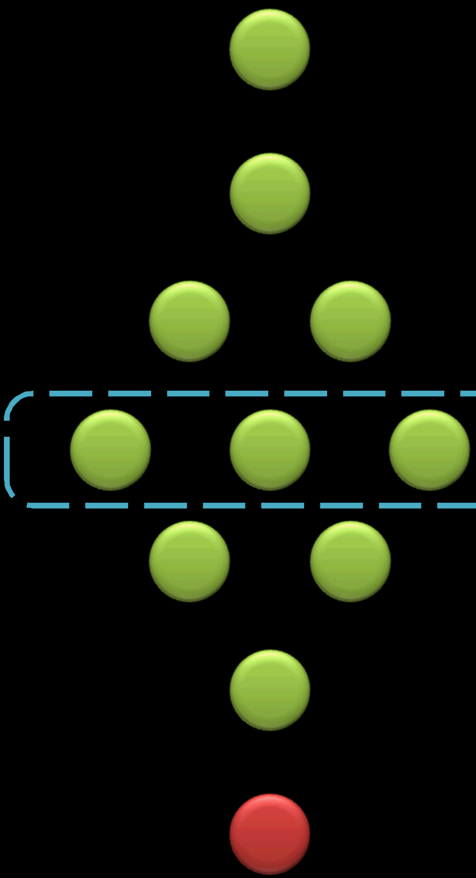
~ **10fJ**, and < **1/10,000** real-time. (1ms spike time vs. <<100ns spike time)

* Heidelberg platform (HBP) - 10pJ/spike and <1/1000 real-time.

Two key enabling concepts:

- **Massive interconnectivity** (>1000 in and out per unit element) and **self-reconfigurability (plasticity)** - necessary to enable neuro-computation, ideally at the lowest level device in the architecture, with low power.

- **Sparse, spatio-temporally coded, hierarchical representation** of information, instantiated by correlated activation of unit elements in a big enough network - necessary for achieving the high level of functionality desired.



1 – physics

1 – existence proof

2 – key functions

3 – ways to build it

2 – things to do with it

1 – evaluation project

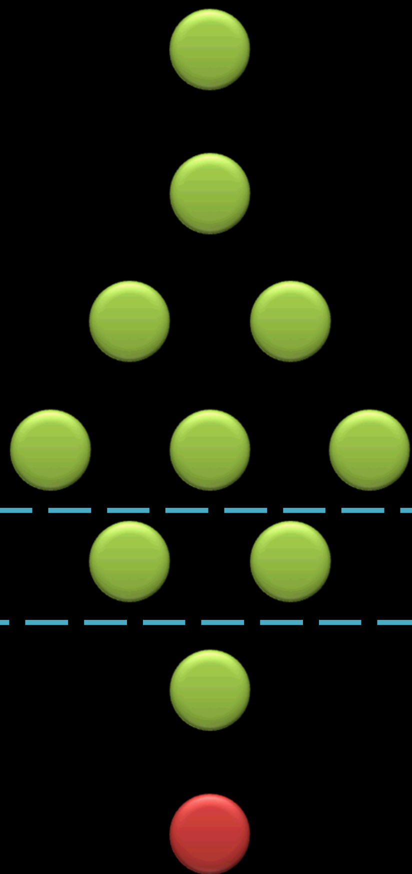
1 – ?

What would you do with such a system?

2 things:

1) use features of neural computation to analyze, predict and control systems in ways currently not possible (power, speed, size, functionality, …)

2) study and understand further details of how the brain/neural systems work



1 – physics

1 – existence proof

2 – key functions

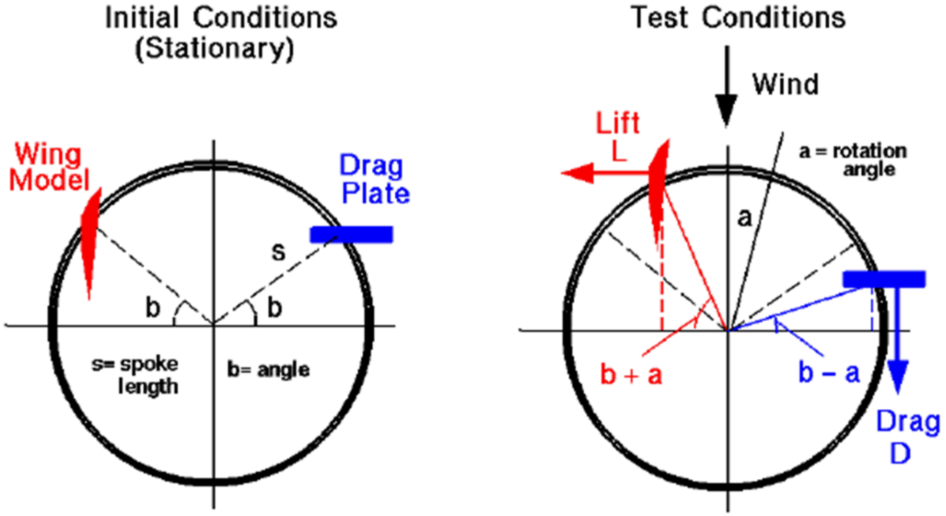
3 – ways to build it

2 – things to do with it

1 – evaluation project

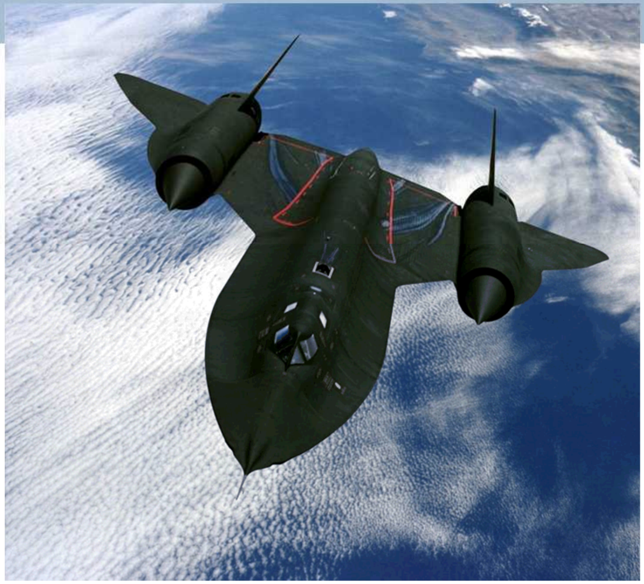
1 – ?

Wright Brothers' first wind tunnel



$$L \sin(b + a) = D \cos(b - a)$$

Wright Brothers' first wind tunnel – to :



Neuro-Inspired Systems Evaluation Project

1) How can we access the hardware and what are the associated costs and timelines? Are there software emulation tools available?

2) Which algorithms have you used to benchmark this system? Especially of interest are algorithms aimed towards image processing.

3) Can we get the benchmark results and associated metrics?

4) How do these results compare to conventional computing (CPU/GPU, etc.)?

NiCE – Neuro-inspired Computational Engines

GOALS:

1) Analyze, predict and control natural and engineered systems, in ways not possible or feasible with conventional computing.

2) Understand the physical mechanisms underlying thought, consciousness, …

OUTCOME:

"<u>predict the future</u> in the most efficient manner possible"

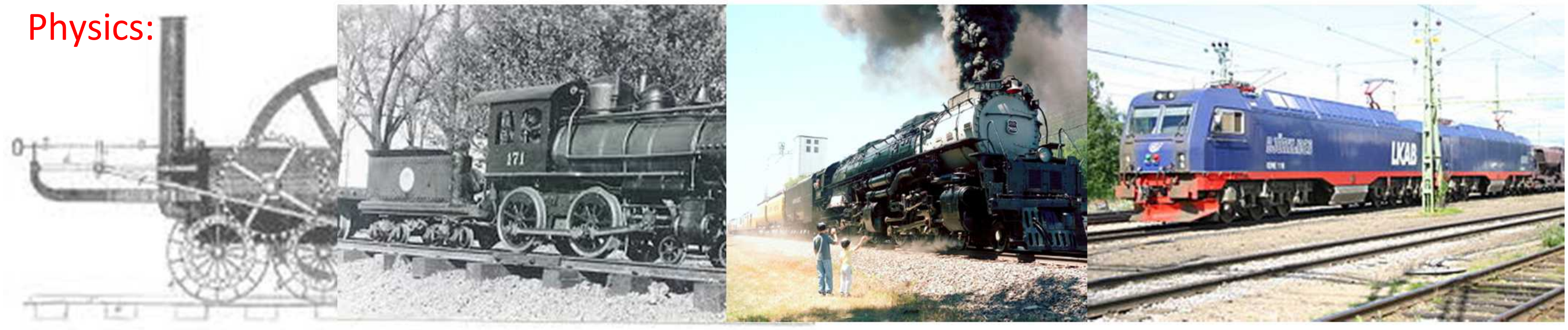
Applications

- <u>Sensor systems:</u>
  High pixel counts, high data rates, limited bandwidth for comm, limited on-board power.
- <u>Unmanned/remote systems:</u>
  High consequence features, time critical response, limited bandwidth for comm, limited on-board power.
- <u>Big Data/Cyber (graph-like):</u>
  Massive data rates, low probability, high consequence features.
- <u>Complex, adaptive systems (graph-like):</u>
  Massive simulations, critical dynamic patterns determine and indicate future behavior.
- <u>Neural interfaces/neuroscience:</u>
  Yet to be uncovered primitives for information encoding and processing, efficient coupling into central/peripheral nervous system, platform for testing hypotheses ("Wright Brothers' wind tunnel").

Couple Points on Moore's Law – and limits

Three reasons the limits are being (have been) reached:

Physics:



Economics:

"Moore's law (scaling) will not end because of physics, but because of economics."
Gordon Moore

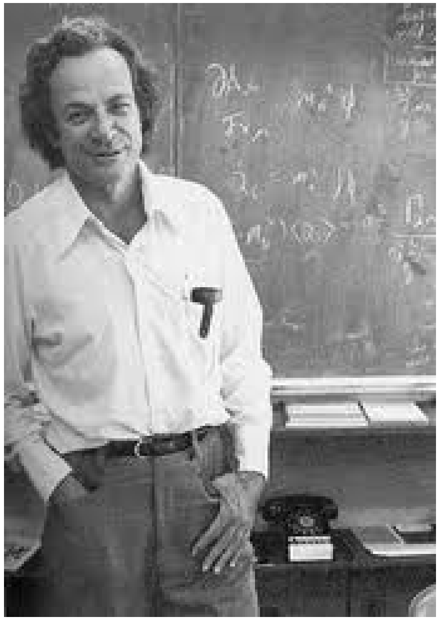
Functionality:

Symbolic computation (a Turing machine) is not capable of providing the level of sophistication and energy efficiency needed for next generation analysis, prediction and control systems.

Why? - How? - What?

- Why should we consider neuro-inspired systems?
  - Why do we need neuro-inspired systems?
- How are we going to build these neuro-inspired systems?
- What are we going to do with them?

What we really are going to do with it…



Feynman's Corollary on new technology

"Like everything else new in our civilization, it will be used for entertainment."

Feynman's second nanotechnology talk, 1983