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

UNM Mind Research Network Presentation

Murat Okandan

Sandia National Laboratories

January 23, 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.

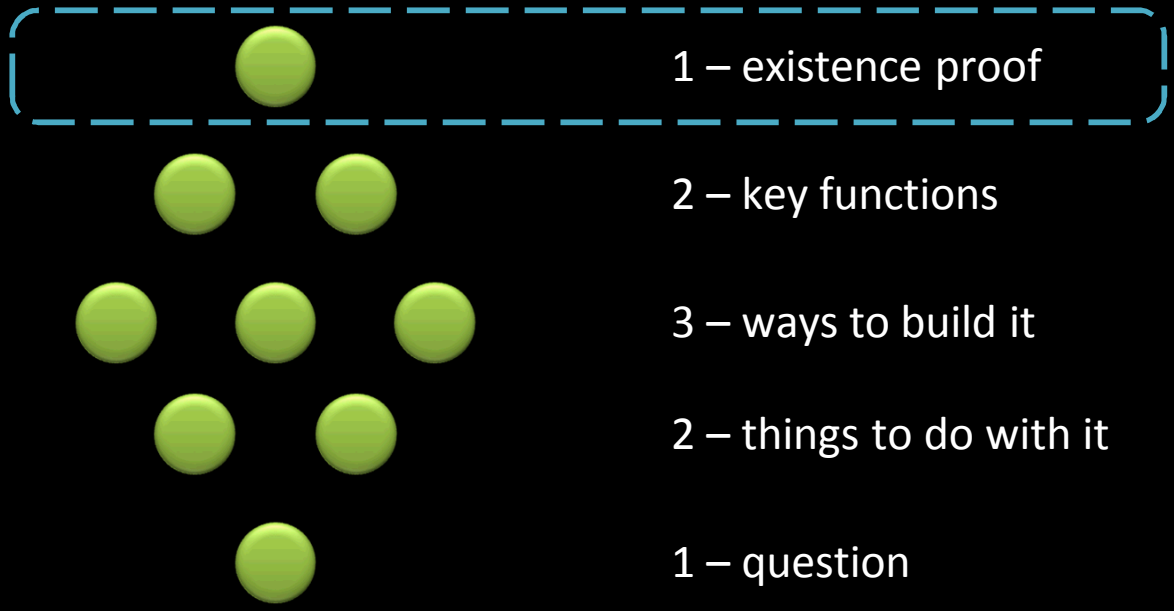
NiCE – Neuro-inspired Computational Engines

GOAL:

Analyze, predict and control systems in ways we can not do with conventional computing.

"predict the future in the most efficient manner possible"

Neuro-inspired Computational Elements Workshop:
http://nice.sandia.gov

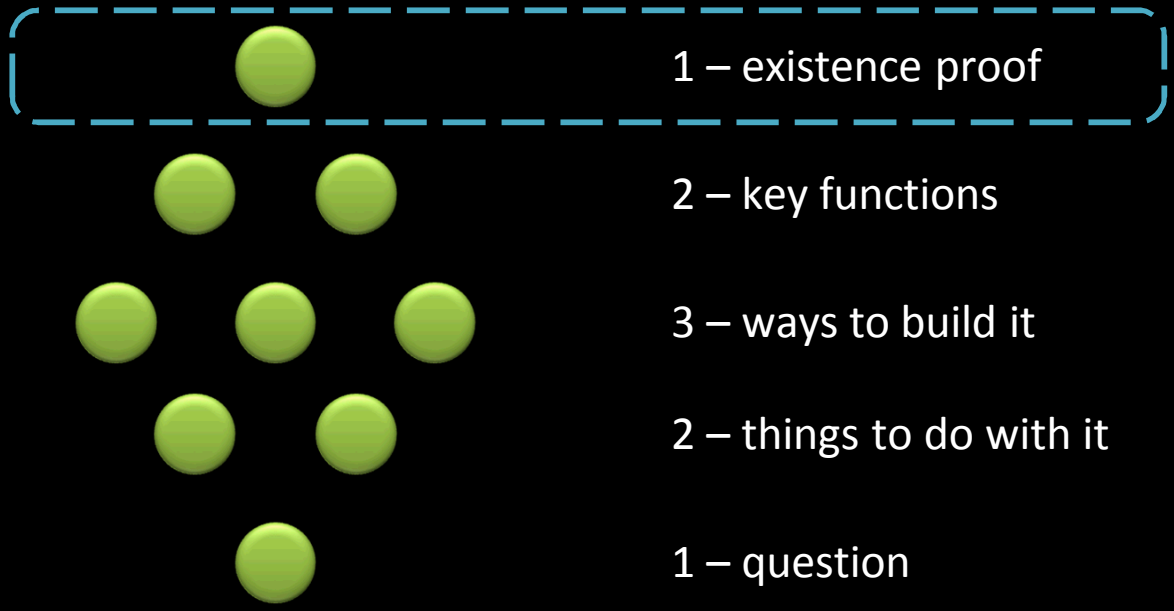

1 – existence proof
2 – key functions
3 – ways to build it
2 – things to do with it
1 – question

what gives you your biggest survival advantage?

Sensori-motor genesis of cortex

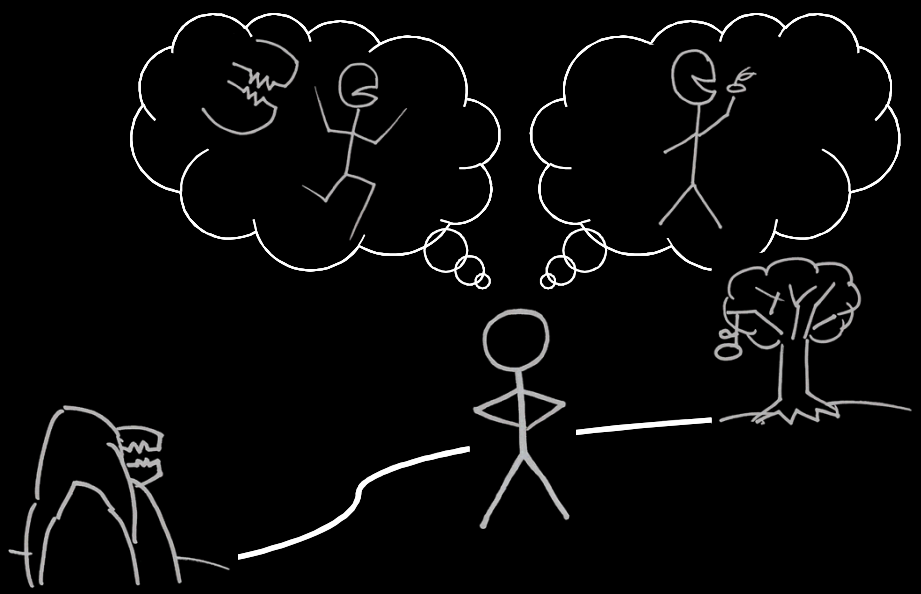
http://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains.html

shortcut: google “wolpert brains”

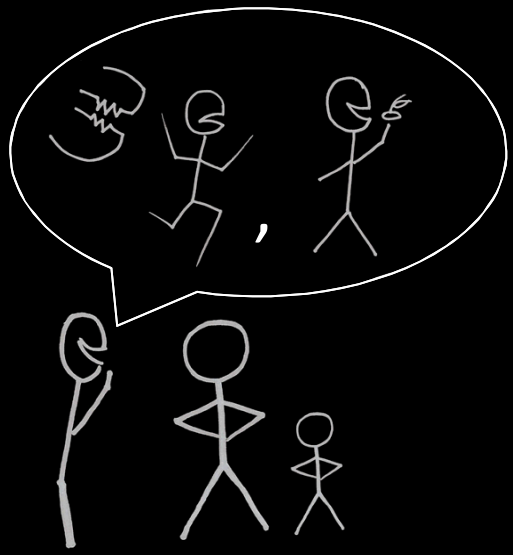

1 – existence proof
2 – key functions
3 – ways to build it
2 – things to do with it
1 – question

what are the two most important functions, capabilities that your brain gives you?

1) Learn and Predict the Future



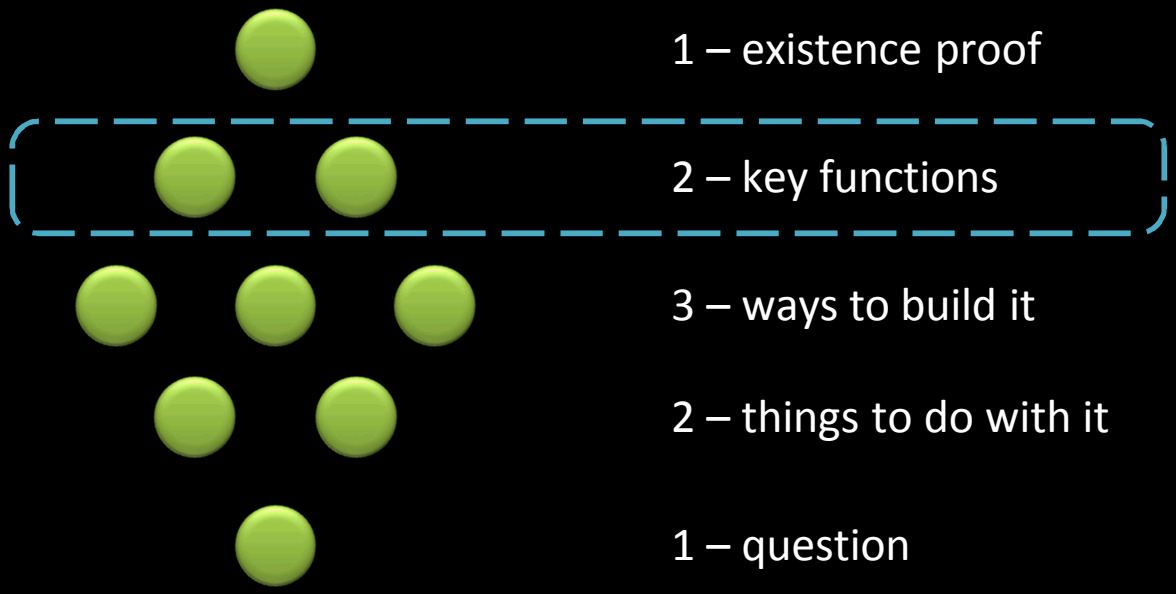
2) Communicate



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

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

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


1 – existence proof
2 – key functions
3 – ways to build it
2 – things to do with it
1 – question

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)

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) has 10pJ/spike and <1/1000 real-time

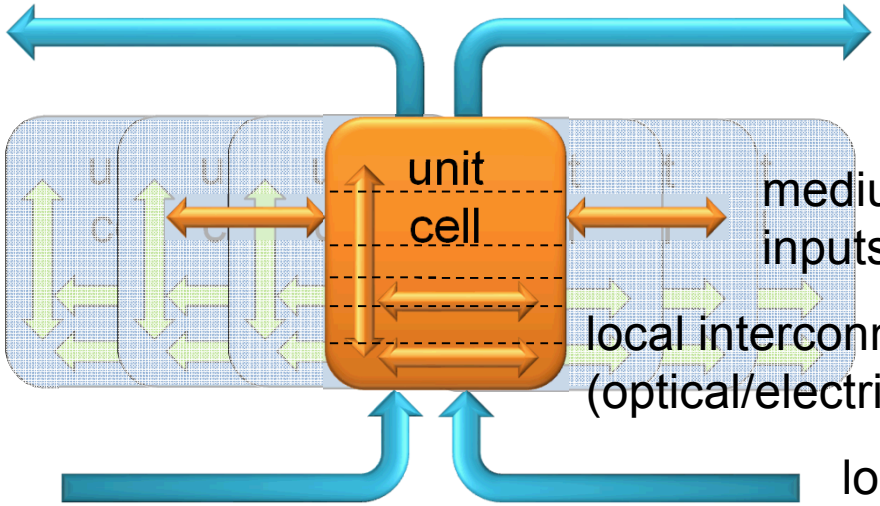
Two key enabling concepts:

- Massive interconnectivity (>1000 in and out per unit element) and self-reconfigurability (plasticity) - needed 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 (prediction of future states).

Neuro-inspired Computational Engines

A new substrate for representing and processing information



long range outputs (optical)

medium range
inputs/outputs (optical)

local interconnects
(optical/electrical)

long range inputs (optical)

"cortical column" - hierarchical, temporal memory

3D hybrid integration – opto-electronics, TSV, novel devices, ...

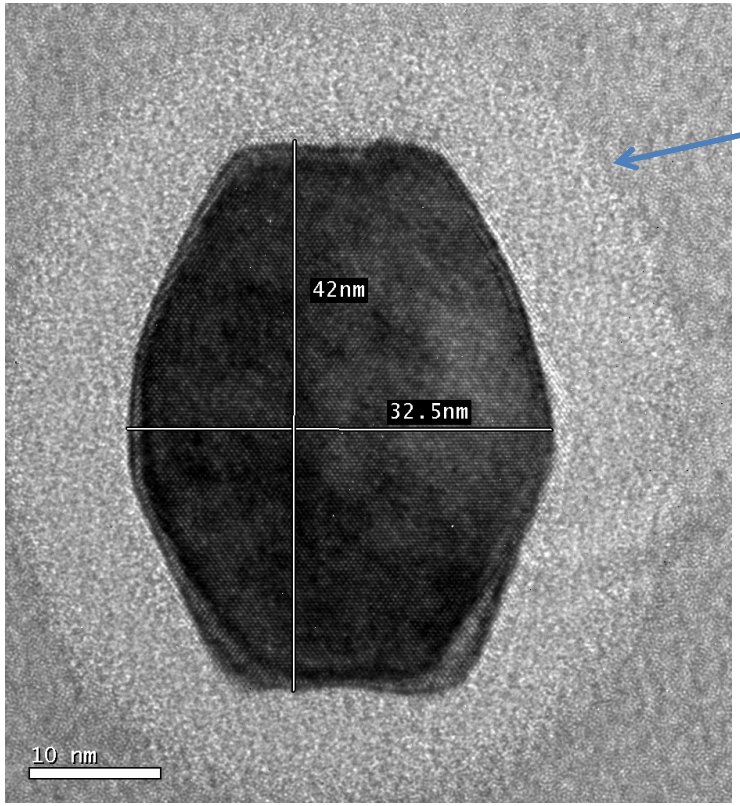
key characteristics:
- Plasticity/adaptability at native (device) level functionality
- Massive interconnect/fanout at system level

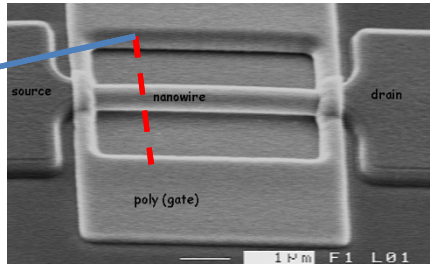
imagine a 10 x 10 x 10 “brain-cube” [10^3]

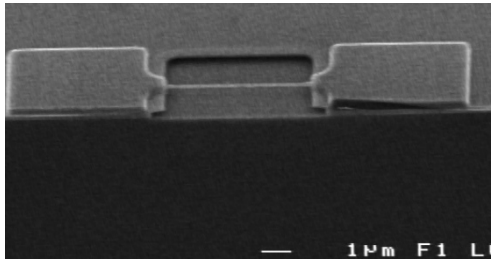
now imagine 100 x 100 x 1 “brain-cube array” [10^7]

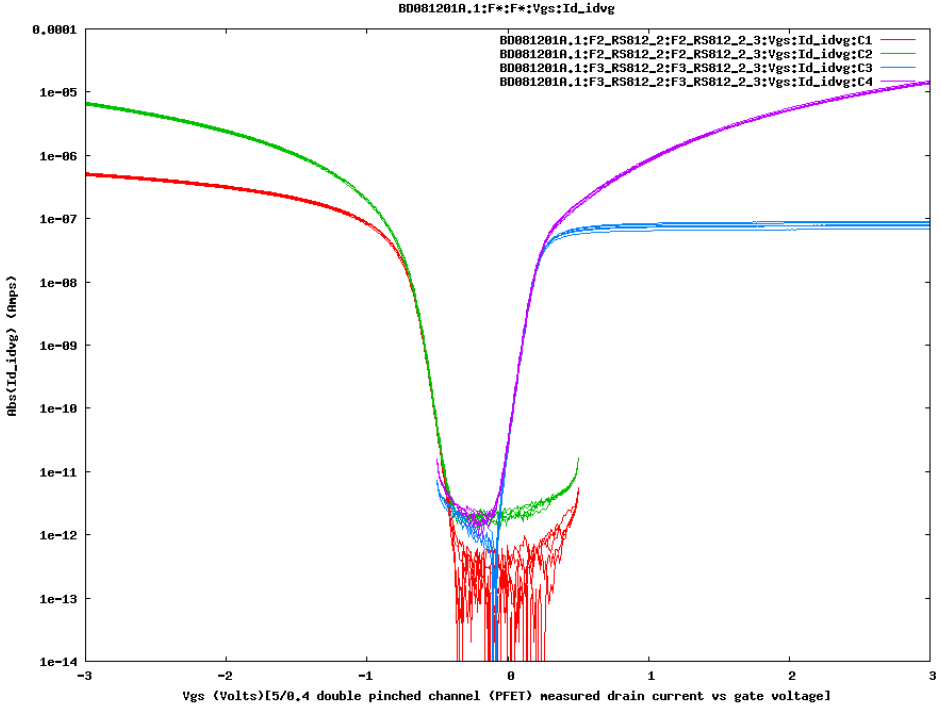
1000 x 1000 x 100 “brain-cube array” [10^{11}]

CMOS Front-end manufactured Si Nanowires

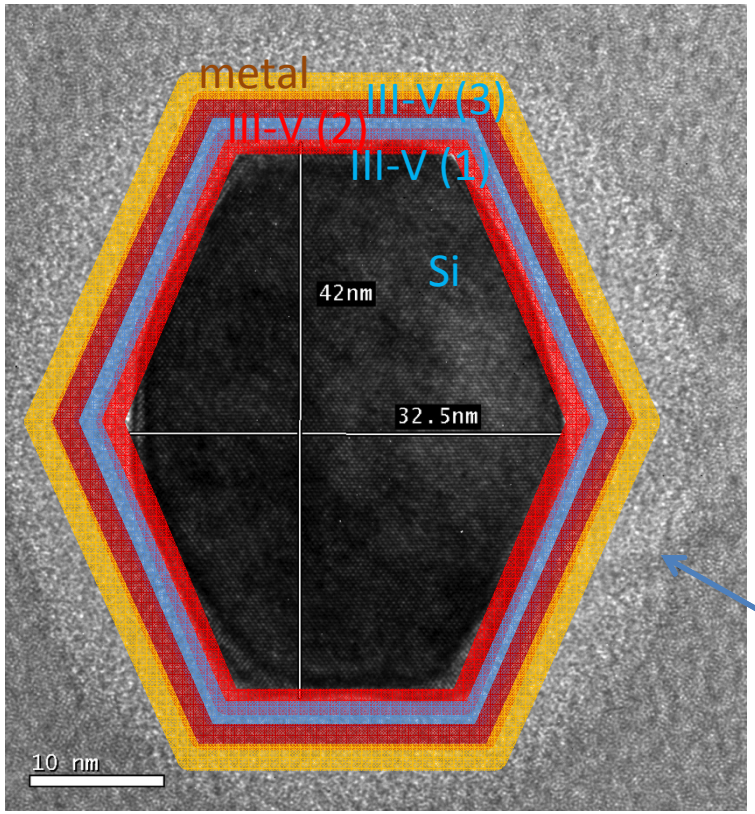






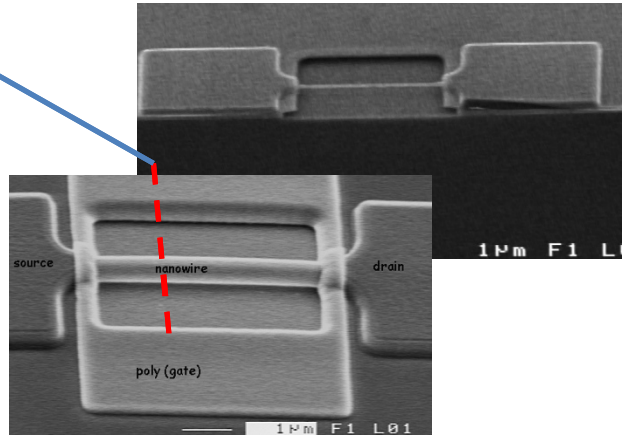


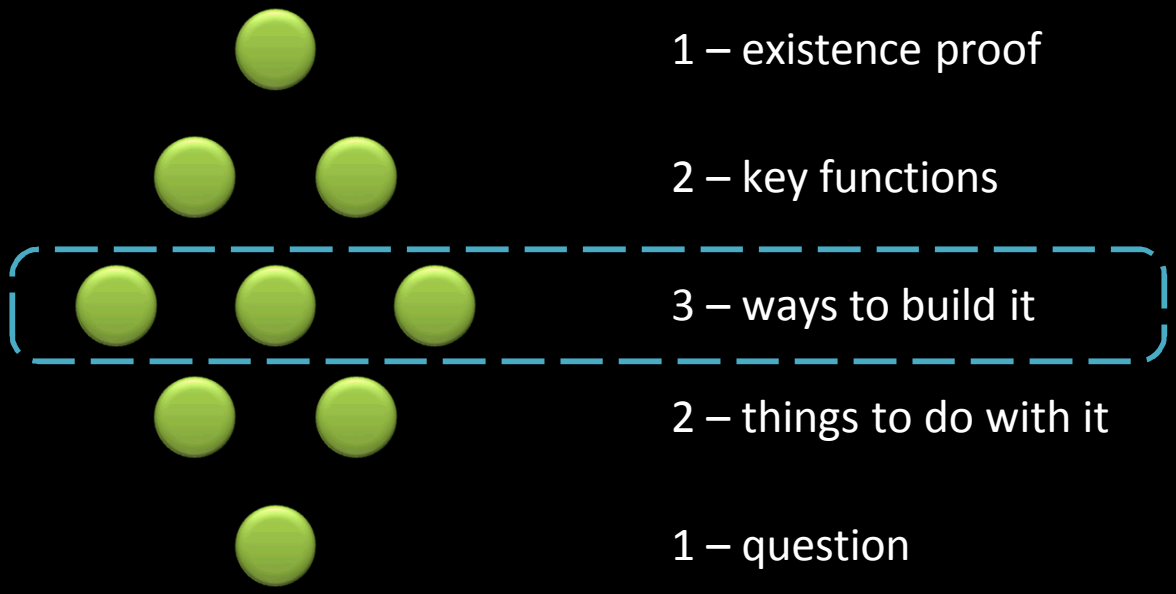
CMOS Embedded Light Source – Si Photonics



Key characteristics:

- Plasticity/adaptability at native (device) level functionality
- Massive interconnect/fanout at system level




1 – existence proof
2 – key functions
3 – ways to build it
2 – things to do with it
1 – question

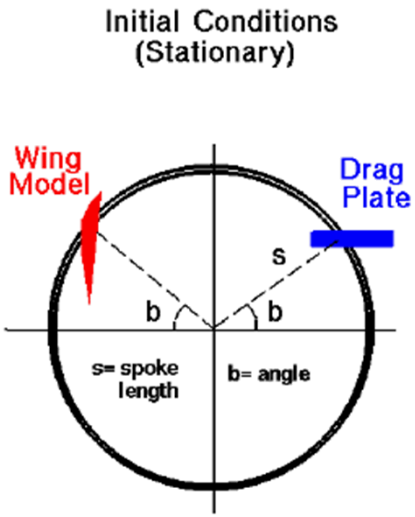
what would you do with such a system?

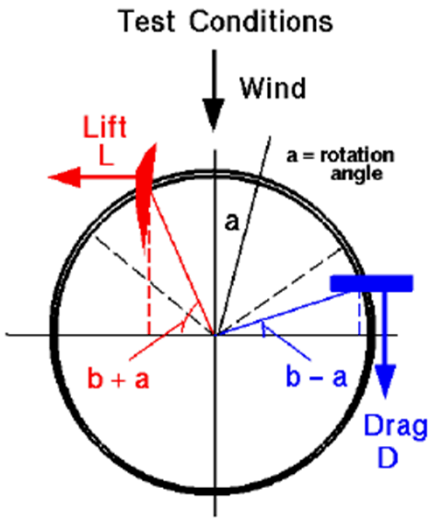
2 things:

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

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

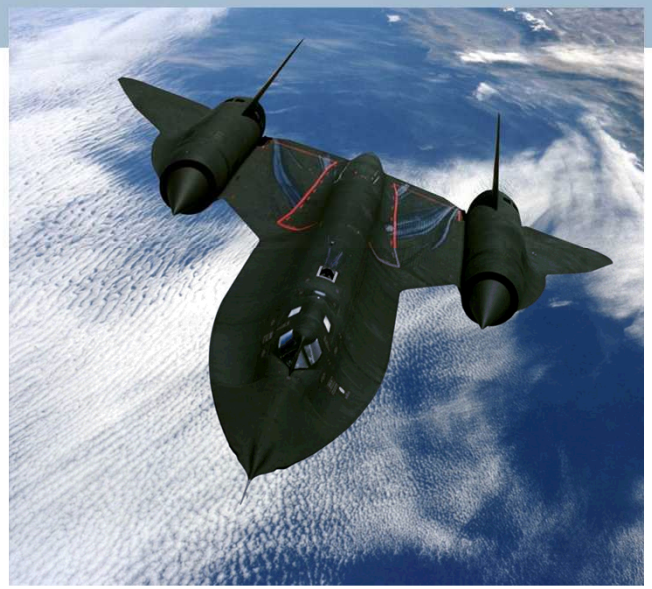
Wright Brothers' first wind tunnel

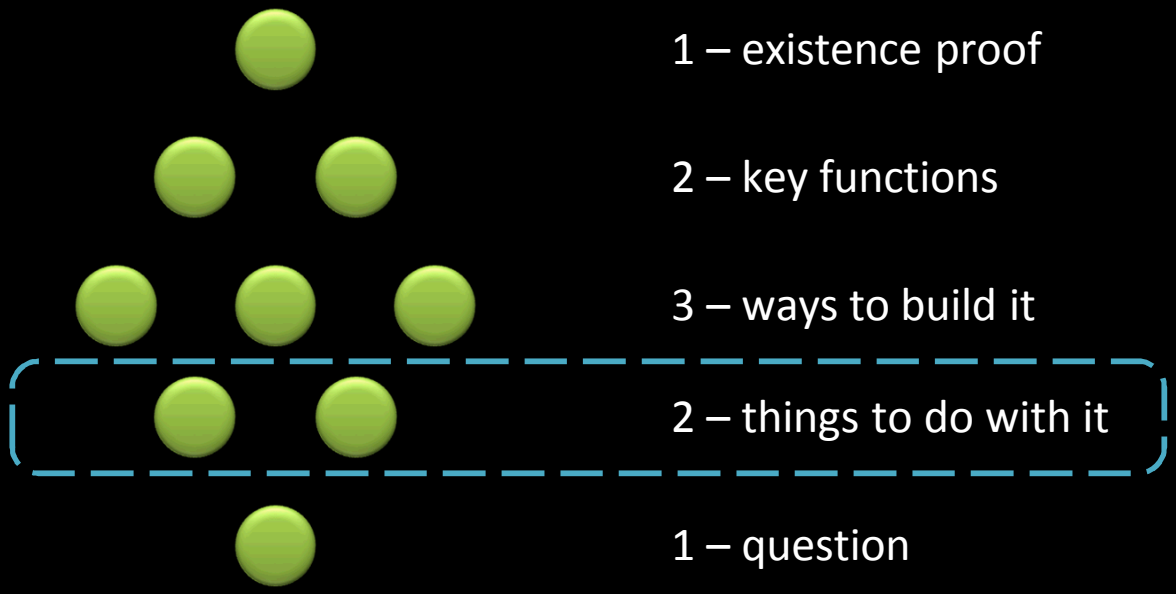




$$L \sin(b + a) = D \cos(b - a)$$

Wright Brothers' first wind tunnel – to :




1 – existence proof
2 – key functions
3 – ways to build it
2 – things to do with it
1 – question

What will happen next?

What will you do?

What will I do?

NiCE – Neuro-inspired Computational Engines

GOAL:

Analyze, predict and control systems in ways we can not do with conventional computing.

"predict the future in the most efficient manner possible"

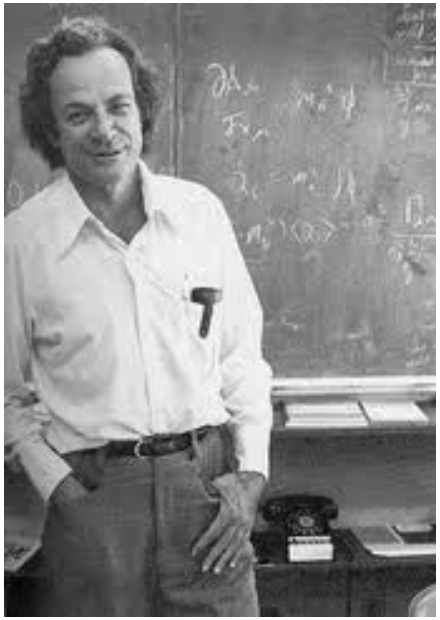
Applications

- Sensor systems:
  High pixel counts, high data rates, limited bandwidth for comm, limited on-board power.
- Unmanned/remote systems:
  High consequence features, time critical response, limited bandwidth for comm, limited on-board power.
- Big Data/Cyber (graph-like):
  Massive data rates, low probability, high consequence features.
- Complex, adaptive systems (graph-like):
  Massive simulations, critical dynamic patterns determine and indicate future behavior.
- Neural interfaces/neuroscience:
  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”).

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