Neuro-Inspired Non-Conventional Computational Approaches

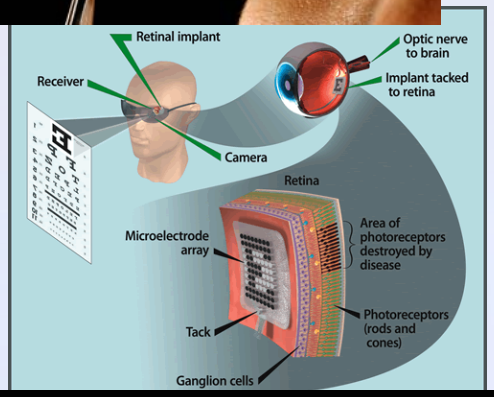
NeuINCC (pronounced “noink”)

SAND2011-5410P

Murat Okandan
Sandia National Labs

Drivers :
- End of physical device scaling
- Power limitations
- Applications with large, incomplete, noisy (“natural”) data sets

Inspiration



Current approaches

Probabilistic -Bayesian Computation
Novel Architectures

Lyric
Probability processing

Navia systems
Probabilistic computing

Numenta
Intelligent computing

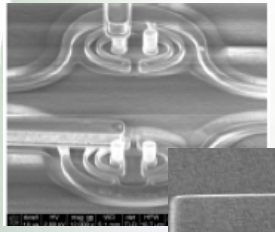
achronix SEMICONDUCTOR CORPORATION
Asynchronous high speed PLD

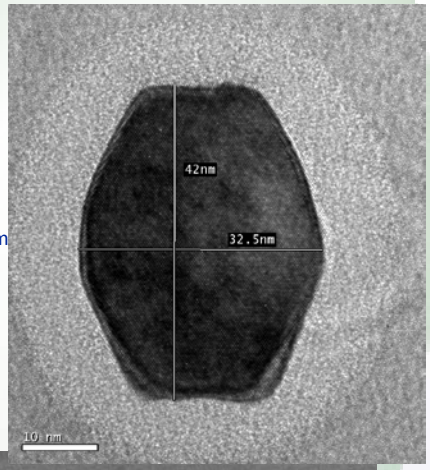
LUXTERA
CMOS integrated optical comm.

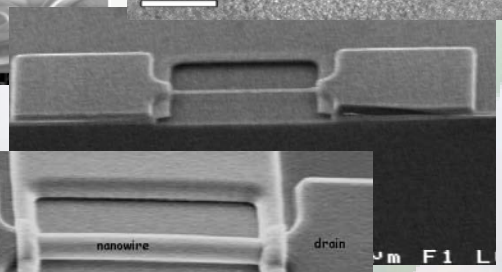
Future

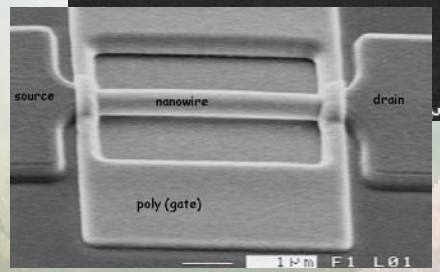
Emulating pattern recognition, abstraction and prediction model of neural systems – from device physics up to systems level

computation not with “1s-and-0s” but 1s-0s in and around the system to provide coupling to conventional systems









Sandia National Laboratories

- Why we are going to do it
  - What we are going to do
    - What are we going to do with it

Why?

Two technology (physics) reasons and one applications reason

- Device scaling reaching final physical limits (beyond 11nm?)

  might be able to address this through 3D integration (TSV, multi-layer CMOS, etc.)

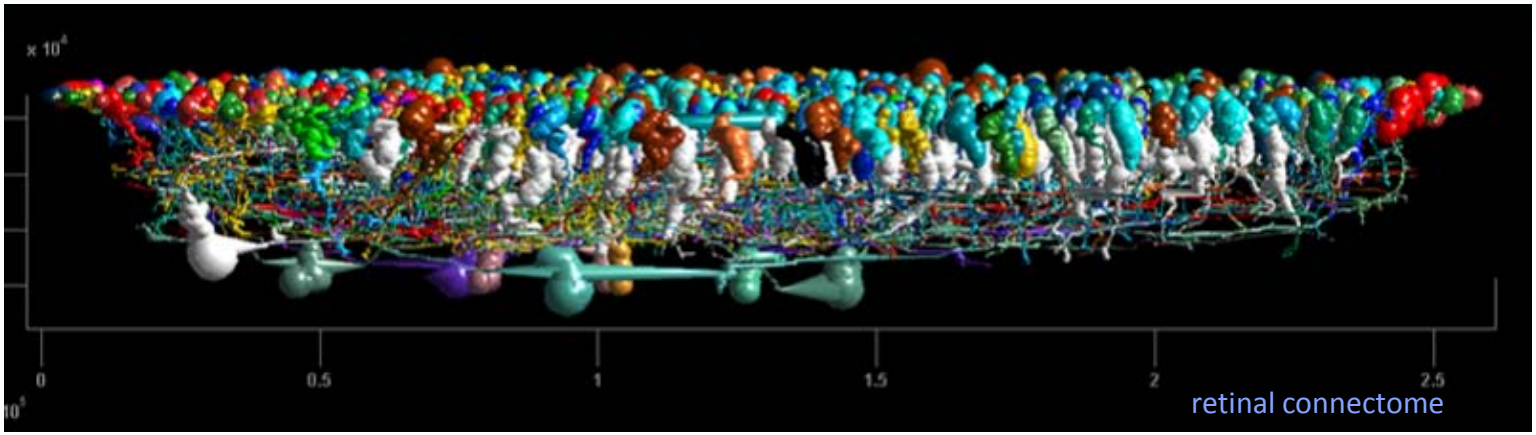
- Power density / power consumption per operation

  even at 100kT (or ~1 kT) per operation, exa-scale computers will require >MW power levels.

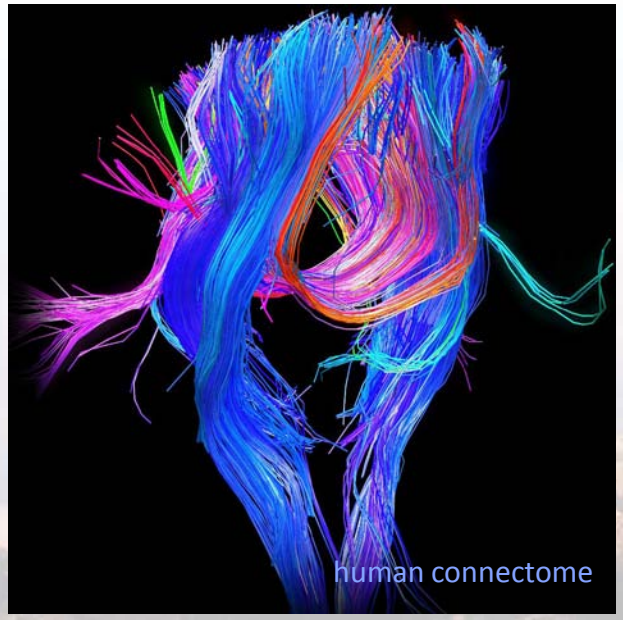
- Most challenging problems involve large, noisy, incomplete data sets

  conventional (von Neumann) architecture systems can only perform exact / algorithmic calculations and have trouble with pattern recognition – and have no native path for abstraction

Model System



Robert Marc group, University of Utah



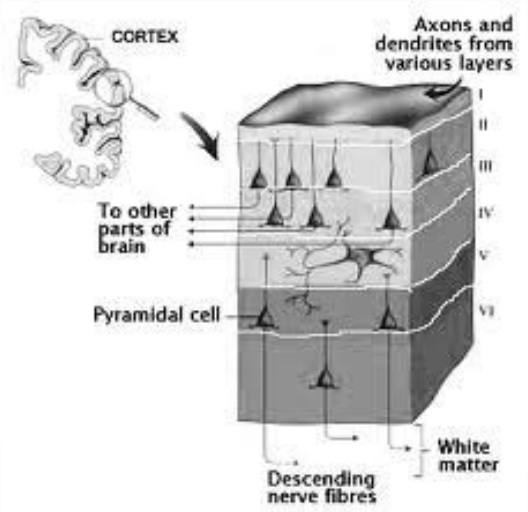
Human Connectome Project

Harvard-MGH/UCLA

WU/Minn

Sandia National Laboratories

Neo-Cortex – CPU/GPU Comparison



~75% of human brain volume

10 billion neurons (10^{10})

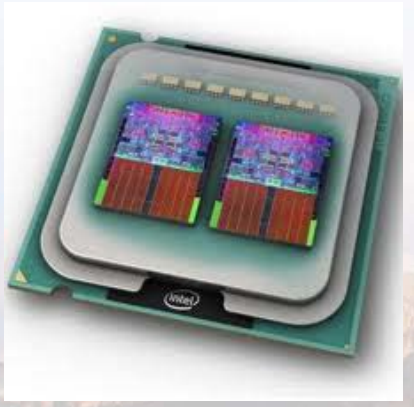
10-100 trillion synapses (10^{13}-10^{14})

2500 cm^2 x 0.4 cm thick

(large dinner napkin, 1/6 inch thick)

massively interconnected, (10^4 synapses per neuron)
dynamic, fault tolerant, low power (~ 25W)
~10ms per unit cell, ~100ms for perception

(photons in → object recognition)



3-7 cm^2 x 200um thick

(~10um active thickness, transistors + metal stack)

1-4 billion transistors

2-4 GHz

40-100W

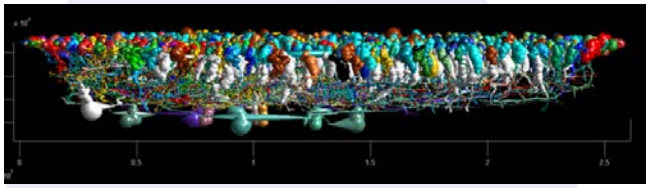
Neo-Cortex – Neural Computation

- not a von Neumann architecture system
- Pattern recognition – abstraction – prediction
  - dynamic
  - multi-dimensional
    - difference engine + accumulator with goals: *
      acquire sustenance, avoid predators, reproduce
- How is the data represented, stored and processed?
  - still an open question
  - multiple mechanisms and time scales

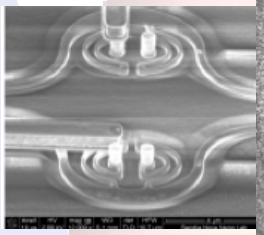
* not necessarily the optimum solution, it worked – and it is conserved

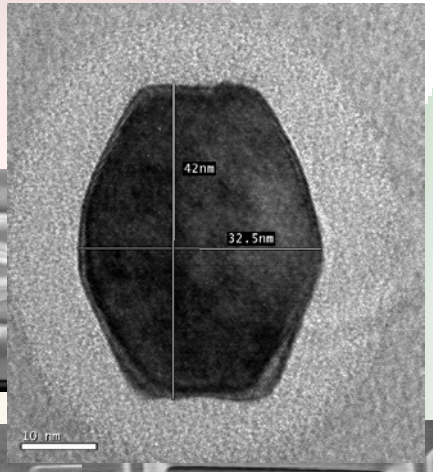
What we are going to do…

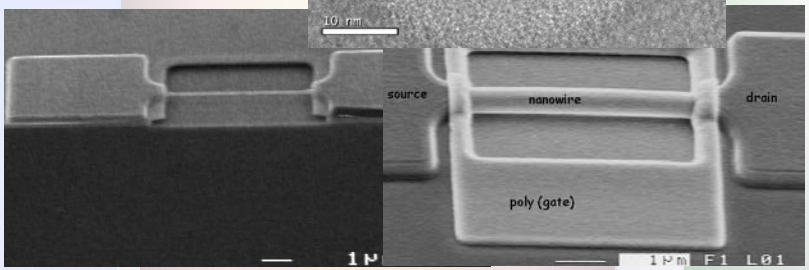
1) Model neural systems



2) Develop new devices and architectures







achronix SEMICONDUCTOR CORPORATION

Asynchronous high speed PLD

LUXTERA

CMOS integrated optical comm.

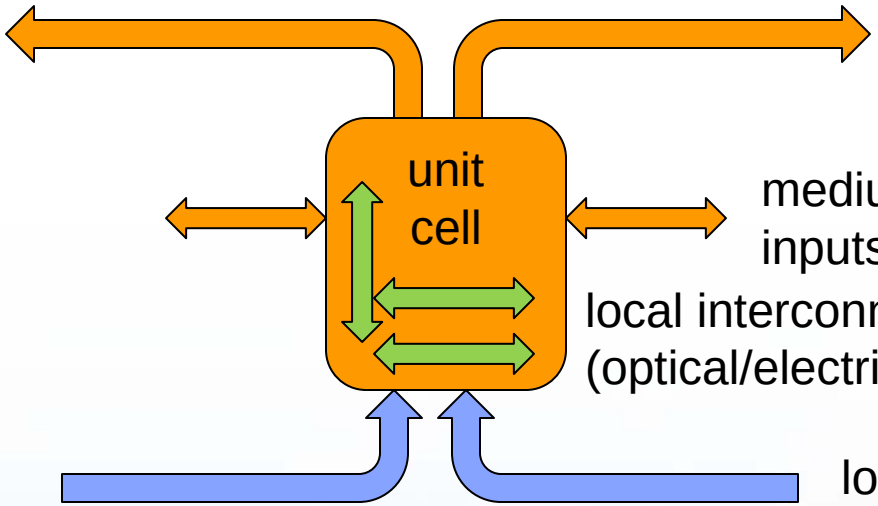
3) Use sub-systems for specific applications

Lyric

Probability processing

Sandia National Laboratories

2) Develop new devices and architectures



long range outputs (optical)

medium range
inputs/outputs (optical)

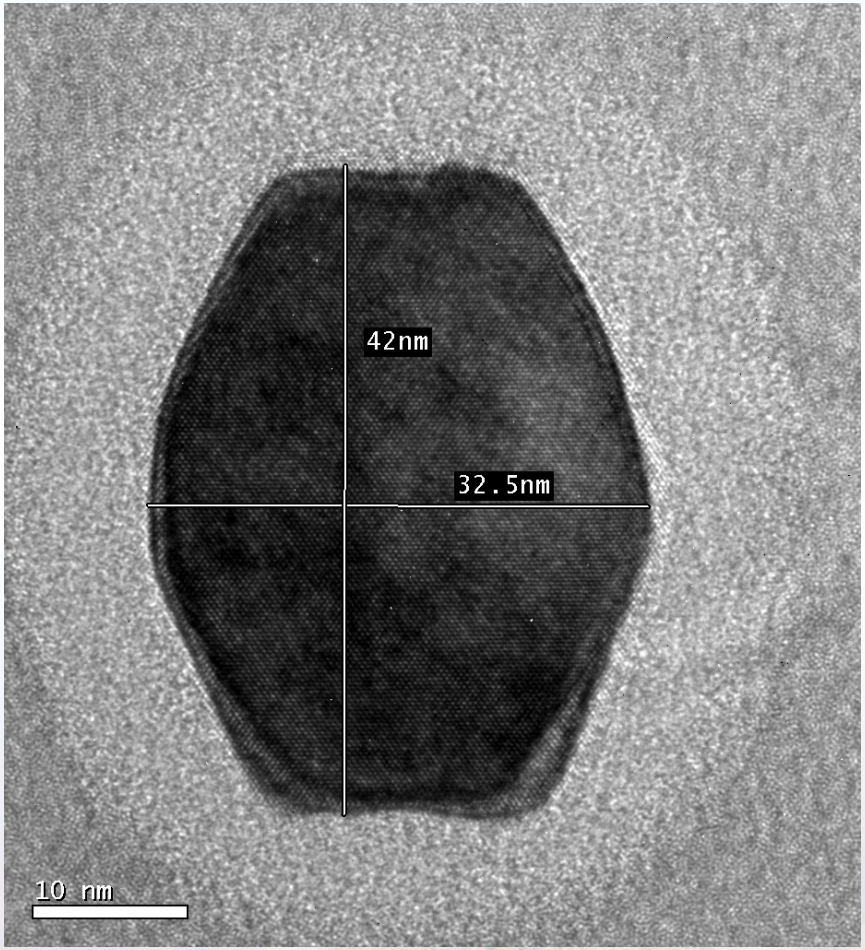
local interconnects
(optical/electrical)

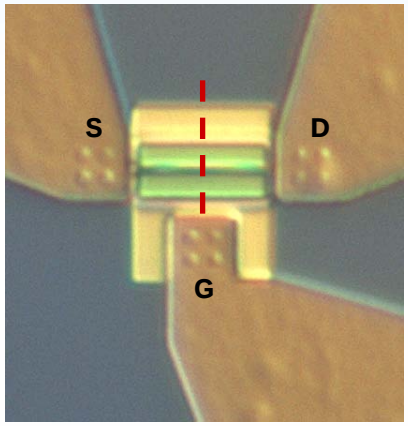
long range inputs (optical)

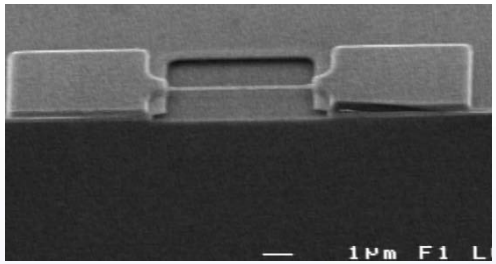
“cortical column” - hierarchical, temporal memory
(On Intelligence, Jeff Hawkins)

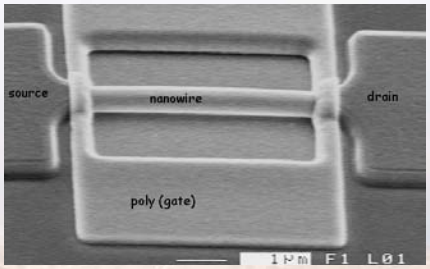
2) Develop new devices and architectures

Si nanowire devices



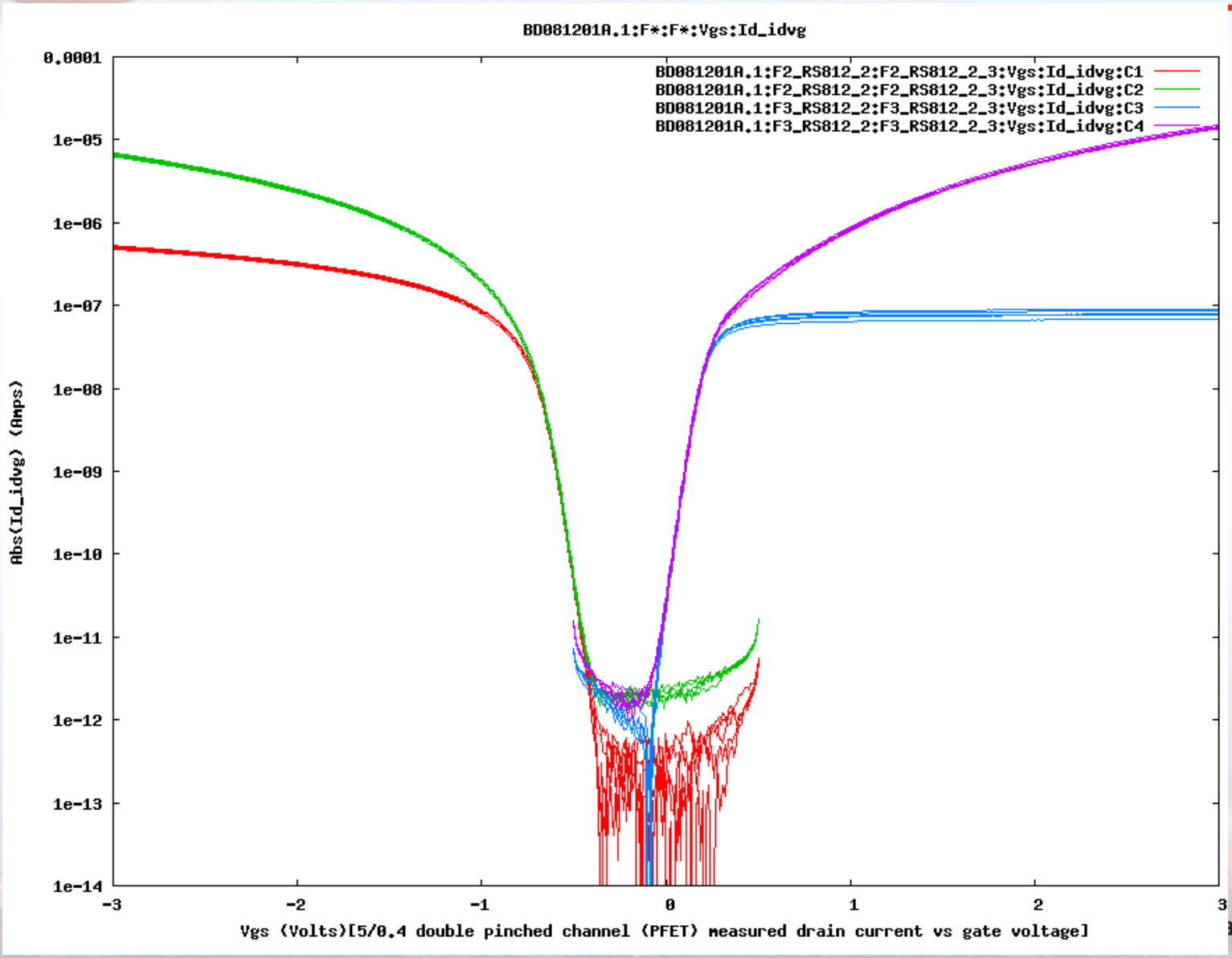






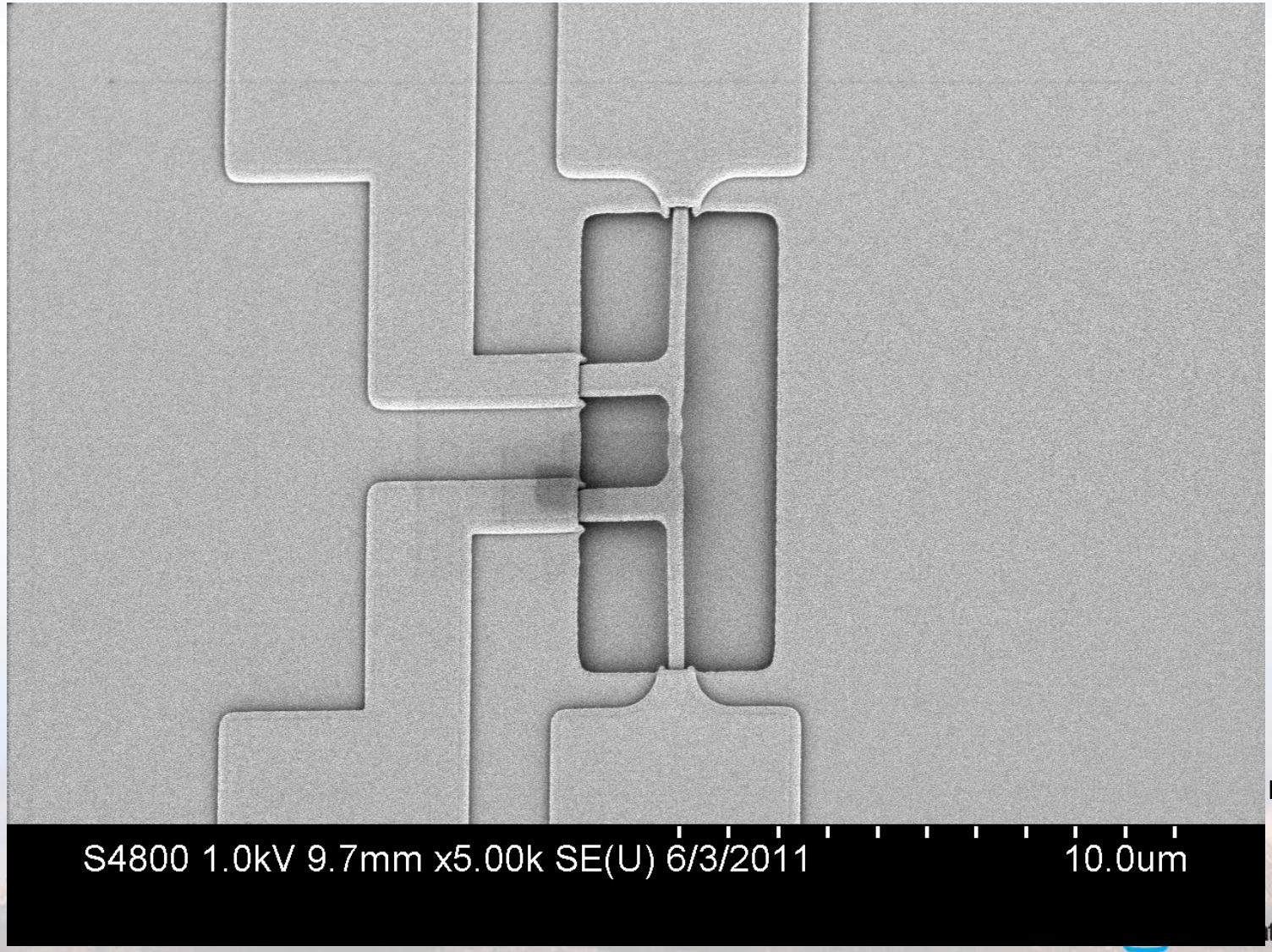
2) Develop new devices and architectures

Si nanowire devices



2) Develop new devices and architectures

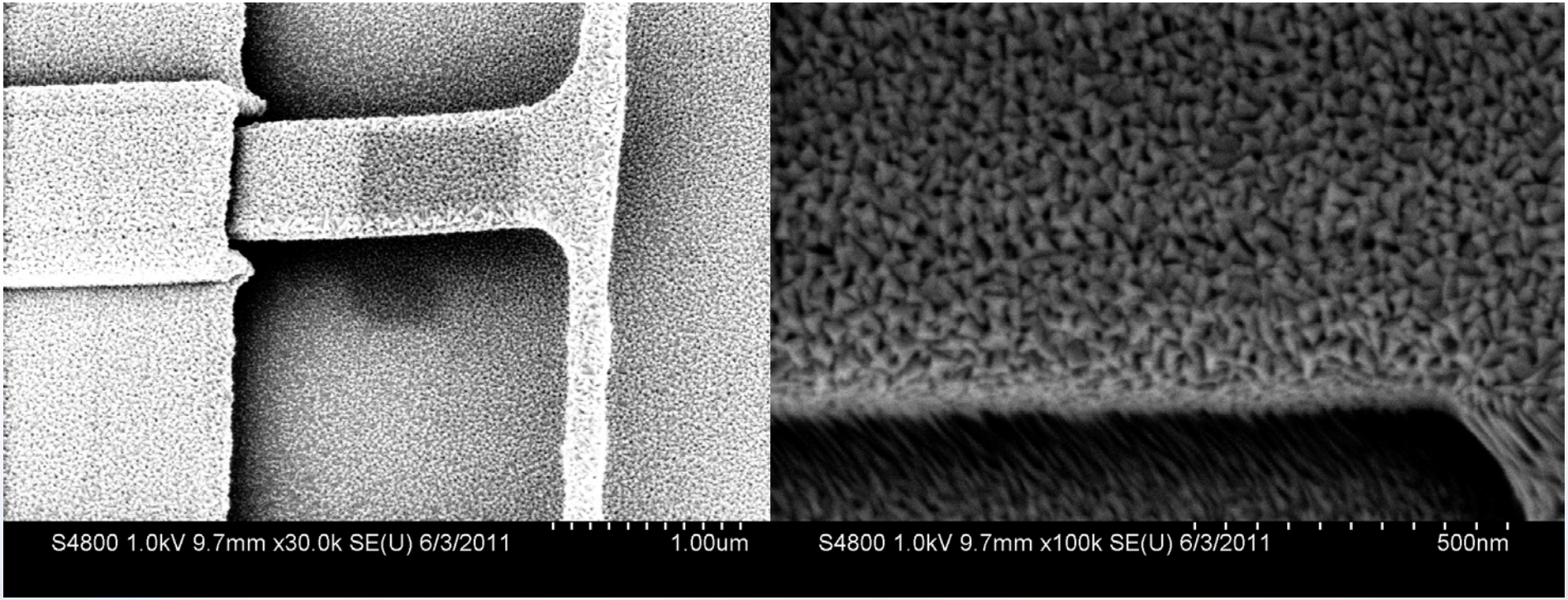
Si nanowire devices



D. Henry, 01748

2) Develop new devices and architectures

Si nanowire devices



D. Henry, 01748

3) Sub-systems for specific applications

Firefox

Lyric Semiconductor | Technology: Gates

http://www.lyricsemiconductor.com/technology-gates.htm

Google

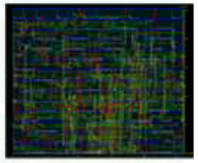
Lyric

Background | Technology | Company | Contact | News

- Overview
- Gates
- Architecture
- Programs
- Processing
- FAQ

Gates: The fundamental building blocks

Traditional LDPC primitive in TSMC 65 nm



scale: 20um

Lyric's LDPC primitive in TSMC 65nm

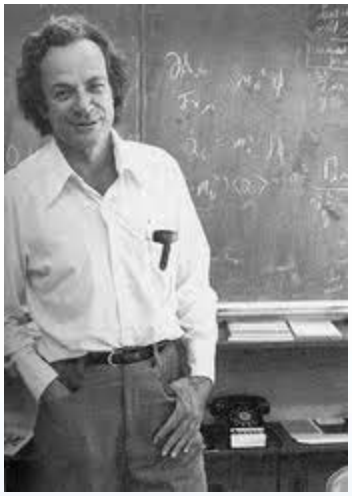
Smaller and lower power than a traditional implementation using 15nm CMOS

At the most fundamental level, computers are an assembly of gates that are used to perform the basic operations required to execute a program. For problems in the probability domain, even the values used in these most basic operations are not constrained to be either a 0 or a 1. Instead, the basic gates must determine the *probability* that a bit is a 1, or the *probability* that it is a 0.

Lyric's gates are designed to model relationships between probabilities natively in the device physics. For this reason, Lyric can perform mathematical operations in the probability domain with just a handful of transistors – creating power and area savings of more than 10X over traditional implementations.

Phone: 617-401-2750 | info@lyricsemiconductor.com

What we 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

Potential Applications

Native probabilistic computation

- low power, dynamic
- robotics, communications, …

Large datasets

- Fraud prevention, anomaly detection, …
- without needing >MW power levels

Modeling large, complex, probabilistic systems

- Physics, anthropology, economics, markets, (history?), …
- Uncovering patterns not readily observable (comprehendible?)

Link between electronic and biological (neuro) systems

- Neural prosthesis
- Augmentation