Introduction to VLSI Systems

Andreas G. Andreou
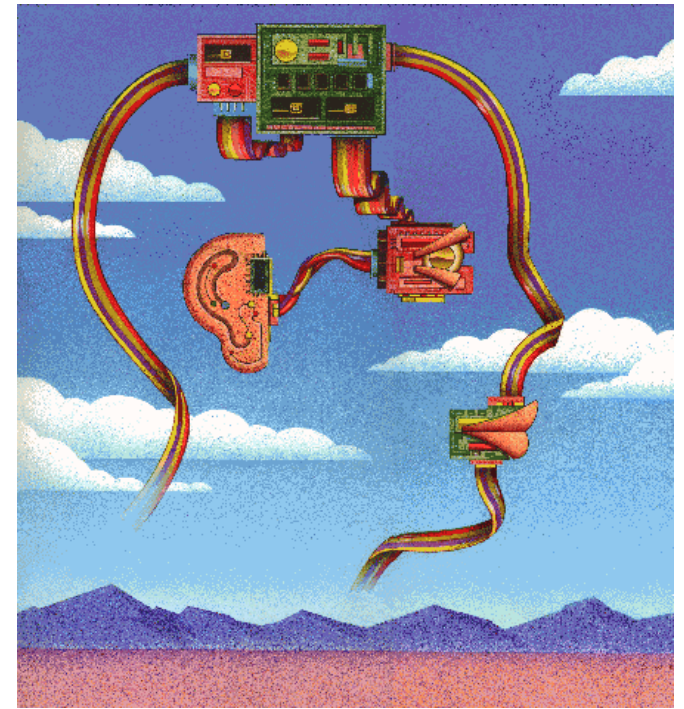
andreou@jhu.edu
Electrical and Computer Engineering
Center for Language and Speech Processing
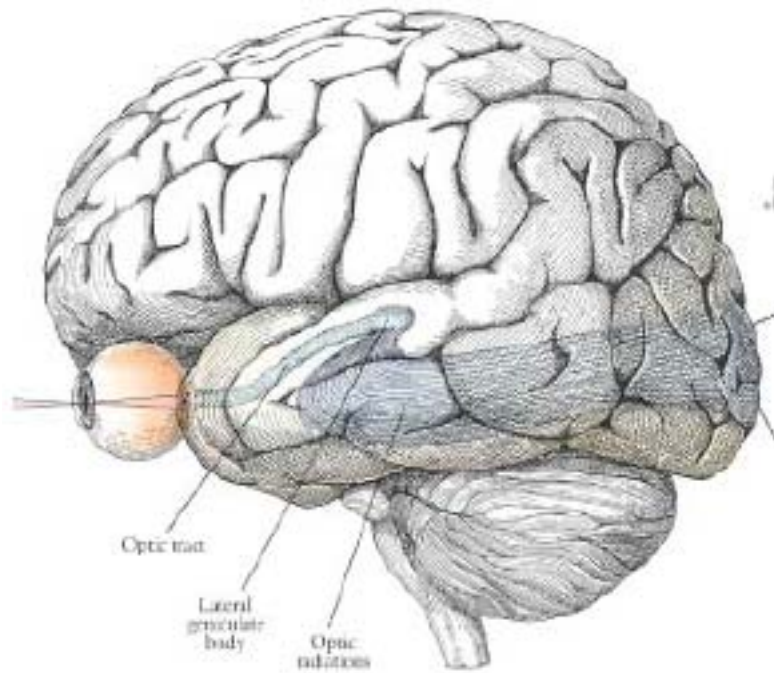Johns Hopkins University
http://www.ece.jhu.edu/faculty/andreou/AGA/index.htm

natural and synthetic Very Large Scale Integrated (VLSI) systems

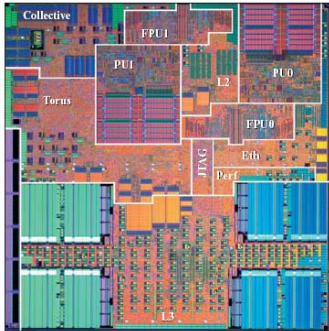The Brain   IBM Blue Gene/L
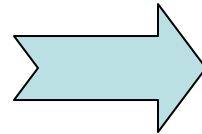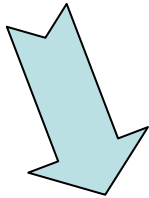

exist in three dimensional physical space but can
deal with problems in hyper-dimensional spaces

let's see what's inside Blue Gene/L supercomputer

15 W
130nm Bulk CMOS

25KW

- ac–dc supply
- Link cards, cables not shown
- Node cards
- Blue Gene/L midplane
- Ethernet switch card
- Gigabit Ethernet I/O cables (green)
- Air intake, plenum removed
- Fan modules
- Hot-air plenum
- Clock distribution (yellow)
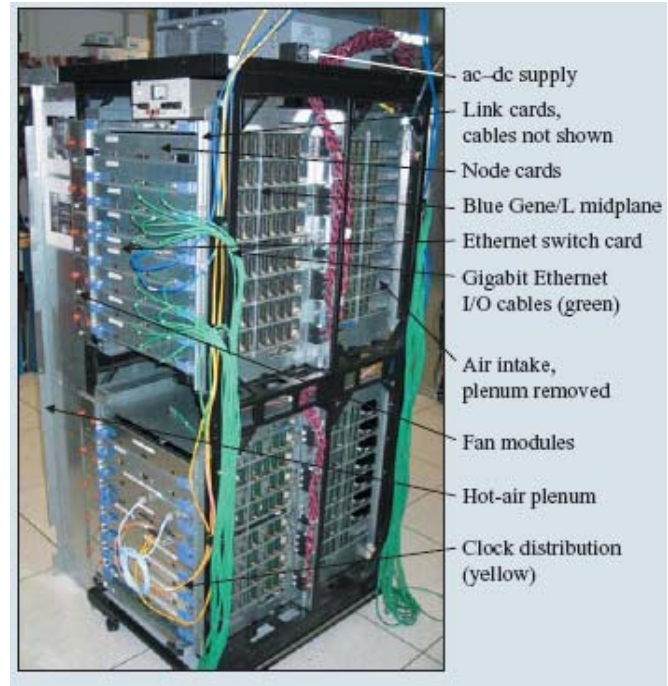
Coteus et.al., IBM J Res Dev, vol. 49, No.2, 2005

Moore's law: more of the same or no moore

## Cramming more components onto integrated circuits
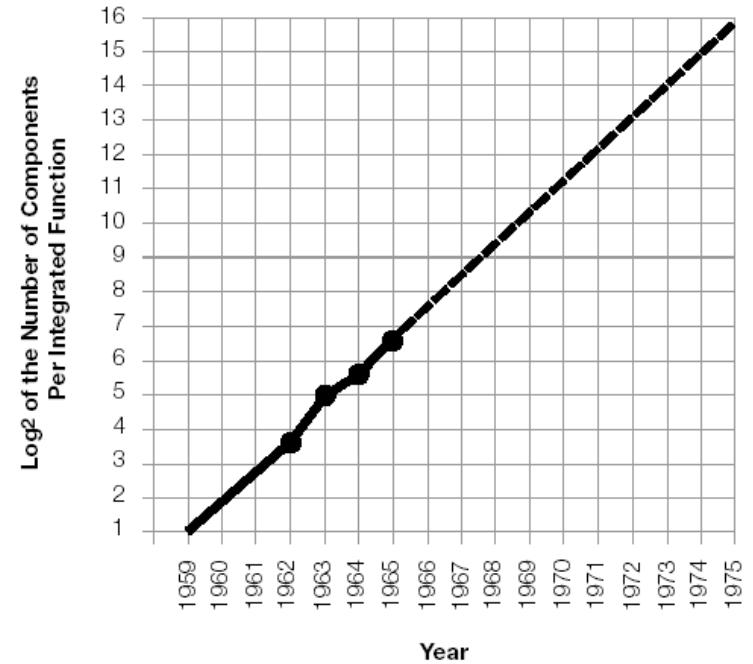
With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Electronics, Volume 38, Number 8, April 19, 1965

1. More transistors per unit silicon area
2. Lower energy costs for computation

Intel Core 2 Duo "Extreme" CPU
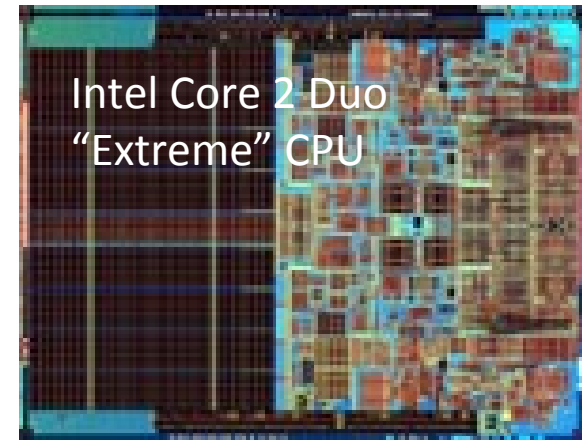
200,000,000 components

CCD to CMOS: the paradigm shift in camera technologies
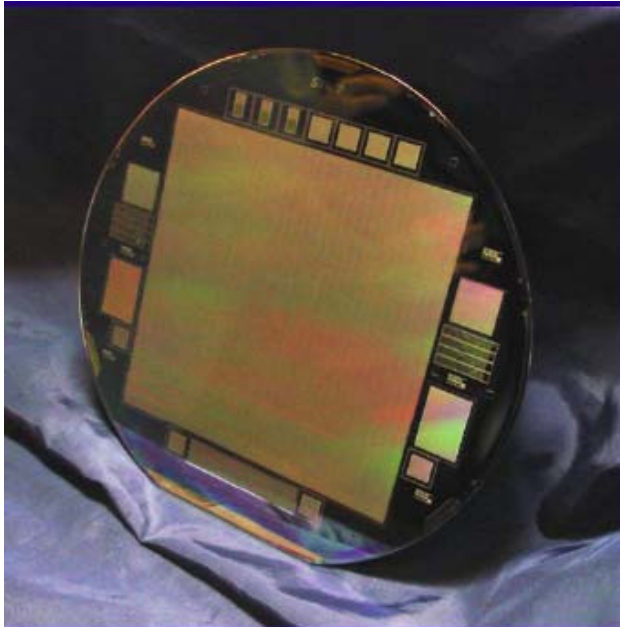
CCD state of the art

Full 6 inch wafer
111,000,000 pixels
1 frame per second!
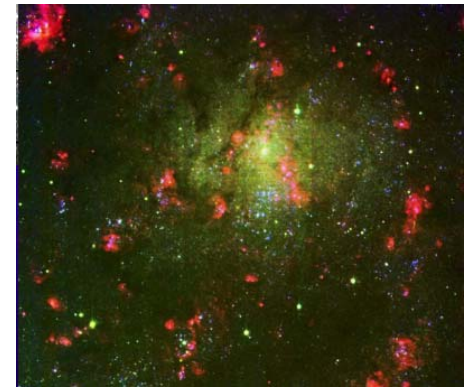
Semiconductor Technology Associates

$10,000,000

CMOS Cameras

1,300,000 pixels

$10

8,100,000 pixels

$ 100

21,000,000 pixels

$ 5000

and moore or less

…and the end of the single core processor paradigm

2005-2007: Sun, IBM, AMD, Intel
shift to shipping multicore processors
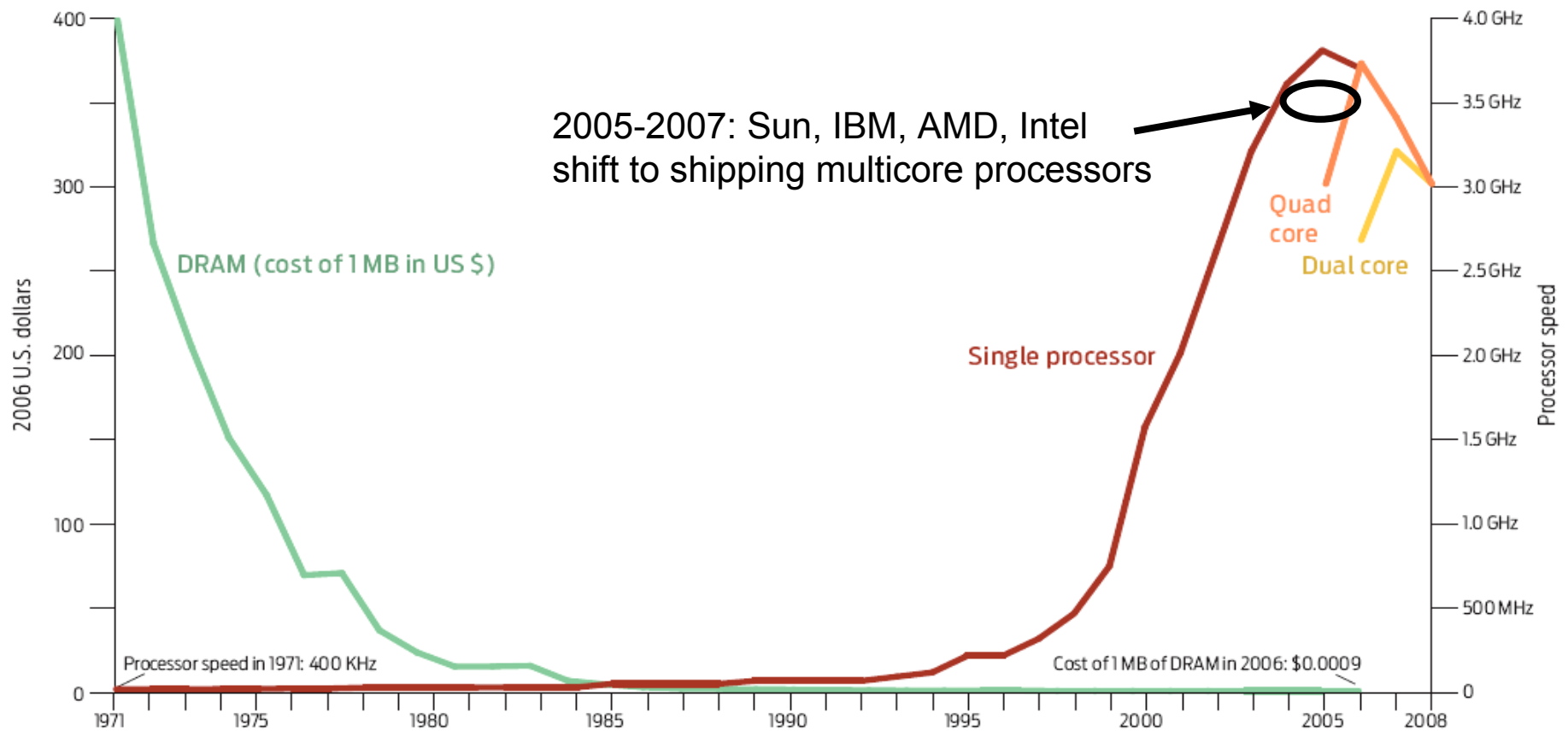
S. Adee, "The Data: 37 Years of Moore's Law," Spectrum, IEEE, vol. 45, no. 5, pp. 56, May 2008.

CPUs, DSPs, FPGAs and FPNAs
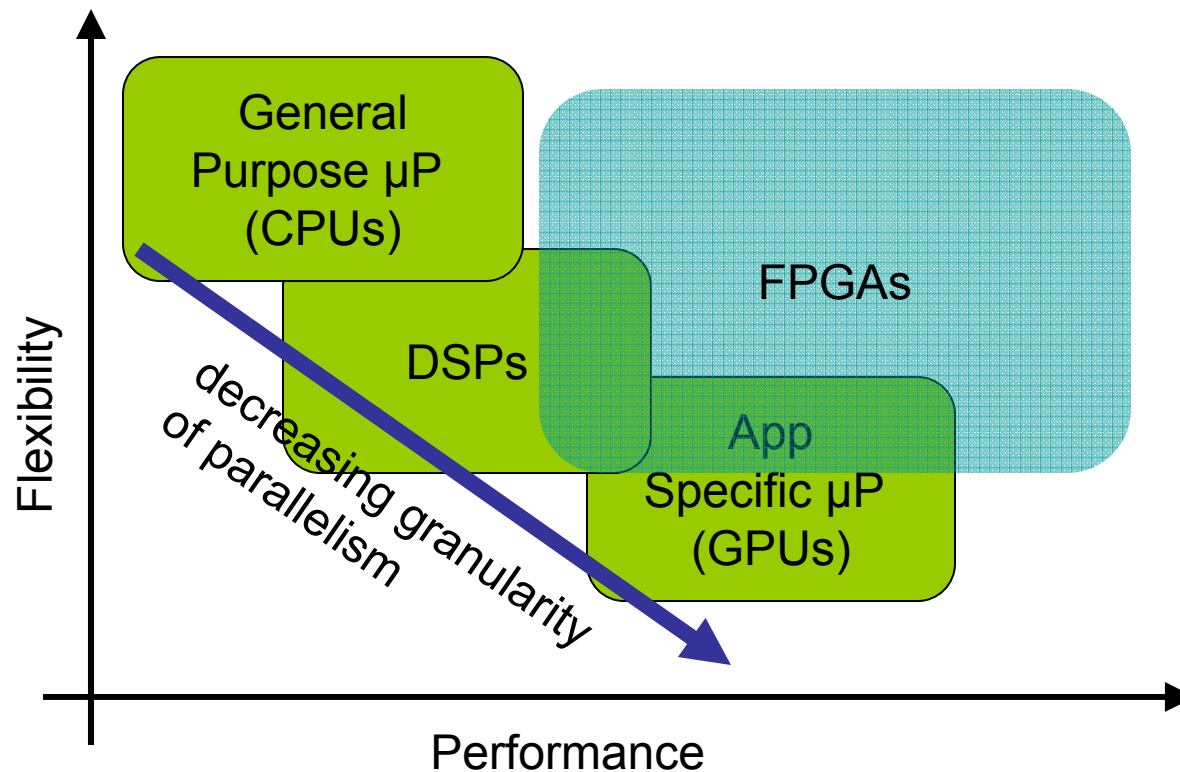
- Field Programmable Gate Arrays (FPGAs); Why? fine-grain parallelism with efficient communication flexibility (software approaches)

The Future

nano-CMOS
SOI-CMOS
3D-CMOS

Coulomb Blockade
Interband Tunneling
Resonance Tunneling
Giant Magnetoresistance

Emerging

Plastic Electronics
Photonic Crystals

Microsystems

Self Assembly
Soft Lithography
uFluidics

Technologies

Carbon Nanotubes
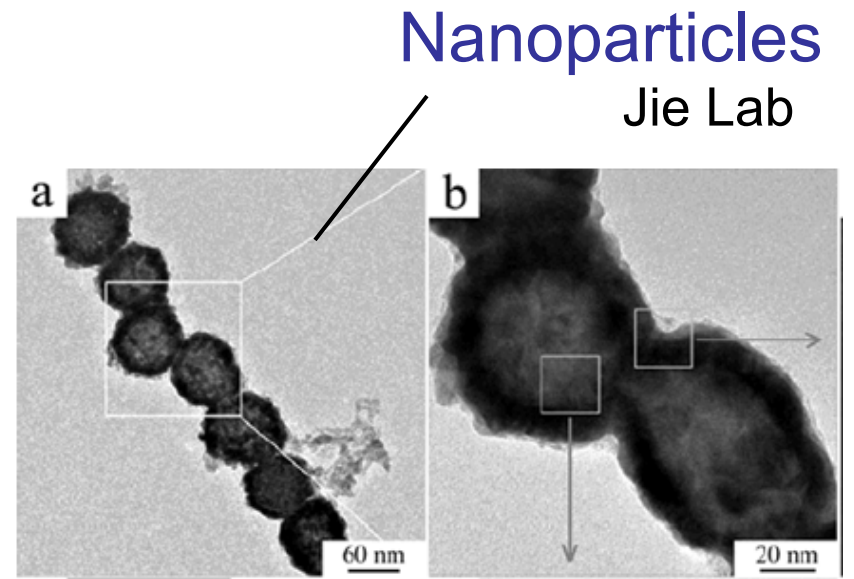**Molecular Devices**

Quantum Computing
DNA Computing
Quantum Cellular Automata

Single Quantum Flux
Electron Interference
Spintronics
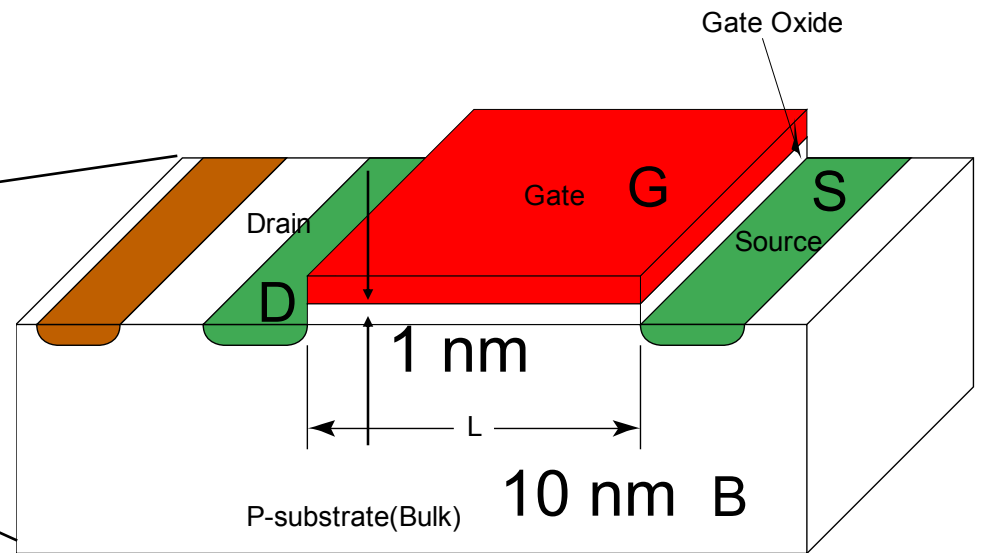
technologies: year 2027

Nanoparticles

Jie Lab

Blue light

a

b

60 nm

20 nm

500 nm

MOS transistor

100 μm

Human Hair

Gate Oxide

10 μm

Human Cell

100000000 X

Drain

Gate

G

S

D

Source

1 nm

L

P-substrate(Bulk)

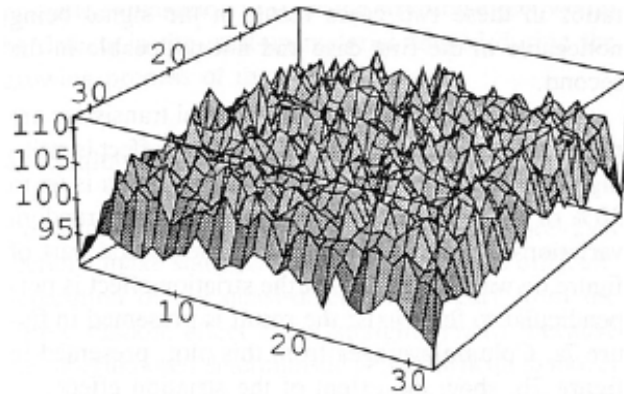10 nm B
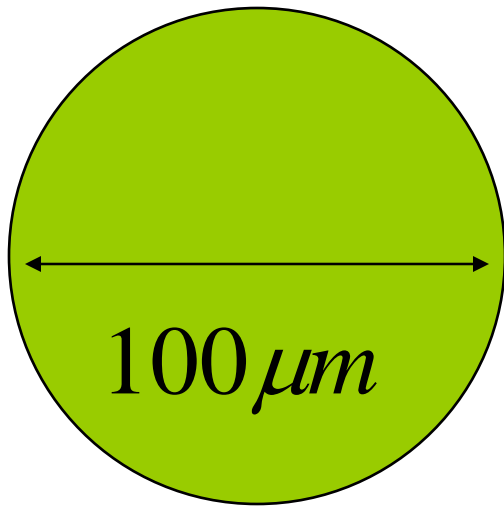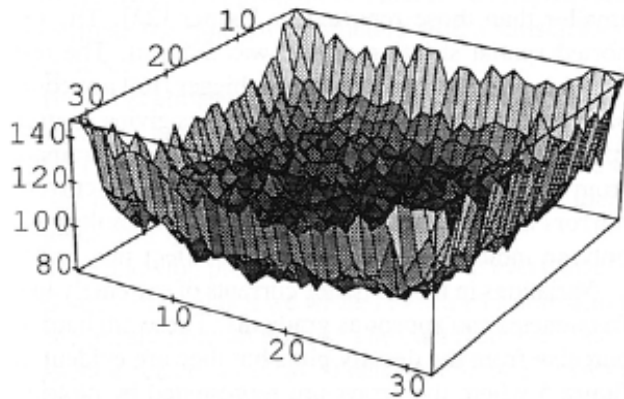
device variability and stochasticity

Fig. 4. Drain current (in pA) as a function of transistor location: (a) n-channel transistors in an n-well process, (b) p-channel transistors in an n-well process. Both arrays are located on the same test strip.

Pavasovic, Andreou, JVLSI 1994

Device mismatch!
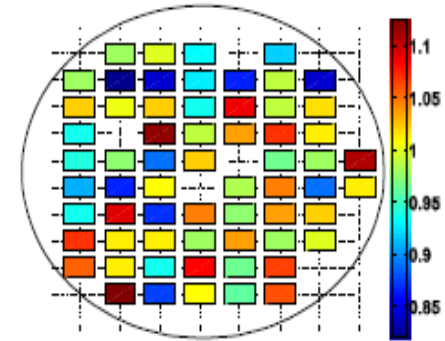
Fig. 3. Standard deviation of local threshold voltage mismatch in sub-100nm PD-SOI technology [4].
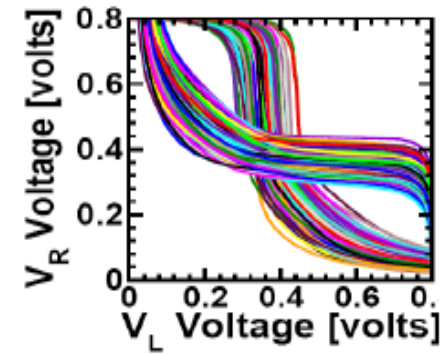
Fig 4: Variation of SRAM static noise margin in sub-100 nm PD-SOI technology [5].

Variability and Power Management in sub-100nm SOI Technology for Reliable High Performance Systems

Koushik Das, Kerry Bernstein, Jeff Burns, Fadi Gebara, Shih-Hsien Lo, Kevin Nowka, Rahul Rao and Michael Rosenfield

IBM Research Division, PO Box 218, Yorktown Heights, NY 10598

IEEE SOI Conference 2008

Introduction to VLSI Systems

11

VLSI systems research in the Andreou Lab

circuits: analog, digital and beyond …

CVDT
Continuous-Value
Discrete-Time

Continuous-Value
Continuous-Time
CVCT

CCD
Switched Capacitor

Linear and non-linear analog

Binary digital
Multivalue digital

Asynchronous digital
Neuron spikes
EPSP
Anisochronous Pulse Time Modulation

DVDT
Discrete-Value
Discrete-Time

Discrete-Value
Continuous-Time
DVCT
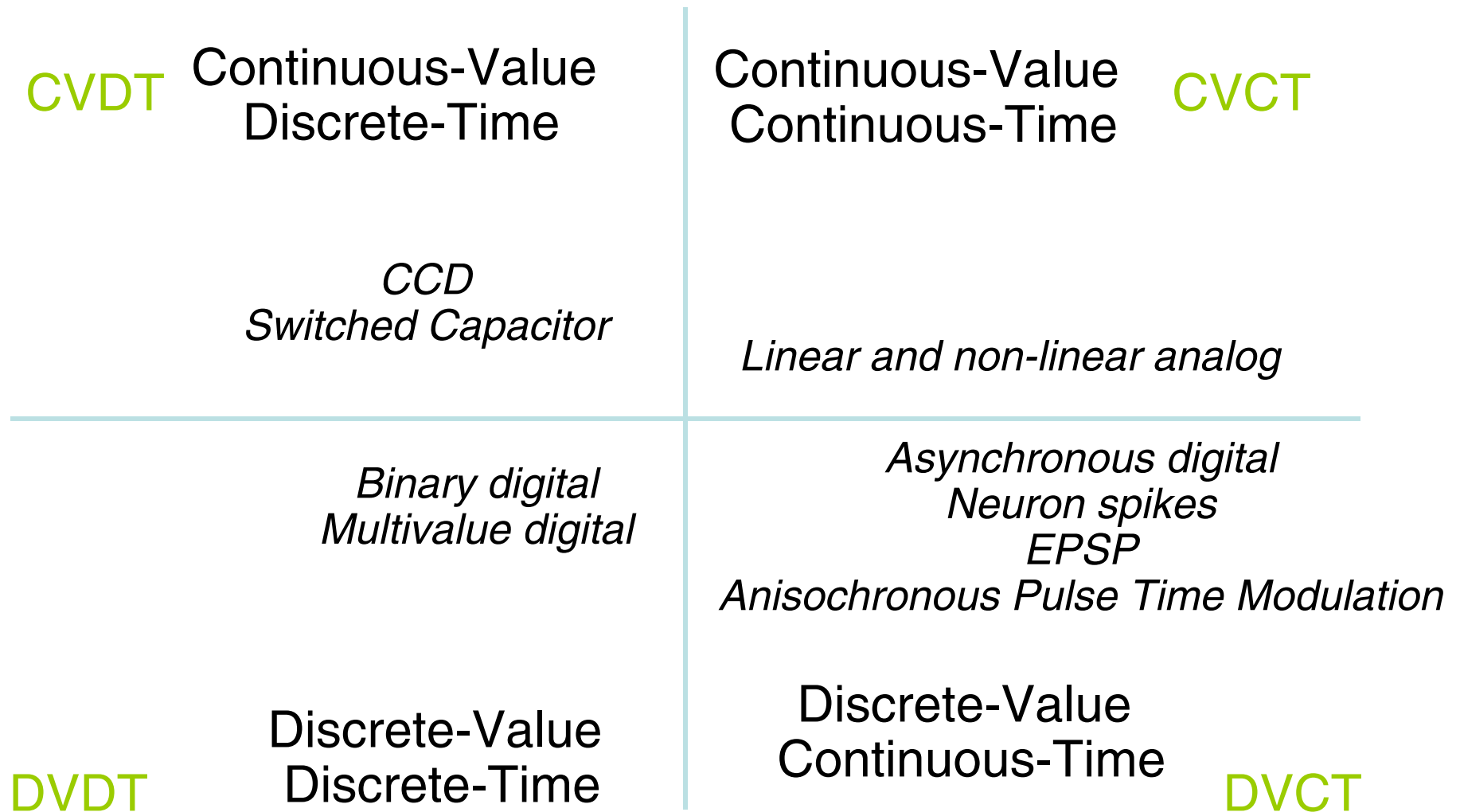
P.M. Furth and A.G. Andreou, "Comparing the bit-energy of continuous and discrete signal representations," *Proceedings of the Fourth Workshop on Physics and Computation* (PhysComp96), T.Toffoli, M. Biafore and J. Leao eds., New England Complex Systems Institute, pp. 127-133, Boston, MA, November 1996.

Introduction to VLSI Systems

1986: Let the physics do the work!

$$\max_{i=1,N}\left(\frac{|\vec{x} \wedge \vec{c_i}|}{a + |\vec{c_i}|}\right)$$

System / Architecture

Circuits

Devices / Technology

$$\sum_i Q_i = 0$$

$$\sum_i I_i = 0$$

$$V(t) = \frac{1}{C}\int_0^t I(t)\,dt$$

$$\sum_i V_i = 0$$

$$I_D \quad S\,I_{n0}\exp\frac{\kappa_n V_{GS}}{V_t}$$

Analog VLSI and Neural Systems

Carver Mead
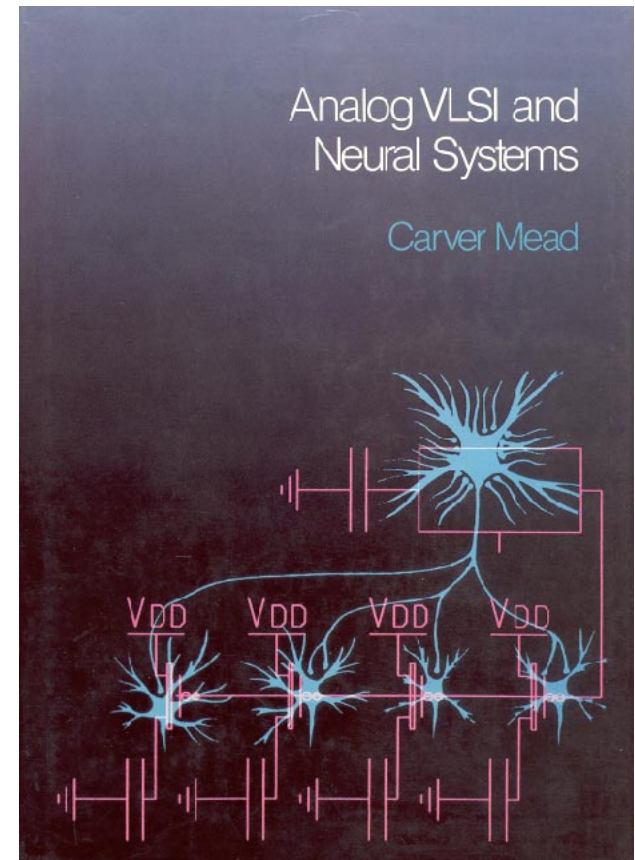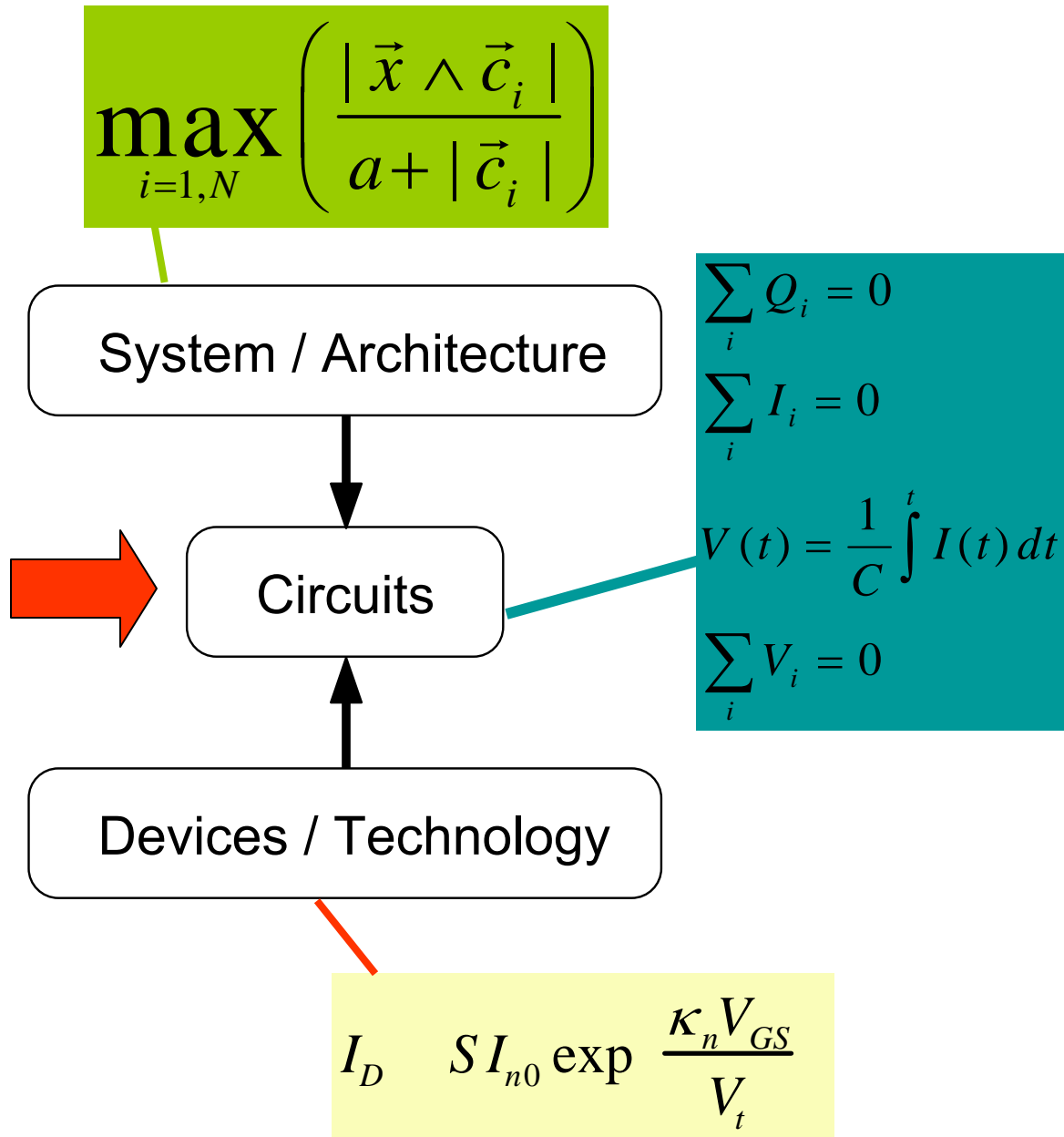
October 1986 (1st Draft)

embedded analog computing in digital memories

$$\max_{i=1,N} \left(\frac{|\vec{x} \wedge \vec{c}_i|}{a + |\vec{c}_i|} \right)$$

**Winner-Takes-All Associative Memory:
A Hamming Distance Vector Quantizer**

PHILIPPE O. POULIQUEN,[1] ANDREAS G. ANDREOU[1], AND KIM STROHBEHN[2]

[1]{philippe,andreou}@olympus.ece.jhu.edu, [2]aleph0@apicomm.jhuapl.edu
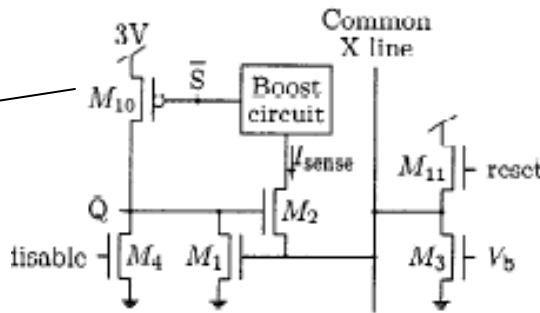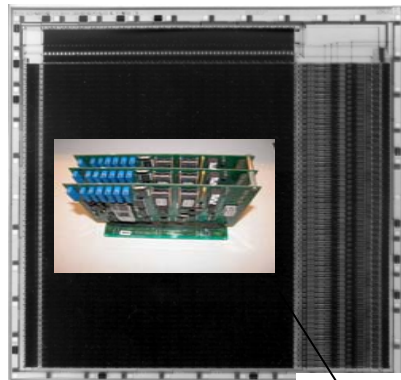[1]Electrical and Computer Engineering, Center for Language and Speech Processing, Johns Hopkins University, 3400 N. Charles Street, Baltimore MD
[2]Applied Physics Laboratory, Johns Hopkins University, Laurel MD 20723 USA
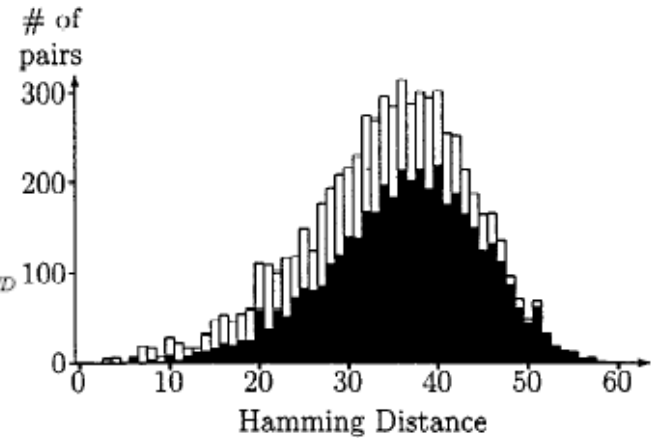
exploiting problem statistics!

pose processor.) In an DEC-Alpha based general purpose computer it takes 10000 cycles to do a single pattern matching computation and thus it takes a total of $20\mu s$ per classification. Power dissipation is 30W at 500 MHz and therefore the energy per classification is $600\mu J$ The Pentium-Pro is worse, because it requires 30W at 150 MHz and more than 10000 cycles for a single pattern matching. In contrast, the total current in the WAM is: $(124 \times 116 \times 10)$ nA continuous bias current for the memory cells at 5V. Computation time is approximately $70\mu s$ for a total energy per classification of approximately 100 nJ. The power dissipation in
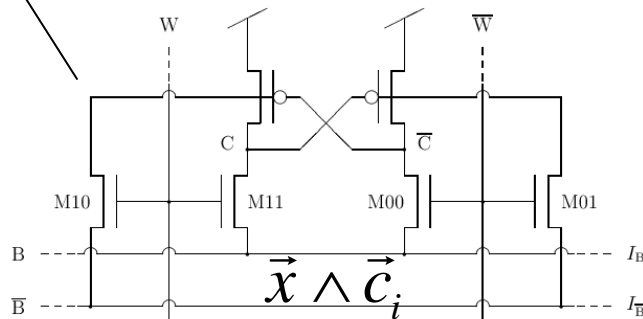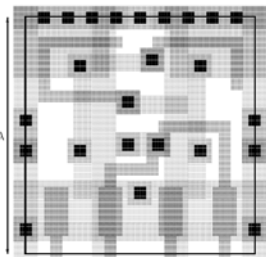
minimal complexity CMOS circuits

$$\vec{x} \wedge \vec{c}_i$$

what did we learn?

1. Memory and processing are integrated in a single structure; this is analogous to the synapse in biology.

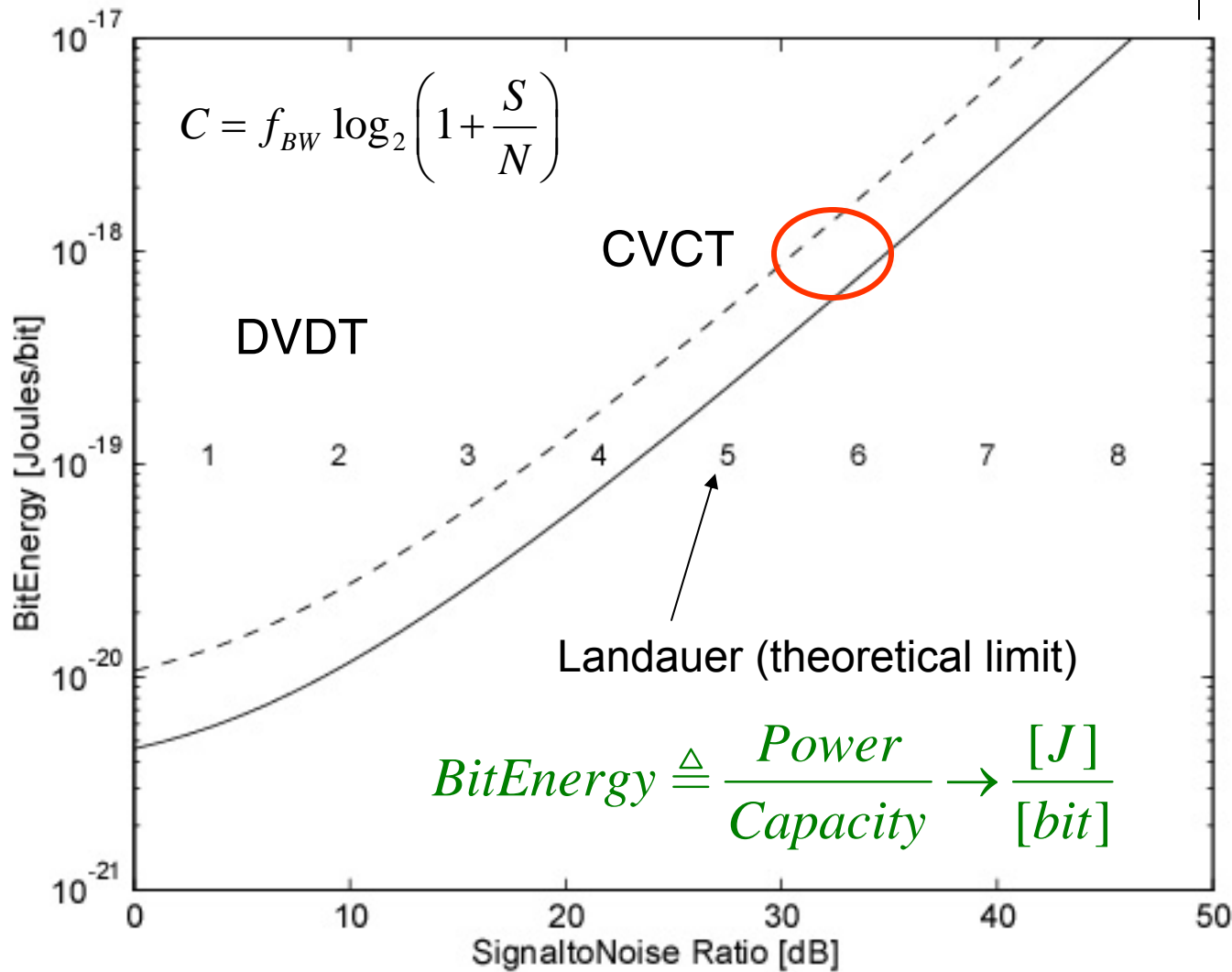2. The system has an internal model that is related to the problem to be solved (*prior* knowledge). This is the template set of patterns to be classified.

3. The system is capable of learning i.e. templates can be changed to adapt to a different character set (different problem). This is done at the expense of storage capacity—we use a RAM based cell instead of a more compact ROM cell–.

4. The system processes information in a parallel and hierarchical fashion in a variable precision architecture. I.e. given the statistics of the problem, most of the computation is carried out with low precision (three or four bit) analog hardware.

5. The system is fault tolerant and gracefully degrades. The same structures that is used in the *precision-on-demand* architecture can also be used to reconfigure the system for defects in the fabrication process. The components of the chip that are worse matched can be disabled during operation.

the energy costs of computing

8-9 bits

DVDT practical limit at 10nm CMOS

$\sim 10^{-16}$

$$C = f_{BW} \log_2\left(1 + \frac{S}{N}\right)$$

CVCT

DVDT

Landauer (theoretical limit)

$$BitEnergy \triangleq \frac{Power}{Capacity} \rightarrow \frac{[J]}{[bit]}$$

BitEnergy [Joules/bit]

SignaltoNoise Ratio [dB]

Izhikevich neuron model

- Fast variable (v), slow variable (u) dynamics:

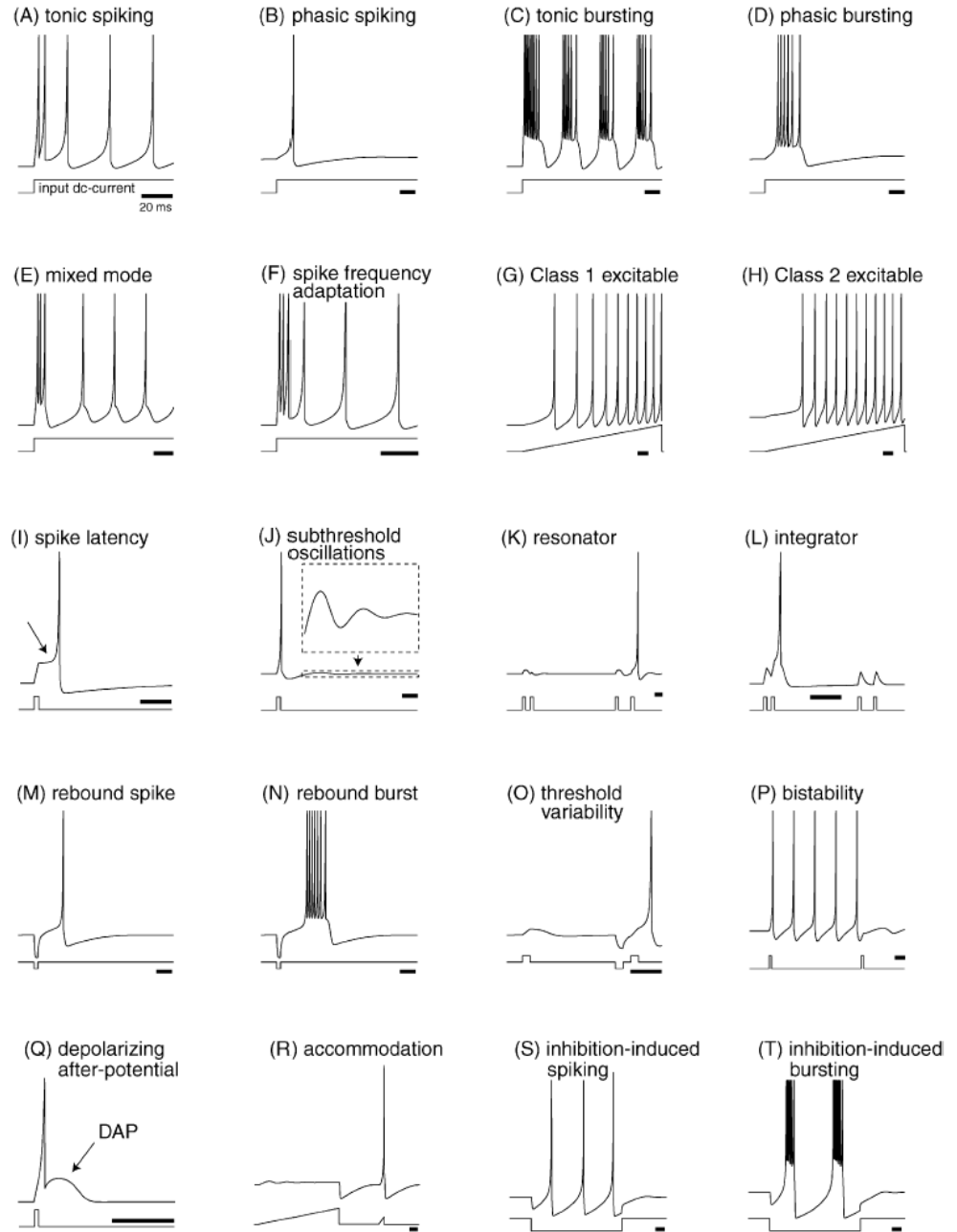$$v' = 0.04v^2 + 5v + 140 - u + I$$
$$u' = a(bv - u)$$

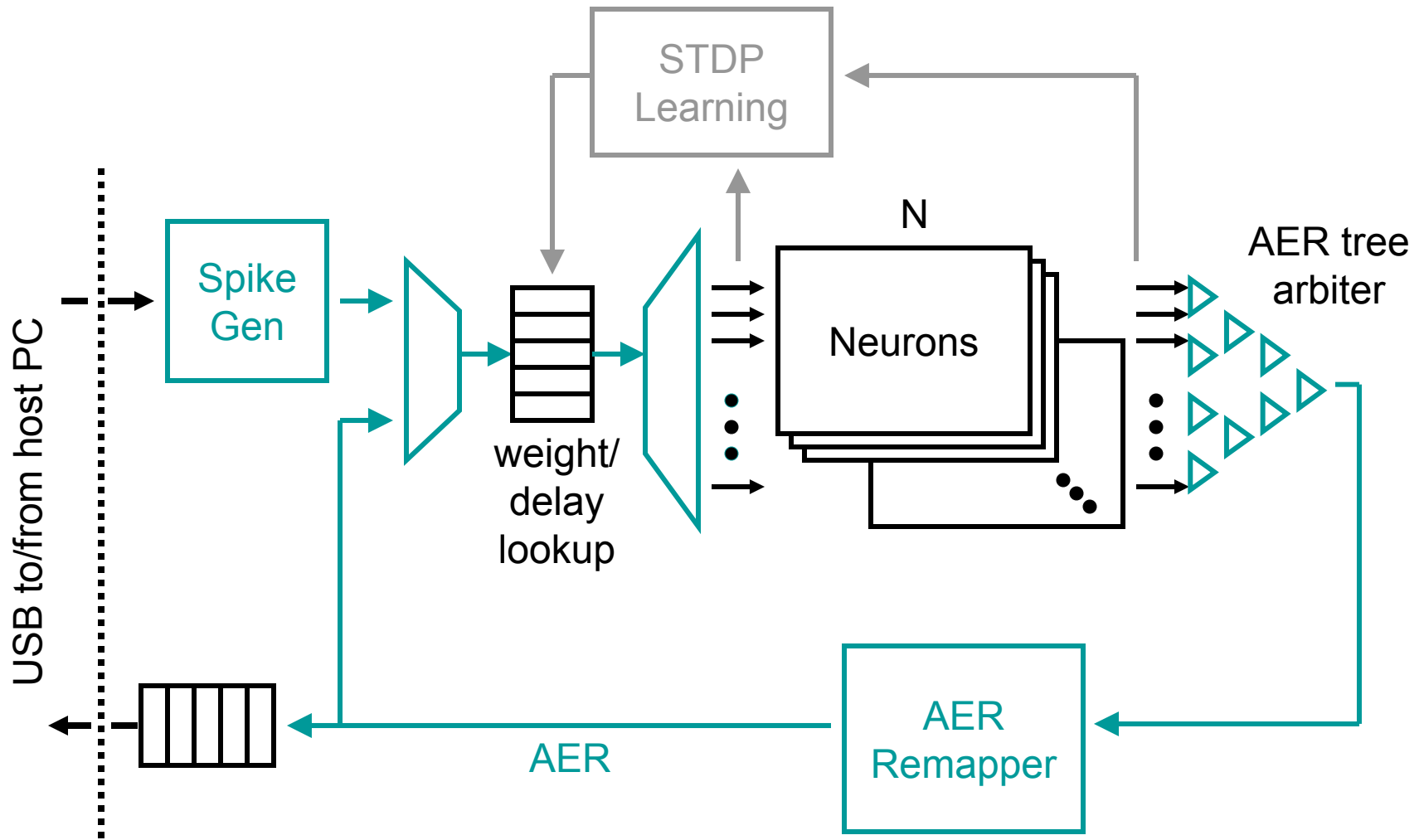- Reset condition:

if $v \geq +30$ mV, then:
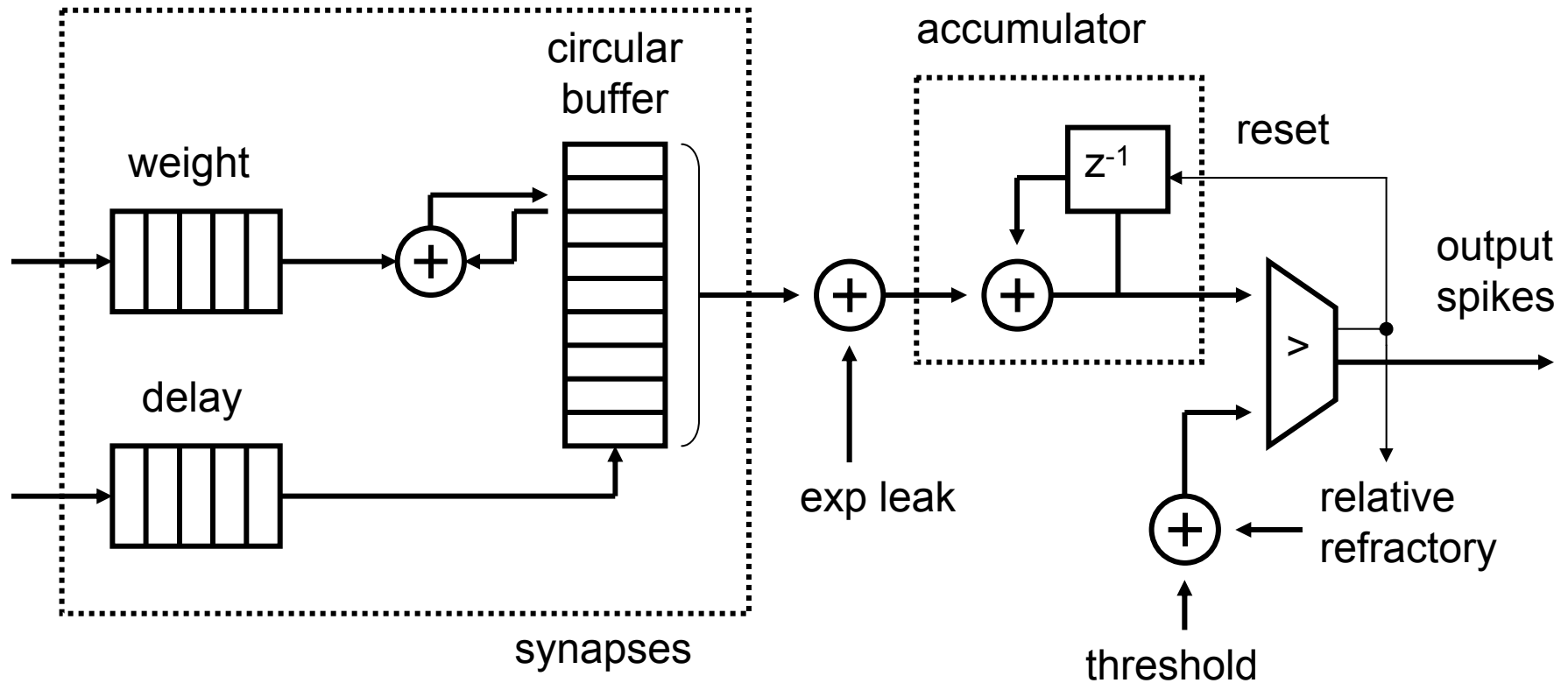
$$v \leftarrow c$$
$$u \leftarrow u + d$$

E. Izhikevich, "Which model to use for cortical spiking neurons?" IEEE Transactions on Neural Networks, vol. 15, no. 5, pp. 1063–1070, Sept. 2004.

(A) tonic spiking

input dc-current
20 ms

(B) phasic spiking

(C) tonic bursting

(D) phasic bursting

(E) mixed mode

(F) spike frequency adaptation

(G) Class 1 excitable

(H) Class 2 excitable

(I) spike latency

(J) subthreshold oscillations

(K) resonator

(L) integrator

(M) rebound spike

(N) rebound burst

(O) threshold variability

(P) bistability

(Q) depolarizing after-potential

DAP

(R) accommodation

(S) inhibition-induced spiking

(T) inhibition-induced bursting
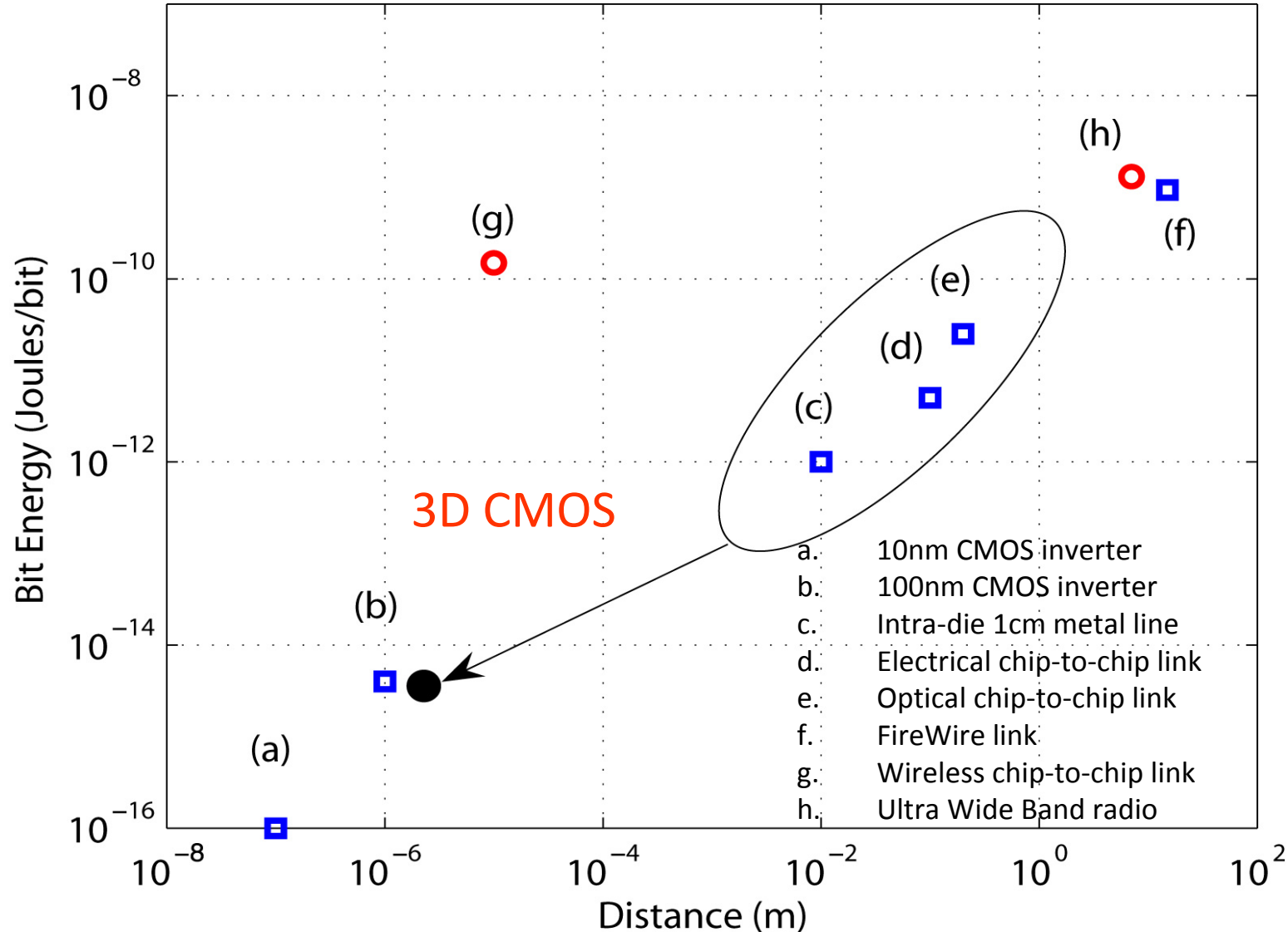
system architecture

micro-architecture

- 16-bit accumulator
- 8-bit synaptic weights
- 128 synapses per neuron

Andrew Cassidy, and Andreas G. Andreou. "Dynamical Digital Silicon Neurons." IEEE International Workshop on Biomedical Circuits and Systems (BIOCAS'2008).

the energy costs of communication

a.   10nm CMOS inverter
b.   100nm CMOS inverter
c.   Intra-die 1cm metal line
d.   Electrical chip-to-chip link
e.   Optical chip-to-chip link
f.   FireWire link
g.   Wireless chip-to-chip link
h.   Ultra Wide Band radio

M.A Marwick and A.G. Andreou, "Retinomorphic system design in three dimensional SOI-CMOS,"
Proceedings of the 2006   IEEE International Symposium on Circuits and Systems.

3D CMOS

MIT Lincoln Labs
3Tier CMOS
180 nm SOI technology
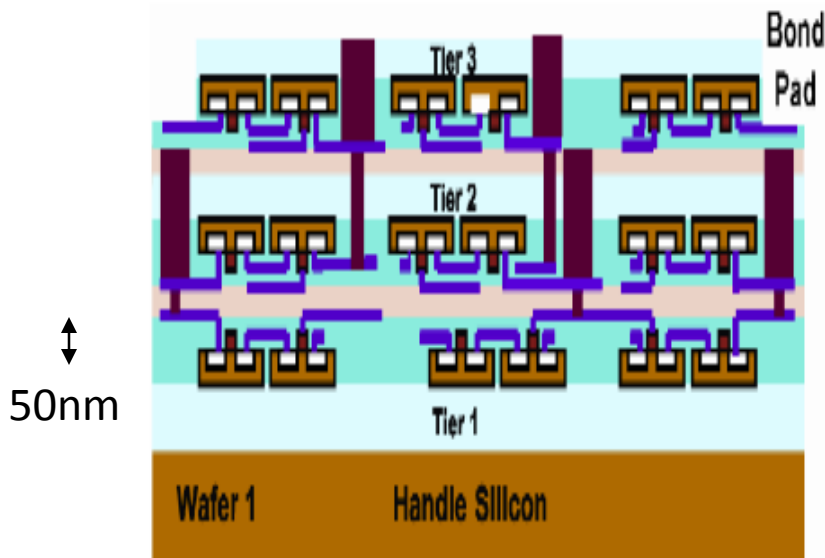
- Buried oxide thickness = 400nm

- Silicon substrate thickness = 40nm

- Inter-tier distance = ~7um

- Inter-wafer via = 1.75um x 1.75um

- Inter-wafer via pitch = 1.5 um

- Gate oxide thickness = 4.2nm

- 1.5 Volts, 3M1P process

Enables seamless, integration of heterogeneous wafers.
- Multi Vdd and Tox CMOS
- Multi material systems

multiproject 3D SOI-CMOS run

• Cadence Design Kit for multiple tier CMOS design environment

• 1st Multiproject Run: May 2005

• Chips back April and August 2006

• 2nd Multiproject Run: November 2006

• 3rd Multiproject Run: November 2008
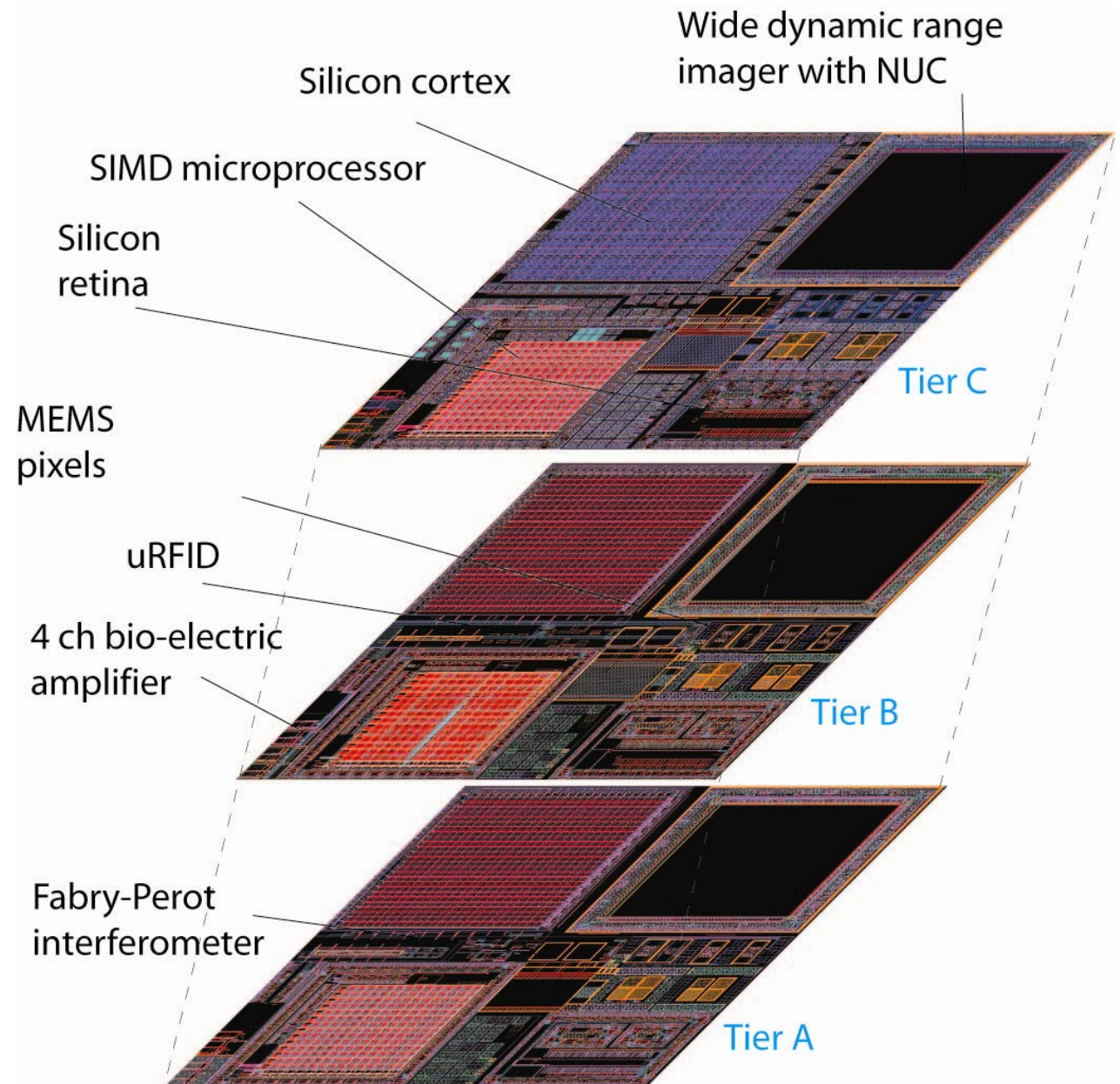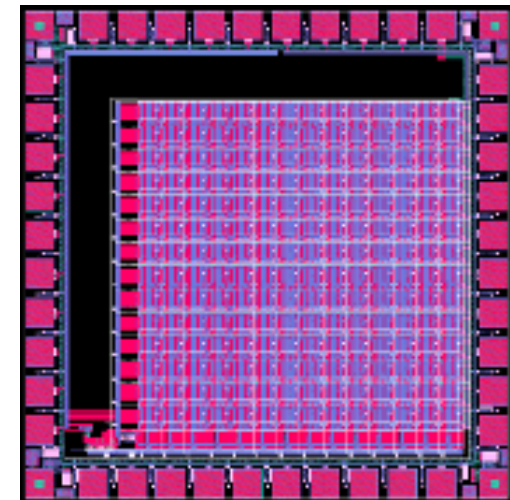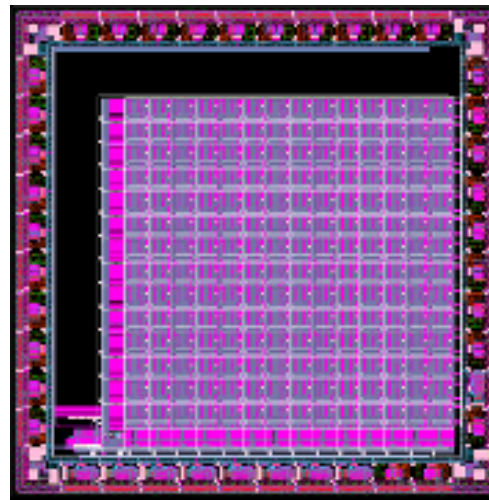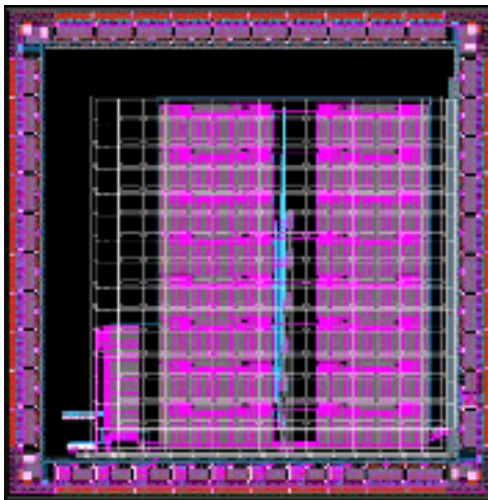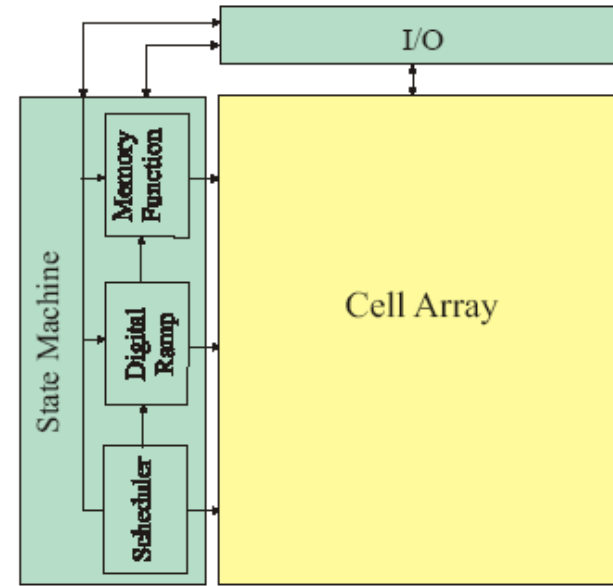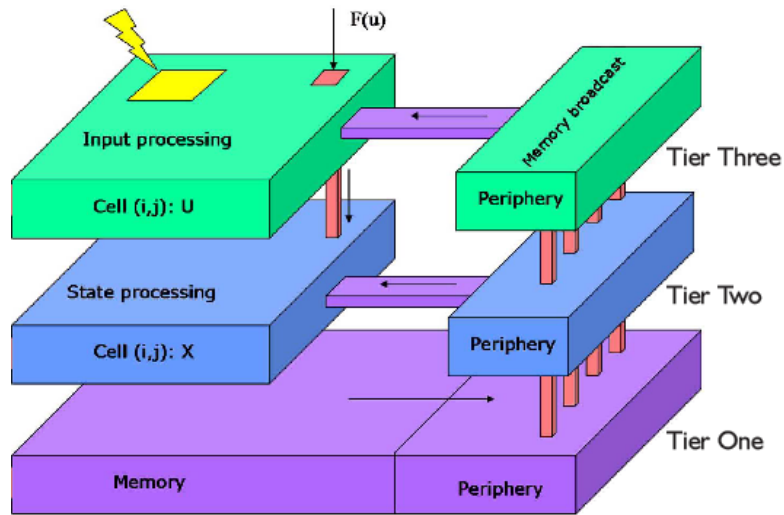
IBM Journal of Research and Development

3D Chip Technology

Historically, the steady growth of computer system performance depended on the performance of microprocessors, which depended on the scaling of devices and circuits to smaller dimensions. As scaling on the 2D surface of chips approaches practical limits, 3D technologies offer an opportunity for continued system improvements, even as the progress of scaling slows down. The eight papers in this issue describe the system design opportunities and challenges of 3D chip technology, as well as methods for producing dense arrays of through-silicon vias, thinned silicon, dense area-array silicon–silicon interconnection, chip stacking, and 3D wafer integration. Thermomechanical modeling and the implementation of 3D structures in products are also described.

digital 3D SIMD processor

asynchronous circuits in 3D CMOS

Results

gold – din
green – din ack
red – dout
purple – dout ack

- **Block Diagram**
  - 5 asynchronous handshake buffers in each path
    - (4 deep FIFO + 1 MUX)
  - Utilizes all three tiers
    - Handshake between tiers

So how do fabricate our own chips?

http://www.mosis.org/

Key idea: Use a number of different manufacturers to contribute manufacturing capacity to multiuser projects.

- AMI 1500, 500 and 350 nanometer CMOS
- TSMC 350, 250, 180 nanometer CMOS
- IBM SiGe 250 and 180 nanometer BiCMOS
- IBM 45 nanometer SOI CMOS!

520.216 will teach you you how to go from a simple idea to a system (chip) that will solve some problem. You will do analysis, design and finally layout and simulation and fabricate your own chip!

Introduction to VLSI Systems

Intro VLSI and her friends