Future chip technologies:
Evolution or Revolution?

Steve Hillenius
Semiconductor Research Corporation

Ireland Summer School for Nanotechnology, Kildare, Ireland, August 28-30, 2011

Session Questions from Organizers

1. What will the driving force be?
2. How can simulation lead the way?
3. Could the transition be 'disruptive'
4. Who will be the leaders, winners or losers

Questions and Answers

1. What will the driving force be?

- **Answers:**
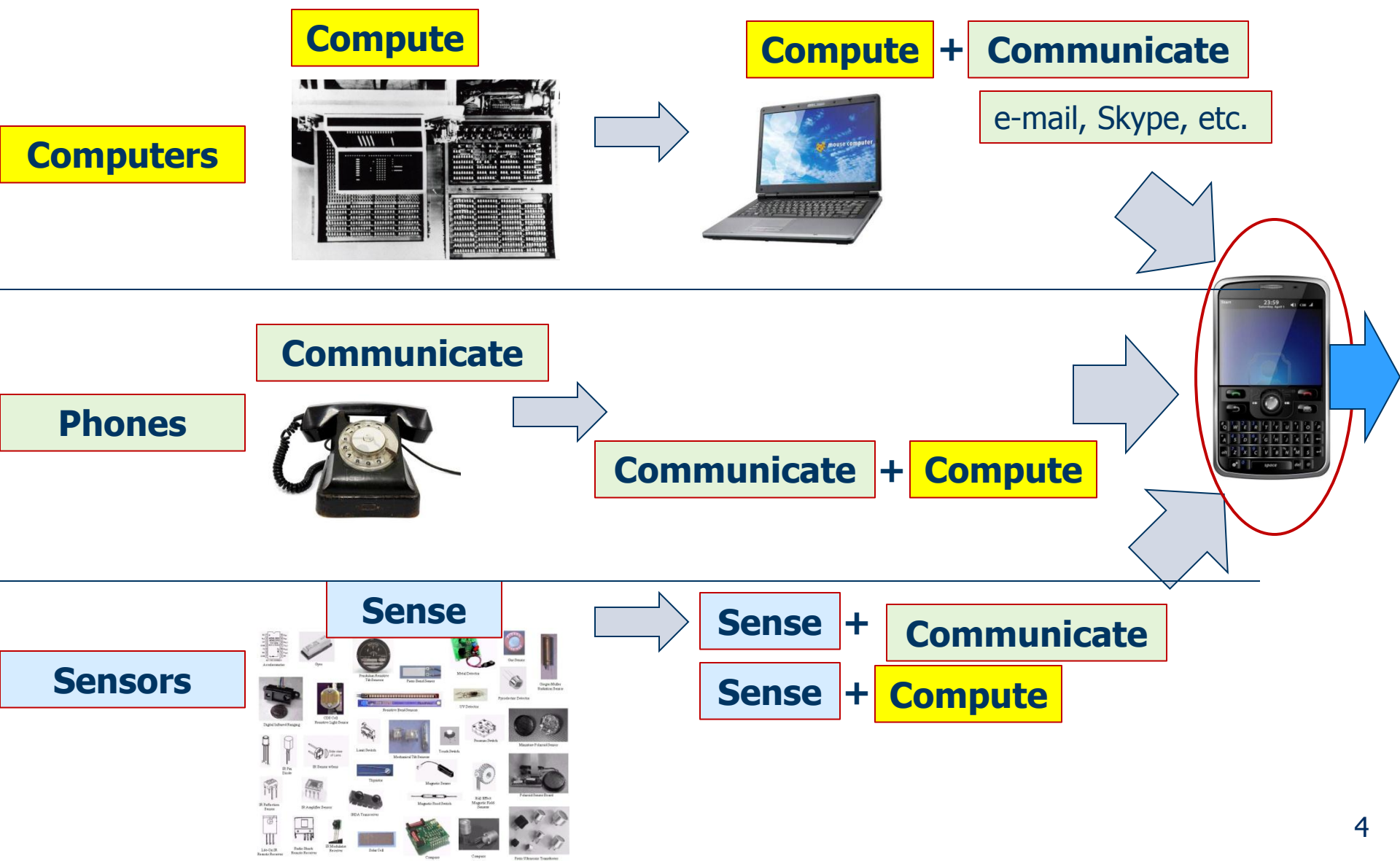 - Smaller size of the system

  - Lower energy

  - More 'work' done:
    - Computing
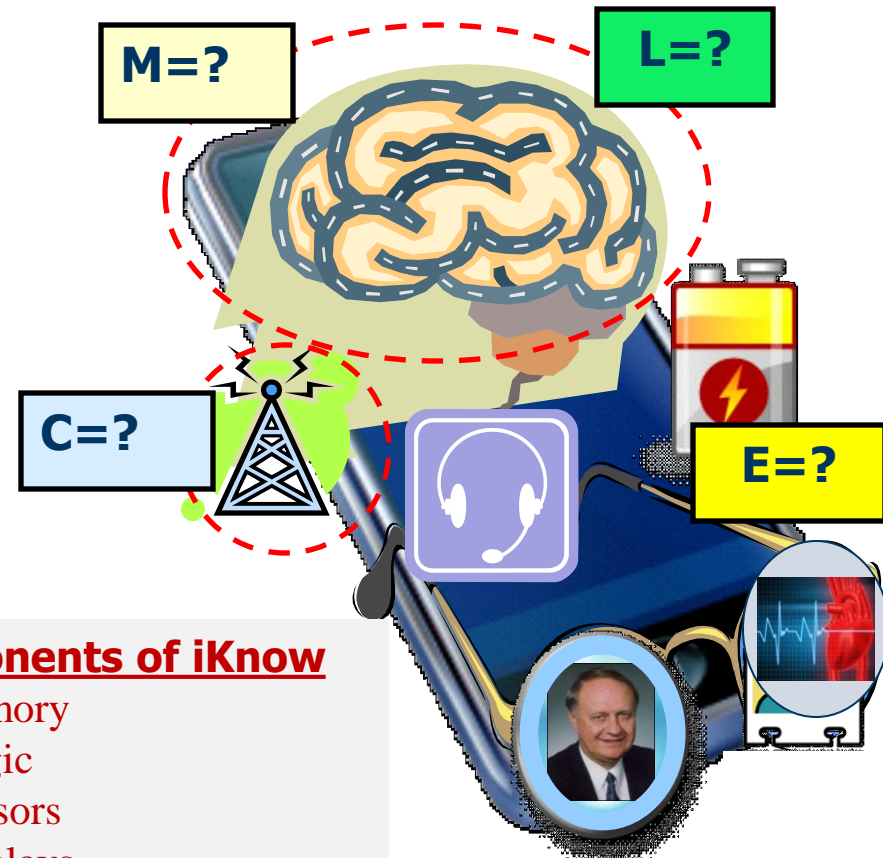   - Communication
    - Sensing

  Convergence!

The Economics of any application will be the overriding moderating factor

Disruption through Functional Convergence for a Feature-Driven Market

Computers — **Compute** → **Compute** + **Communicate** — e-mail, Skype, etc.

Phones — **Communicate** → **Communicate** + **Compute**

Sensors — **Sense** → **Sense** + **Communicate** / **Sense** + **Compute**

Introduction to the *iKnow Machine*

Thought Model - a hypothetical handheld device whose function is to enhance the life experiences of its user by providing a variety of support functions. It is envisioned that the set of support functions could be defined by the users to best meet their perceived needs.

Some of the *iKNOW* functions:

- **Real-time image collection, recognition and labeling**

- **Medical monitoring and reporting**

- **Communication**

- **Support for reasoning about questions across a wide set of problem domains**

- **Real-time language translation**

- **Entertainment etc.**

M=?

L=?

C=?

E=?

Essential components of iKnow

M - Memory
L - Logic
S - Sensors
D - Displays
C - Communication
E - Energy

Bio-inspired device evolution

- Electronic systems like biology: They are evolving to become more intelligent and connected
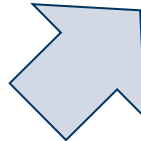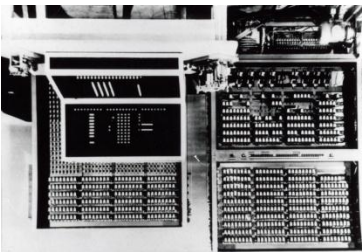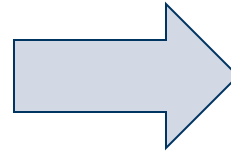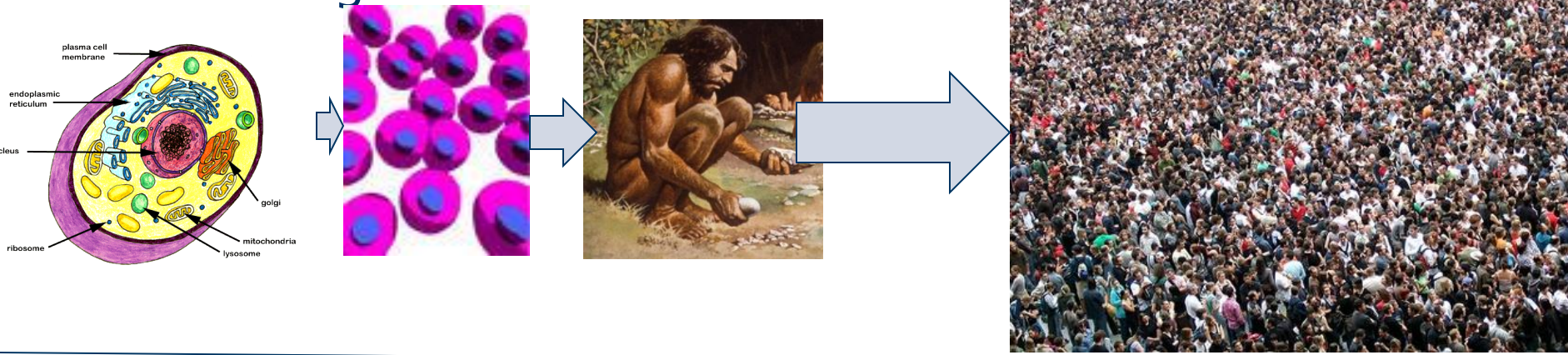
2. How can simulation lead the way?

- Answers:
  - Productive simulation usually occurs **after** you have a technology, not **before**
    - e.g. SPICE
- <u>First:</u> We need to have a design concept
- <u>Second:</u> We need experimental data (either ab-initio or from the design)
- <u>Third:</u> Simulation plays an important role to support understanding and optimization
- **Simulation Challenge:** Are there limits to how predictive simulation tools can be?

3. Could the transition to new IP technologies be 'disruptive' ?

- Answers:
  - Logic – not likely
    - No convincing alternatives yet
  - Memory – likely
    - Recent results on ReRAM are encouraging
  - Architectures – Promising!!!
    - Moving from logic-centric to data-centric?
  - More-than-More – Intelligent Sensor Networks
    - Central Nervous System of the Earth?

Disruptive Memory technologies: *The Past*

Semiconductor Landscape in 1982

256 Kbit DRAM in production
      2 μm minimum features

64 Kbit UV-EPROM in production
      No flash memory as we know it today yet exists

SRC 1982 challenges:
      64 Mbit DRAM
      0.25 μm minimum features

Semiconductor Vision (circa 1982):

- "It is doubtful that one can scale the device dimensions to **below 0.1 μm** and gain any advantage in circuit performance because of several **basic limitations**"

  > **Proc. IEEE (1983):** *A systems approach to 1 μm NMOS* by M.P. Lepselter, D.S. Alles, H. J. Levinstein, **G. E. Smith (2009 Nobel Prize Recipient)**, H. A. Watson

- "MOS gate lengths of about **0.25 micrometer** are **the practical scaling limit**" (1st SRC Annual Report–1984)
  - The SRC 0.25 micrometer CMOS research thrust is centered at Cornell University, with contributing projects at Wisconsin, Illinois, Stanford, Colorado State, Arizona, Yale, and Notre Dame

Table 3 Evolution of the Silicon MOS Random Access Memory (RAM) (1969–1988)

Disclosure Dates[a] Subm.	Pub.	Authors-Inventors Development Team	Institutions or Locations[b]	Memory Density (bit/chip)	Device and Technology[c]	Reduction to Practice[d]	Ref.
	1985	IBM Essex Junction		1M	DRAM NMOS 1T2d SAMOS	Prod	[154]
	1985	ATT, Fujitsu, Hitachi, Toshiba		1M	DRAM NMOS 1T2d	Eng	[154]
	1985	TI		1M	DRAM NMOS 1T3d trench-C	Eng	[154]
	1985	IBM E. Fiskill, Yorktown Ht		64k-4M	DRAM PMOS 1T3d trench-C	Lab	[169]
	1985	Hitachi, Toshiba, NEC		4M	DRAM NMOS 1t3d trench-C	Lab	[170]
	1985	Chatterjee et al.	TI	4M	DRAM NMOS 1T3d >1μm TCT	Lab	[173]
	1985	IBM Research	IBM	16M	DRAM NMOS 1T2d 0.5μm EB	Est	[198]
	1986	IBM-3090	IBM	1M	DRAM NMOS 1T2d SAMOS	Prod	[154]
	1986	MicroVAX-2 Toshiba/Chrislin		1M	DRAM NMOS 16MB/card	Prod	[154]
	1988	Matsushita, Toshiba, Hitachi		16M	DRAM CMOS 1T3d trench-C	Eng	[177]–[179]
	1995	SRC University Research		64M	DRAM CMOS 1T3d 0.25μm	Est	[130]

11

CT Sah, *Proc. IEEE* **(1988)**

Flash memory enabled:

The iPod was un-imaginable circa 1980

2006

iPod(5G)
80GB

126 IBM 3350 units needed!

$9,000,000 !!!
in 1976 dollars
(storage only)

Best available storage technology in 1976: IBM 3350

$70,000
$70,000
$70,000
$70,000
$70,000
$ $ $ $ $70,000

Each unit: 635 MB

Do Basic Research and Applications/Markets will follow!

3

Disruptive Memory technologies: *The Future*

Trend: The amount of data being created is exploding, growing significantly faster than Moore's law

Need: Data-centric information processing technologies

Flash may not get us there…

Flash memory: Device Challenges

- NAND flash is currently the workhorse NVM technology
- NAND flash has recently become an alternative storage technology
  - faster access times and smaller size, as compared to HDD.

- The NAND-based solid state drive (SSD) market has flourished recently.
- **Unfortunately**, there are several **fundamental limitations** of NAND flash for data-centric applications
  - poor endurance ($10^4 - 10^5$ erase cycles),
  - modest retention (typically 10 years on the new device, but only 1 year at the end of rated endurance lifetime),
  - long erase time (~ms), and high operation voltage (~15V). – ENERGY!

Type & Part Number	SLC [37]	MLC$_{\times 2}$ [38]
Endurance (Erase Cycles)	100,000	10,000
Price (US dollars/GB, 2008 Q1)[42]	7.10	2.48
Serial Access	$25ns$	$25ns$
Random Read	$20\mu s$	$60\mu s$
Write/Program	$200\mu s$	$800\mu s$
Erase	$1.5ms$	$1.5ms$

Flash SSD: Architectural Challenges

- Page/block-based architecture,
    - doesn't allow for a direct overwrite of data,
    - requiring sophisticated garbage collection
    - bulk erase procedures,

- Computation-intensive data management
    - Takes extra memory space,
    - Limits performance
    - Accelerates the wearing out of memory cells.
    - Lower power potential **compromised** in current SSD implementations

- Flash memory scaling doesn't improve (degrades!!!) the basic performance characteristics
  - read, write and erase latencies have been nearly constant for over a decade
  - Extreme scaling results in the degradation of retention time and endurance,
    - critical for storage applications!
- Emerging technologies for non-volatile memories have a potential to "take over" the scaling roadmap for flash.
  - may help to overcome the fundamental shortcomings of flash technology.
  - e.g. ReRAM

16

A THEME:
Minimal Memory Element

What is the smallest volume of matter needed for a memory cell?

Al Fazio, Intel Fellow (ITRS ERD meeting, Barza, April 2010)

Information carriers

Flash

Electrons — $V_{min} \sim 10^3$ nm^3

Atoms — $V_{min} < 10$ nm^3

ReRAM
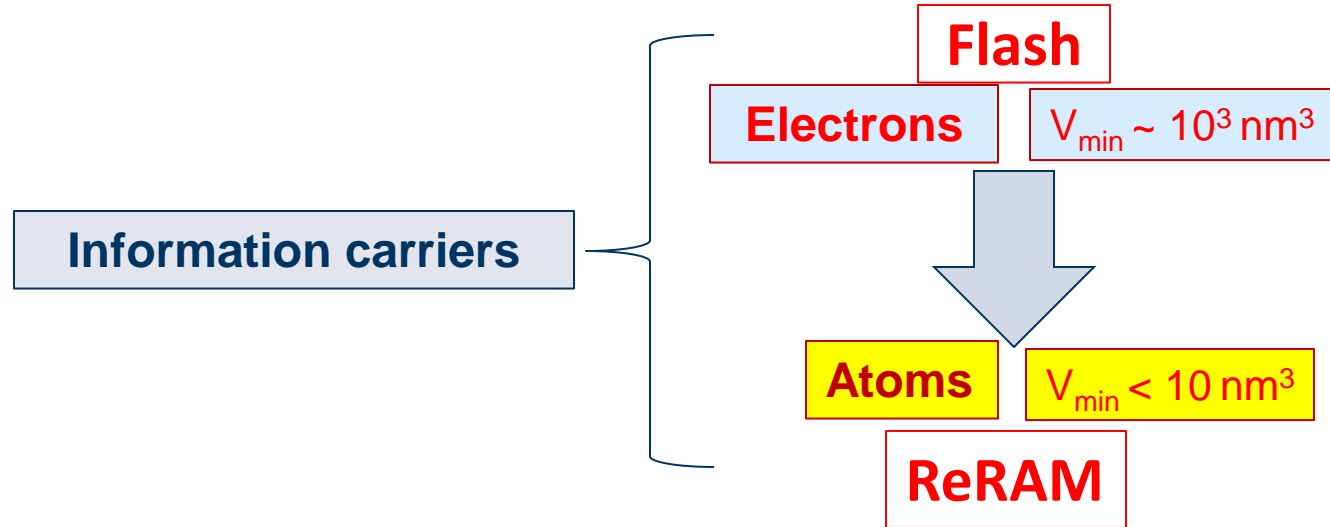
V. V. Zhirnov, R. K. Cavin, S. Menzel, E. Linn, S. Schmelzer, D. Bräuhaus, C. Schindler and R. Waser, "**Memory Devices: Energy-Space-Time Trade-offs**", *Proc. IEEE* 98 (Dec. 2010) 2185

In collaboration with RWTH Aachen Univ / Jülich Res. Ctr.

V. V. Zhirnov, R. Meade, R. K. Cavin, S. Menzel, and G. Sandhu, "**Scaling Limits of Resistive Memories**", *Nanotechnology* 22 (June 2011) 254027

In collaboration with Micron Technology, Inc.

Space-Action Principle for Memory

$$Energy \times time \times Volume = \min$$

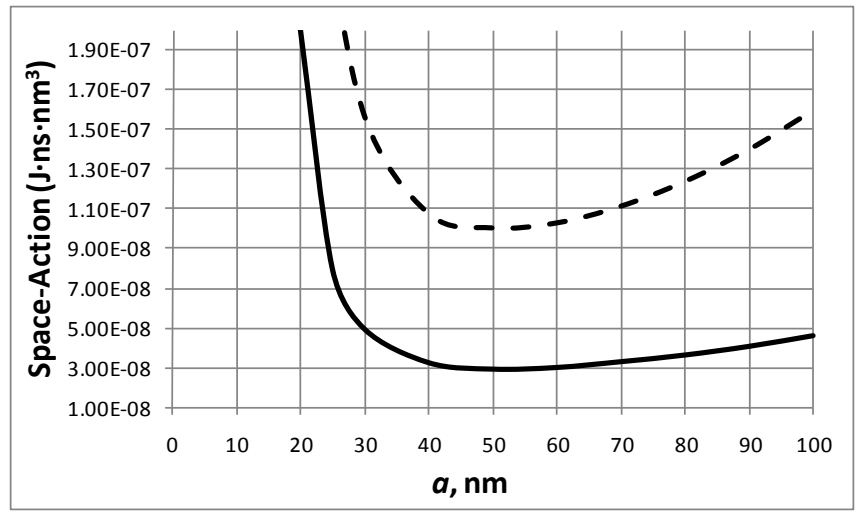$$E \cdot t \cdot V = \min$$ $$E \cdot t \cdot L = \min$$ $$E \cdot t \cdot N_{at} = \min$$

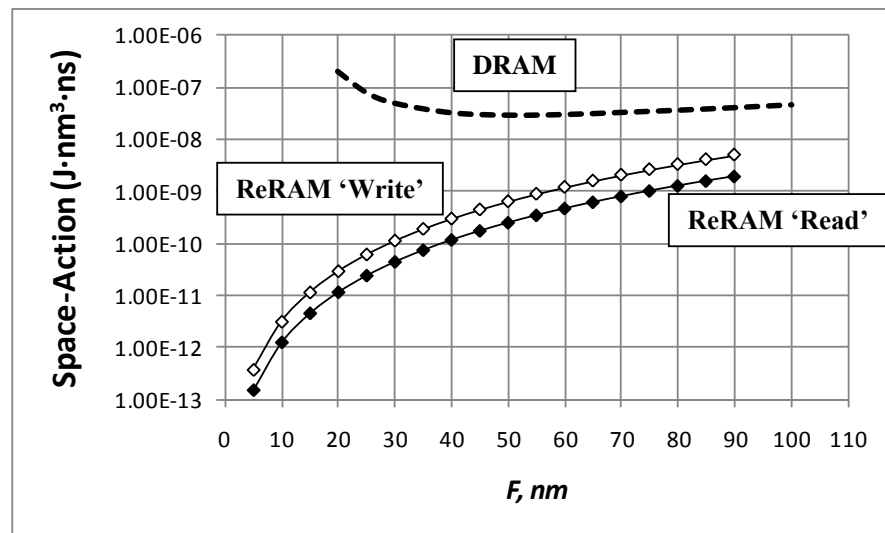**The Least Action principle is a fundamental principle in Physics** $E \cdot t = \min$ **(≥h)**

Plank's constant
$h=6.62 \times 10^{-34}$ Js

Scaling optimization for DRAM based on minimal space action

DRAM vs. ReRAM

# Memory Devices:
Space-Time-Energy Trade-offs

	$N_{carriers}$	V_{stor}, nm³	E_w, J	t_w, ns	*Space-Action, J-ns-nm³*	Critical Component	
DRAM	10^5	10^5	10^{-14}	1 ns	$\sim 10^{-7}$-10^{-8}	Storage Node	
Flash	10	10^3	10^{-16}	10^3 ns	$\sim 10^{-9}$	Sensor	FET
STT-RAM	10^5	10^3	10^{-14}	1 ns	$\sim 10^{-10}$	Selector	FET
ReRAM	100	3	10^{-17}	1 ns	$\sim 10^{-13}$	Selector	FET or 2-t select device

Constraints by remote sensor not considered

$$Energy \times time \times Volume = \min$$

Architectural Implications

- Advances in memory technologies could drive the emerging data-centric chip architectures

  - *Nanostores* - Chips consisting of multiple 3D-stacked layers of dense nonvolatile memory with a top layer of power-efficient processor cores

- *Nanostores* architectures could be an important direction for the future of information processing.

- **Ultra-fast data access**
- **Flattening memory hierarchy**
- **LOW ENERGY!**

Matches with future data-centric workloads

COVER FEATURE

From Micro-processors to Nanostores: Rethinking Data-Centric Systems
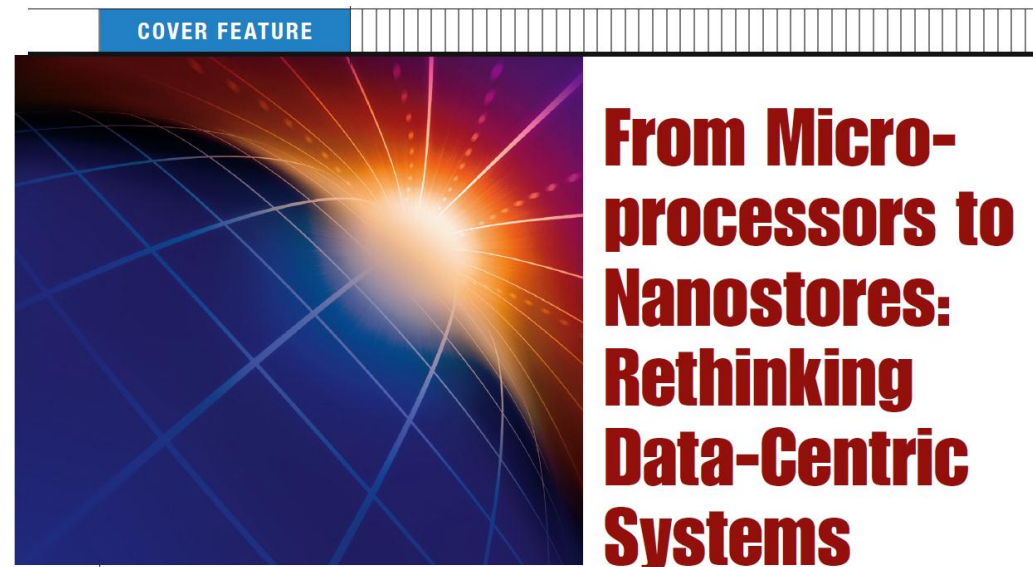
Parthasarathy Ranganathan, *HP Labs*

***Computer*, Jan. 2011**

Disruptive Sensor Technologies: *The Future*

Trend: Sensors will likely be everywhere - from scales ranging from the human cell to outer space; and they will be present in very large numbers.

A Vision: Integrated Sensor Systems

- Highly functional space-limited (e.g. portable) sensor systems
  - Multisensory
  - Operate with extremely low energy consumption.

- Sensor networks could consist of a very large number of sensor nodes talking to one another; sleeping and waking on demand etc.
  - Sensors to support ubiquitous / pervasive autonomic networks
  - Communication energy/volume expenditures is most costly activity – should therefore maximize "node intelligence"

- 'Autonomic'
  - It must be user-transparent, self-healing, self-configuring, self-optimizing, and self-protecting

- Energy sources are key to future integrated sensor systems
 - Diverse energy sources in a small volume

Universal Sensor Platform

- Can we move from the custom-designed, application-specific sensors to a technology framework for universal integrated on-chip multisensor systems?

- Could be next successful technology platform
  - Standard cell library for sensors
  - Standard interfaces

Gedanken '*In Silico*' System

- **Nanomorphic Cell:** A model system, designed to analyze the physical scaling limits of electronic systems,

- Postulated to be confined within a $10\mu m \times 10\mu m \times 10\mu m$ cube.

- An atomic-level integrated, self-sustaining microsystem with **six** primary components: computation, communication, energy supply, sensing, and actuation.

-

Communication
S1
Energy
S2
Control
S3
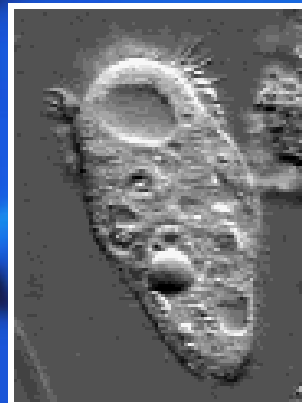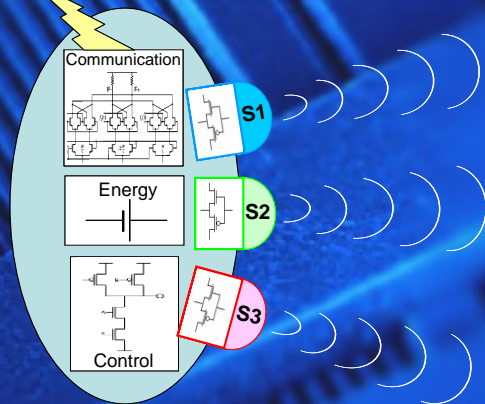
Benchmark: Living cell
In carbo system

**"Microsystems for Bioelectronics: The Nanomorphic Cell",
by Victor V. Zhirnov and Ralph K. Cavin (*Elsevier*, 2010)**

Questions and Answers

- Who will be the leaders or winners
- **Answers:**
- Who changes faster…
- Companies that develop and integrate the folowing technologies:
  - Packaging
    - 3D ICs
    - Heat management
    - Robust technologies for diverse / extreme environments
  - Universal sensor platform
  - Extremely scale energy sources