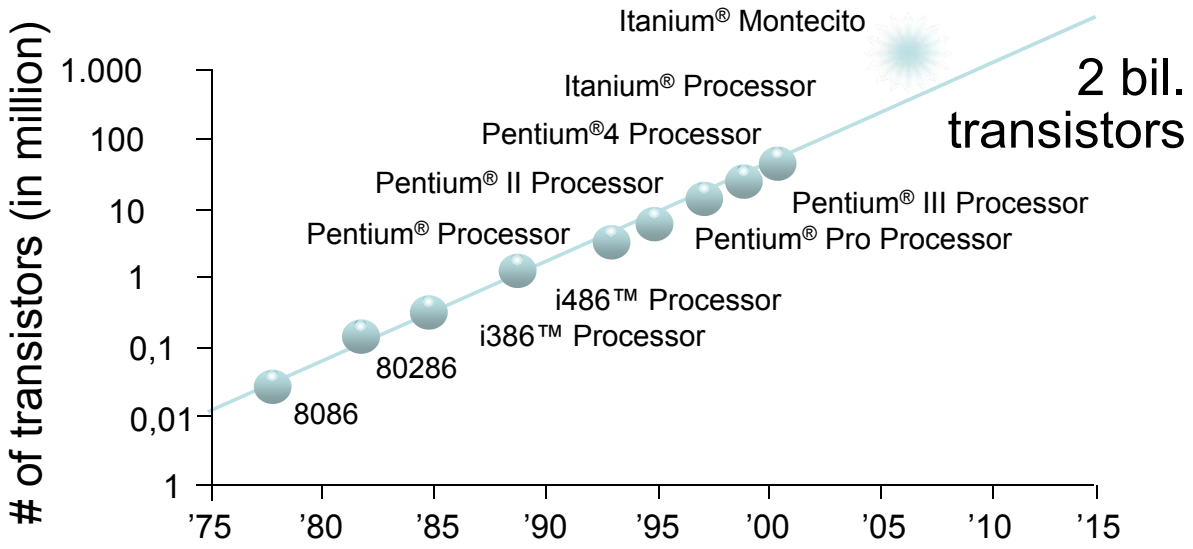
Embedded Systems

Limits of Charge-Based Devices

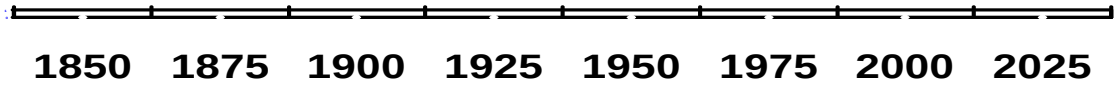
Moore's Law



Source: Intel

docendo discimus
Advanced Microsystems Technologies for Sensor Applications

Evolution of Electronics



Mechanical

Electro-Mechanical

Electronic-VT

Bipolar

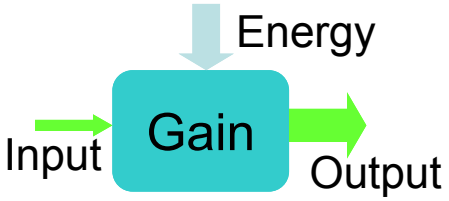
NMOS

CMOS…….⇨ ?

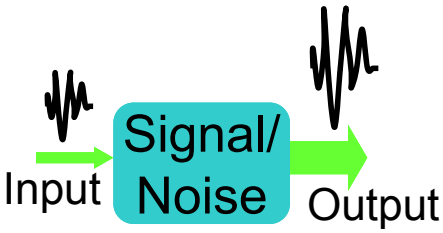
All cross-road technologies show
1. Gain
2. Signal/Noise
3. Scalability

Performance
Energy
Price/Performance

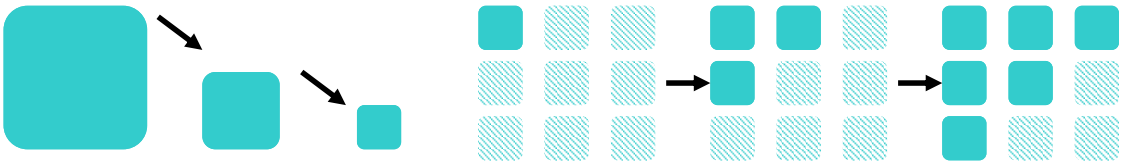
The Three Tenets



(1)



(2)



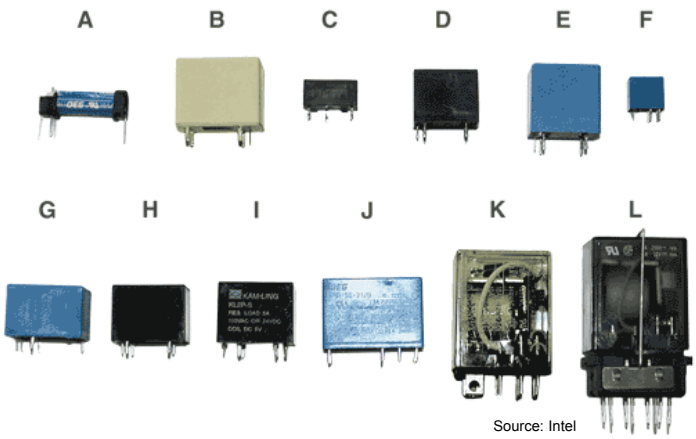
Scalability, in some shape or form

(3)

Source: Intel

Electro-Mechanical scaling—Relays

1928, Otis Elevator



Source: Intel

Vacuum Tubes

1920's

ST, UX base Globe, UV base Globe, UX base

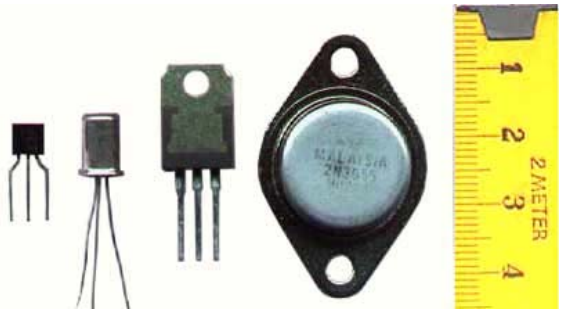
1930's

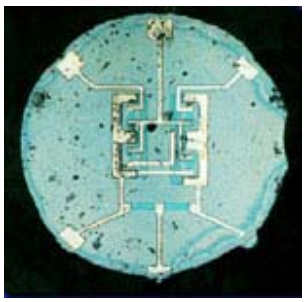
1950's & 60's

Source: Intel

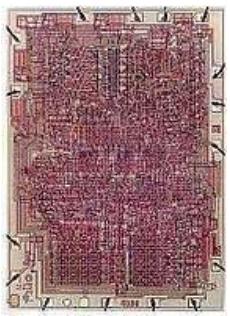
Semiconductors

The first transistor

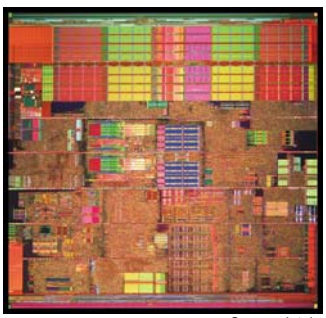




The first integrated circuit



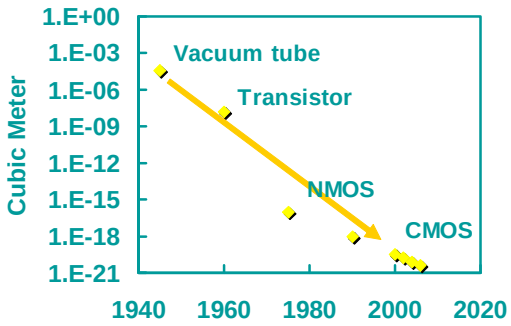
4004

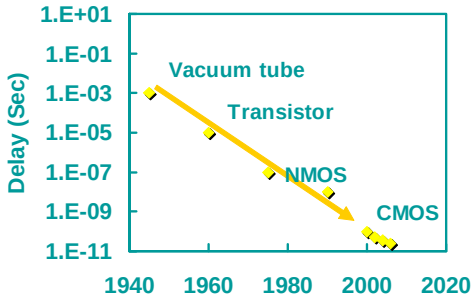


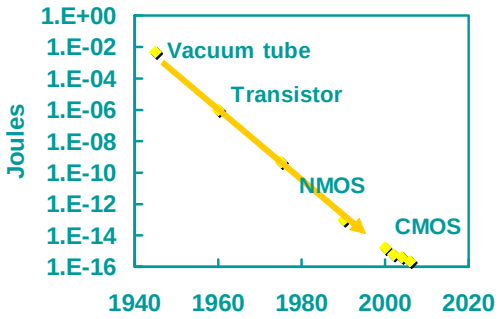
Source: Intel

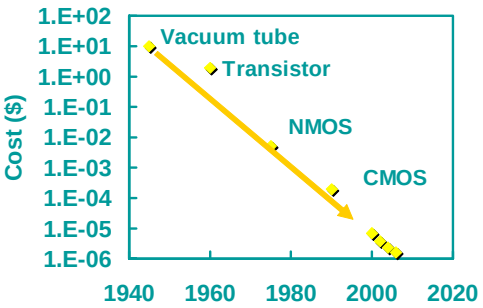
Pentium® 4

Benefits of Scaling

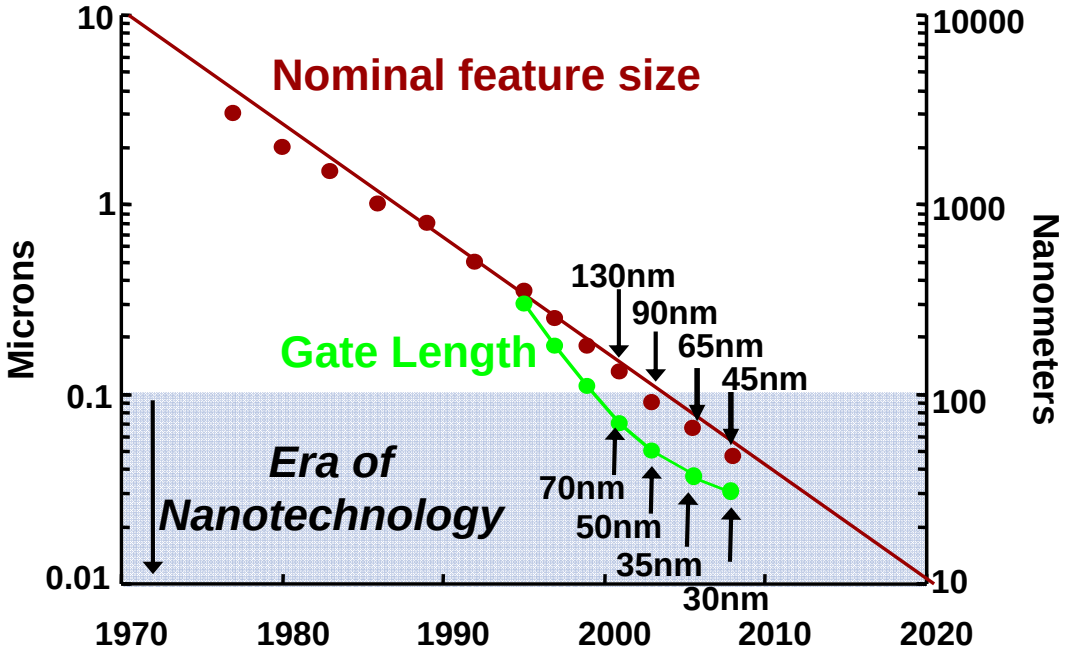








CMOS Scaling



Source: Intel

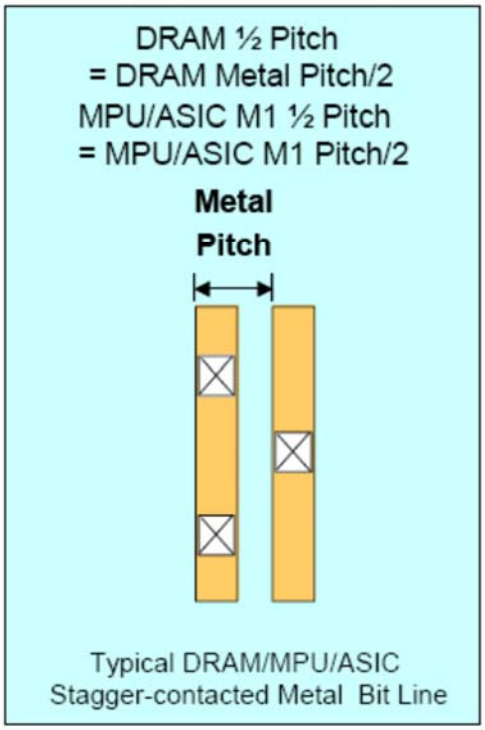
MOS Transistor Scaling
(1974 to present)

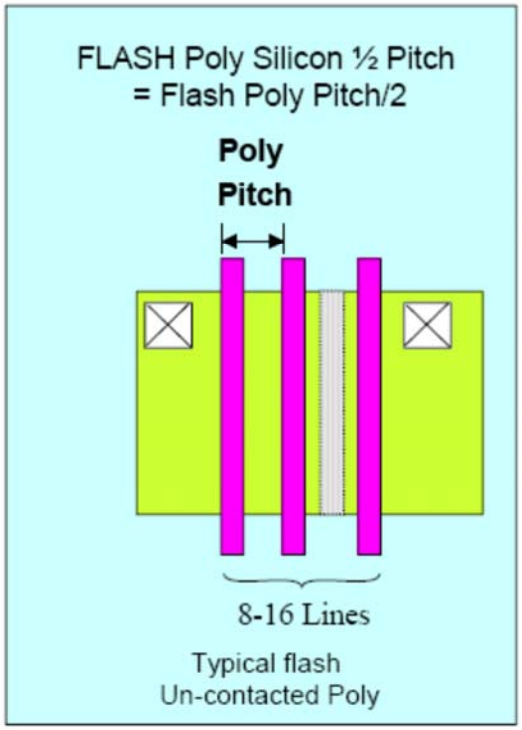
S=0.7

[0.5x per 2 Technology Cycles]

Source: ITRS

Metal and Poly Pitch





Source: ITRS

Scaling Calculator +

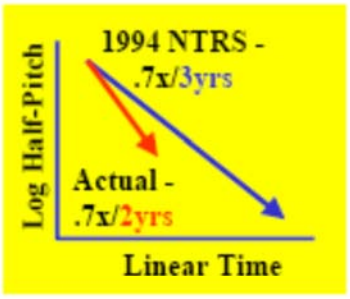
Cycle Time:

0.7x 0.7x

250 -> 180 -> 130 -> 90 -> 65 -> 45 -> 32 -> 22 -> 16

0.5x

N-1 N N+1



* CARR(T) = Compound Annual Reduction Rate (@ cycle time period, T)

Cycle Time (T yrs):

*CARR(T) =

[(0.5)^(1/2T yrs)] - 1

CARR(3 yrs) = -10.9%

CARR(2 yrs) = -15.9%

(DRAM M1 Example)

Source: ITRS

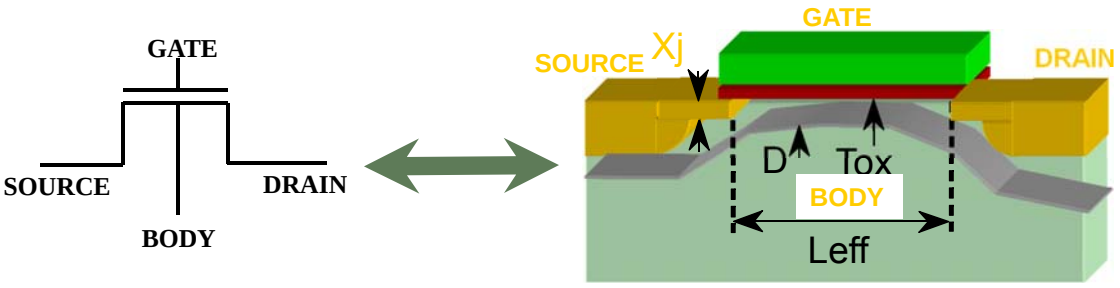
docendo discimus
Advanced Microsystems Technologies for Sensor Applications

Technology Outlook

High Volume Manufacturing	2004	2006	2008	2010	2012	2014	2016	2018
Technology Node (nm)	90	65	45	32	22	16	11	8
Integration Capacity (BT)	2	4	8	16	32	64	128	256
Delay = CV/I scaling	0.7	~0.7	>0.7	Delay scaling will slow down				
Energy/Logic Op scaling	>0.35	>0.5	>0.5	Energy scaling will slow down				
Bulk Planar CMOS	High Probability					Low Probability		
Alternate, 3G etc	Low Probability					High Probability		
Variability	Medium		High			Very High		
ILD (K)	~3	<3	Reduce slowly towards 2-2.5					
RC Delay	1	1	1	1	1	1	1	1
Metal Layers	6-7	7-8	8-9	0.5 to 1 layer per generation				

Source: Intel

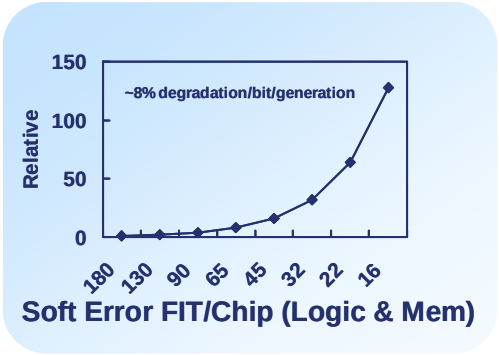
Technology Scaling

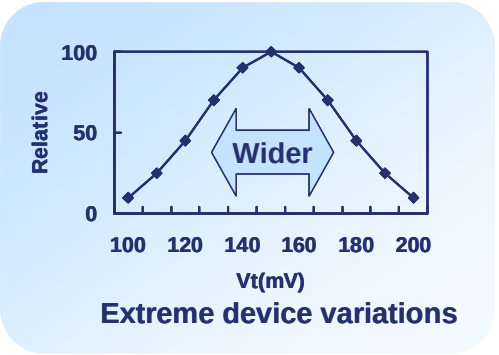


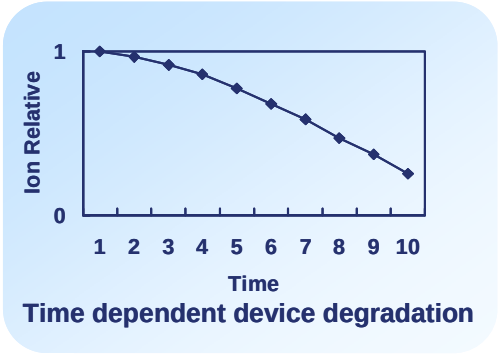
Dimensions scale down by 30%	Doubles transistor density
Oxide thickness scales down	Faster transistor, higher performance
Vdd & Vt scaling	Lower active power

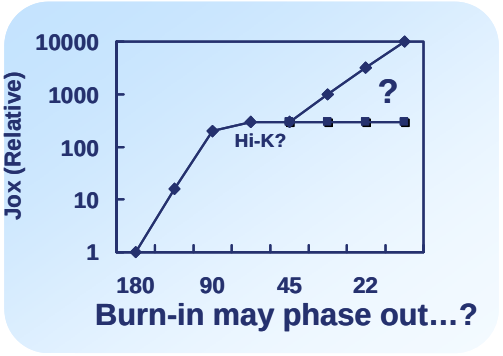
Source: Intel

Reliability



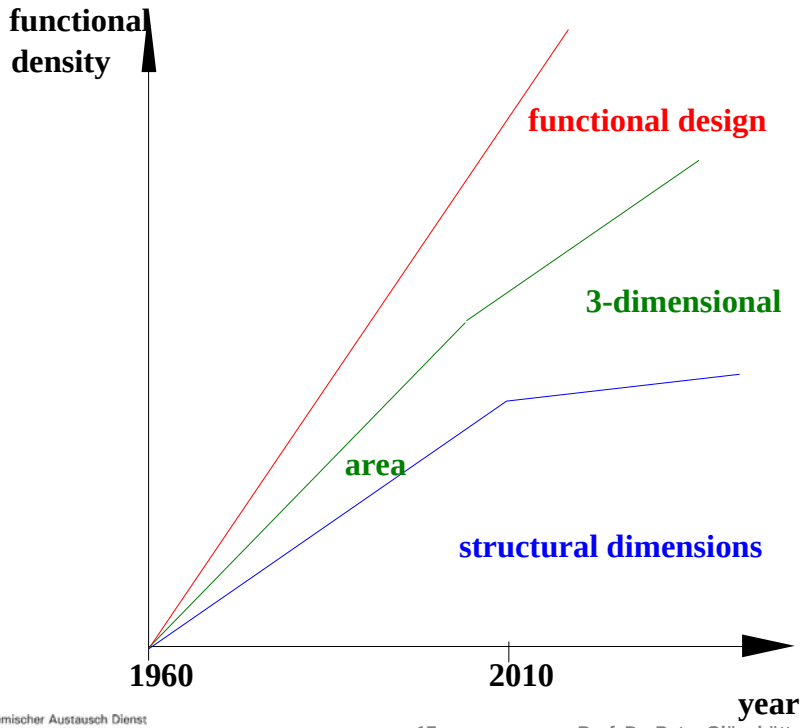




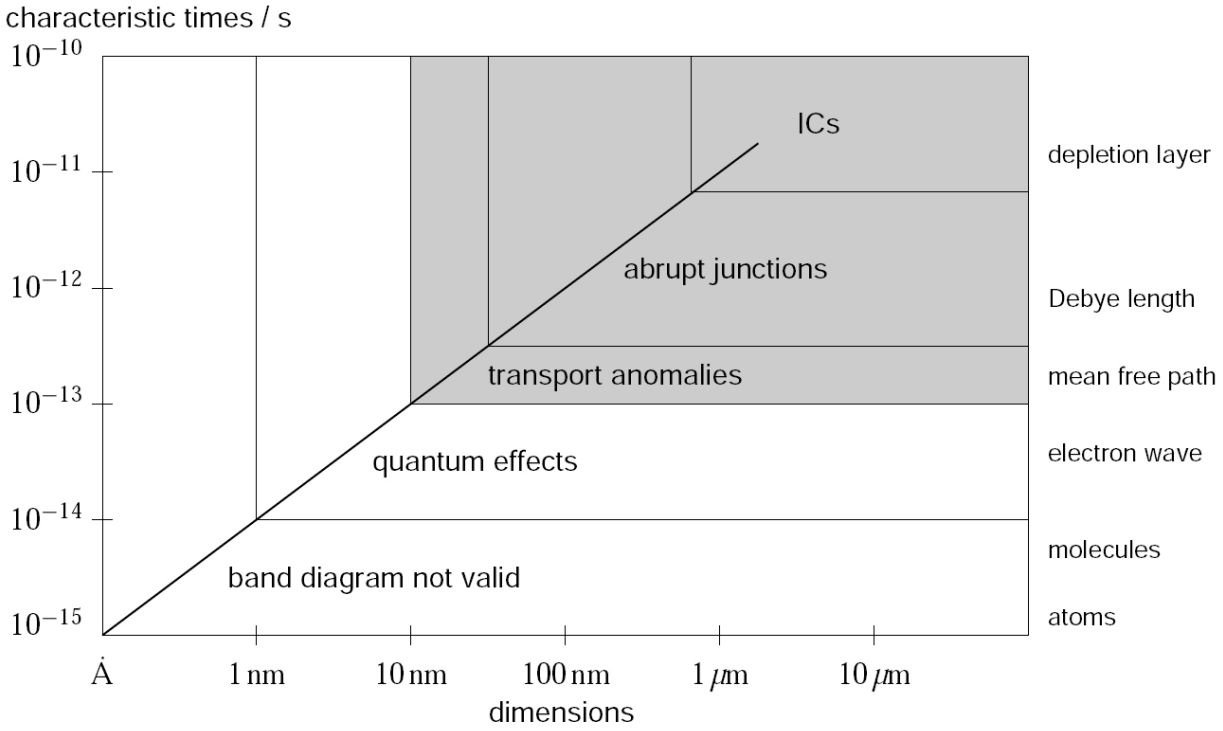


Source: Intel

Increase of functional density over time



Characteristic times and structures



CMOS—Cross Road?

Cross Road False Alarms		
1 μ	Short Channel Effects	Device Engineering
0.5 μ	Interconnects	More metals, Cu Low K ILD
130 nm	SD Leakage	Leakage control, avoidance, tolerance
65 nm	Gate Leakage	Hi-K + Metal Gate
22 nm	Lithography	EUV, Self assembly
…	…	…
< 1.5nm	SD Tunneling	?

What's in sight after CMOS?

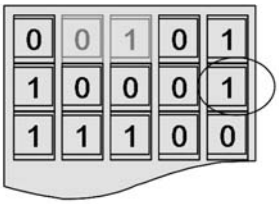
- Which technology shows gain?
- Satisfactory signal to noise ratio?
  - At room temperature?
- Scalability in some shape or form?
  - Performance, Energy, Cost
- Research must continue to find one
- Then it will take 10-15 years to mature
- Until then…

Basics of Nanoelectronics

Important Energy Limits

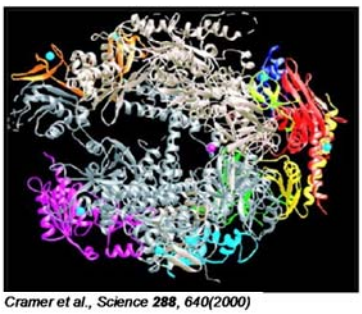
- Near-term leakage-based limit for MOSFETs:
  - May be ~5 aJ, roughly 10× lower than today.
    - 10× faster machines, ~4-8 years left on the clock
- Reliability-based limit on bit energies:
  - Roughly 100 $kT \approx 400$ zJ, ~100× below now.
    - 100× faster machines, ~8-15 years to go…
- Landauer limit on dissipation per bit erasure:
  - About 0.7 $kT \approx 3$ zJ, ~10,000× below today.
    - 10,000× faster machines, ~15-30 years left…
- No limit is known for reversible computing…
  - We need to investigate this alternative further.

Minimal Energy in Logic



*Landauer, IBM J Res Dev, **5**, 183 (1961)*
*Bennett, Int. J. Theor. Phys., **21**, 905 (1982)*

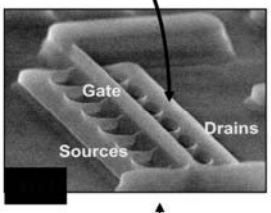
SNL theory ~ kT ln 2 ~ 17 meV
Practical ~ 40 kT ~ 1 eV

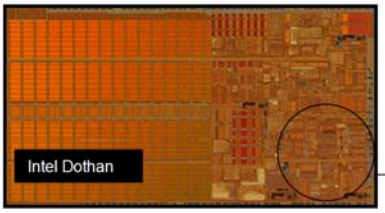


*Cramer et al., Science **288**, 640(2000)*

~ O (10 kT)
per nucleotide[1]

Electronic irreversible computing produces Joule heat

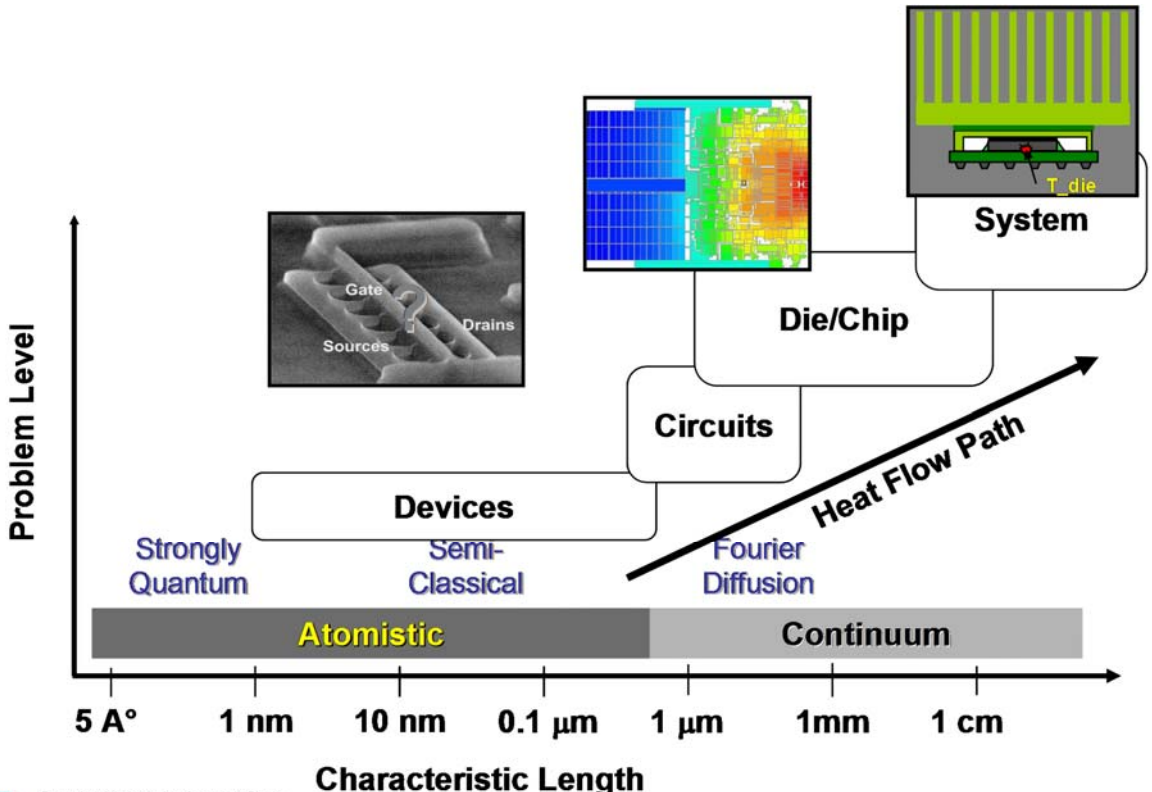




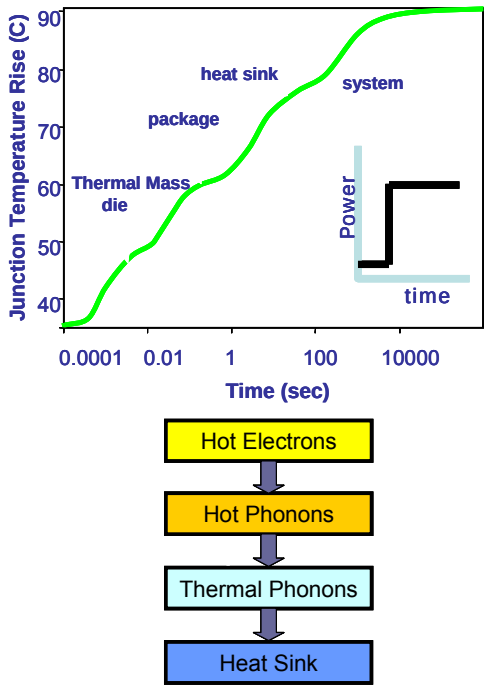
10^6 kT ~ 10 keV

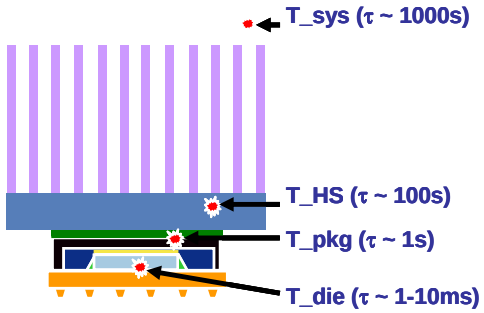
Source: Intel

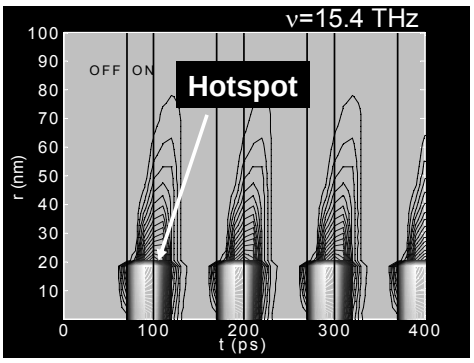
Length Scales in Internal Energy Flow



Time Scales in Internal Energy Flow







Sinha et al, J. Heat Transfer, 128 (2006)

Source: Intel

docendo discimus
Advanced Microsystems Technologies for Sensor Applications

FET Energy Limit

- A practical limit for all transistors based on the *field effect* principle.
  - It's probably not an absolutely unavoidable, fundamental limit.
    - However, it is probably the biggest barrier to further transistor scaling today.
- Eventually, the extra power dissipation from leakage overwhelms the power/performance reductions that we would gain by reducing CV^2!
  - Beyond this point, further transistor scaling hurts us, rather than helping.
    - Transistor scaling then halts, for all practical purposes!

Reliability-Based Limit

- A limit on signal (bit) energy.
- Applies to any mechanism for *storing* a bit whose operation is based on the *latching* principle, namely:
  - We have some physical entity whose state (*e.g.* its location) encodes a bit.
    - *E.g.*, could be a packet of electrons, or a mechanical rod
  - If the bit is 1, the entity gets "pushed into" a particular state and held there by a potential energy difference (between there and not-there) of E.
    - The entity sits in there at thermal equilibrium with its environment.
  - A potential energy barrier is then raised in between the states, to "latch" the entity into place (if present).
    - A transistor is turned off, or a mechanical latching mechanism is locked down


docendo discimus
Advanced Microsystems Technologies for Sensor Applications

Reliability-Based Limit

- The Boltzmann distribution implies that $E > T \log N = kT \ln N$, in order for the probability of incorrect storage to be less than $1/N$.
  - For electrons, we must use the Fermi-Dirac distribution instead…
    - But this gives virtually identical results for large N.
- When erasing a stored bit, typically we would dissipate the energy E.
  - However, this limit might be avoidable via special level-matching, quasi-adiabatic erasure mechanisms, *or* non-equilibrium bit storage mechanisms.

Numerical Example

- Reliability factor of $N=10^{27}$ (e.g., 1 error in a 10^9 gate processor running for ~3 years at 10 GHz):
- The entropy associated with the per-op error probability is then:
  $\log 10^{27} = 27 \log 10 = 27\ k_B \ln 10 \approx 62\ k_B = 8.6\times10^{-22}$ J/K
- Heat that must be output to a room-T (300 K) environment: k_B (300 K) $\ln 10^{27} = 2.6\times10^{-19}$ J (or 260 zJ, or 1.6 eV)
- Sounds small, but…

Von Neumann / Landauer (VNL) bound for bit erasure

- The von Neumann-Landauer (VNL) lower bound for energy dissipation from bit erasure:
  - "Oblivious" erasure/overwriting of a known logical bit moves the information that it previously contained to the environment → The information becomes entropy.
    - Leads to fundamental limit of $kT \ln 2$ for oblivious erasure.
  - This particular limit could *only* possibly be avoidable through *reversible computing*.
    - Reversible computing "de-computes" unwanted bits, rather than obliviously erasing them!
      - This enables the signal energy to be preserved for later re-use, rather than dissipated.

Adiabatic Circuits

- Reversible logic can be implemented today using fairly ordinary voltage-coded CMOS VLSI circuits.
  - With a few changes to the logic-gate/circuit architecture.
- We avoid dissipating most of the circuit node energy when switching, by transferring charges in a nearly *adiabatic* (literally, "without flow of heat") fashion.
  - *I.e.*, asymptotically thermodynamically reversible.
    - In the limit, as various low-level technology parameters are scaled.
- There are many designs for purported "adiabatic" circuits in the literature, but most of them contain fatal design flaws and are not truly adiabatic.
  - Many past designers are unaware of (or accidentally failed to meet) all the requirements for true thermodynamic reversibility.

Physical Fundamentals

$$rotH = J + \frac{\partial D}{\partial t} \qquad rotE = -\frac{\partial B}{\partial t}$$

Description of macroscopic electrical effects → basics of classical electrical engineering

Quantization of charge Q:

→ Coulomb blockade

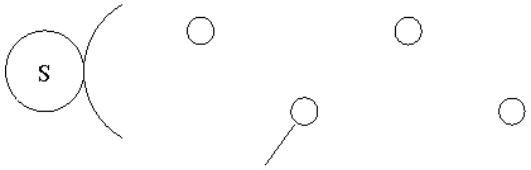
$$W = \frac{1}{2}\frac{(nq)^2}{C}$$

Physical Fundamentals

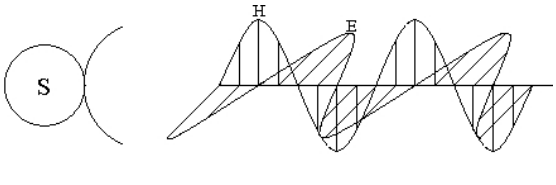
Differential equation:
(derived from Maxwell's law)

$$\frac{\partial^2 E}{\partial^2 x} = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}$$

a) Particle Flow



b) Electromagnetic Wave



$$W = hf = \frac{hc}{\lambda}$$

Quantization

Action: $H = Wt = n_H h$

Common electrical values: $W = \frac{H}{t} = VIt$

$$P = \frac{W}{t} = VI$$

Further quantized values may be derived heuristically.

Schrödinger Equation

Kinetic energy of a single electron: $E = \frac{1}{2} m_e v^2$

Classic model → electron with velocity v has a momentum of $p=m_e v$

➔ Action: $H = m_e v \cdot x$

Action should increase continuously with x, but as we know, action is quantized. → abruptly changing velocity v and discontinuous changes of the electron energy? **How is that possible?**

➔ We have to give up the classical model of a continuous trajectory and we have to move on to quantum mechanics.

Schrödinger Equation

Quantum mechanics describes electrons using successive quantum-mechanical states, which represent a certain probability that the particle may be located in a specific spatial region.

$$-\frac{\hbar^2}{2m}\frac{\partial^2 \Psi(x)}{\partial x^2} + V(x) \cdot \Psi(x) = E \cdot \Psi(x)$$ (time-independent)

For the trivial case V(x)=0 it is solved by the complex wavefunction:

$$\Psi(x) = A \cdot \exp\left(j(kx - \omega t)\right)$$

The probability P to encounter the particle in a specific spatial region is given by:

$$P = \left|\Psi\right|^2 = \Psi \cdot \Psi^*$$

Wavelength & Energy

Let H=h and x=λ, then wavelength of Ψ equals:

$$\lambda = \frac{h}{m \cdot v} \longrightarrow p = \frac{h}{\lambda}$$

De Broglie

$$v^2 = \frac{h^2}{m^2 \cdot \lambda^2}$$

➔ Kinetic energy of a particle:

$$E = \tfrac{1}{2} m v^2 = \frac{h^2}{2m} \frac{1}{\lambda^2} = \frac{\hbar^2}{2m} k^2$$

with: $k = \frac{2\pi}{\lambda}$

$\hbar = \frac{h}{2\pi}$

Uncertainty

$$p = \frac{h}{\lambda}$$

De Broglie

$$\frac{\lambda}{2} \cdot m v = \frac{h}{2}$$

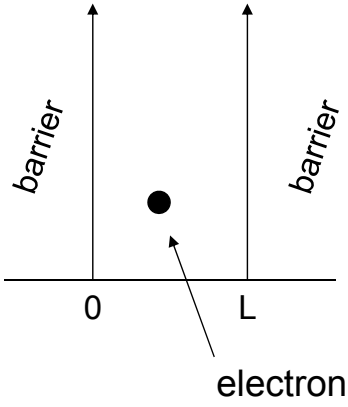
Let $\frac{\lambda}{2} = \Delta x$ and $m \cdot \Delta v = \Delta p$:

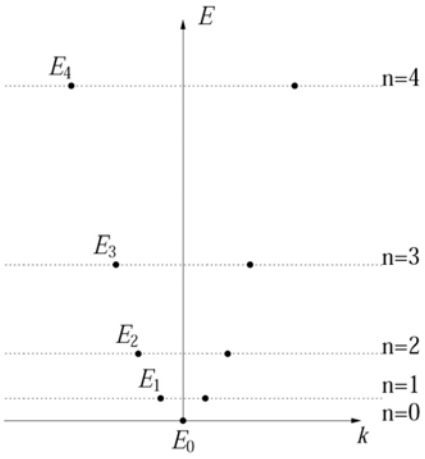
We obtain heuristically the uncertainty principle:

$$\Delta x \cdot \Delta p \geq \frac{h}{2}$$

Electrons in Potential Wells

Discrete energy levels of an electron in an one-dimensional potential well:





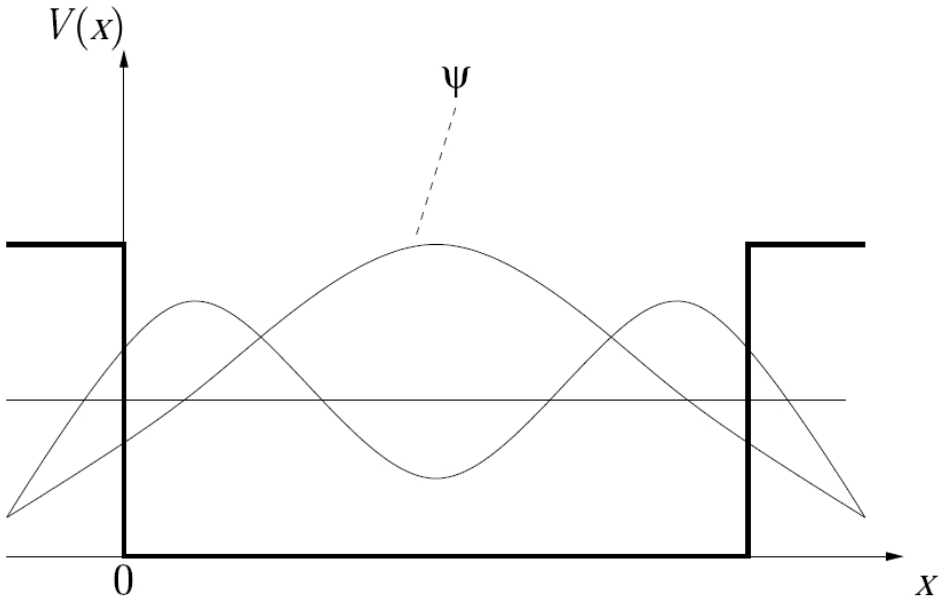
The eigenstates are located on a parabolic curve:

$$E = \frac{\hbar^2}{2m} k^2$$

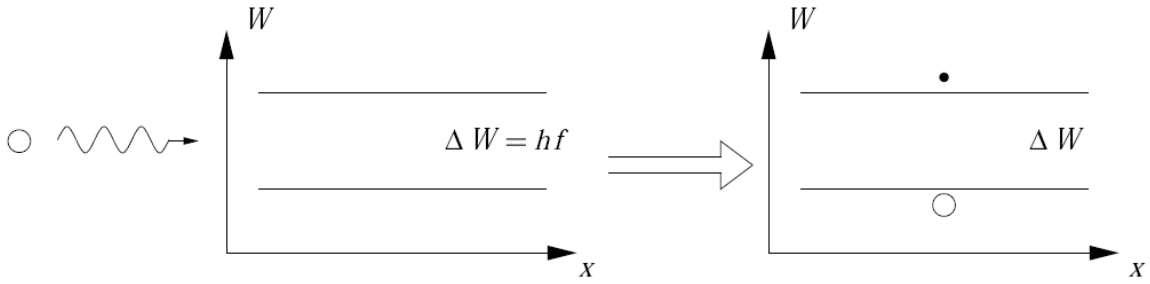
Solution of Schrödinger's equation:

$$\Psi = \frac{2}{L} \sin \frac{n\pi x}{L} \quad with \quad n = 0,1,2...$$

Potential well defined by finite walls



Photons interacting with Electrons in Solids



Action converted during the switching operation: $W_s t_s = nh$

n=1 leads to minimum amount of energy: $W_s \geq \frac{h}{t_s}$

Information Content

- Shannon defined the information content I_r of a pattern X_r that occurs with the probability of P_r:

$$I_r = -K \ln P_r$$

K is a constant factor

- Information content is measured in bits. Presuming that *0* and *1* arrive with the same probability P_r=*0.5*, the factor *K* is determined by:

$$I = -ld\ 0.5 = 1bit$$

Expectation Value

- If the pattern is chosen from a set of m patterns with an identical occurrence probability, the average information content equals *H*:

$$H = -\sum_{i=1}^{m} P_i \cdot ld\ P_i \quad i = 1...r...m$$

H is called the *expectation value* or in terms of thermodynamics *entropy*.

- Now, consider *m* different incidents with the same occurrence probability $P_r=1/m$. The average information content reaches a maximum of:

$$H_0 = -ld\ P_r = ld\ m$$

Minimum Energy for Processing one bit

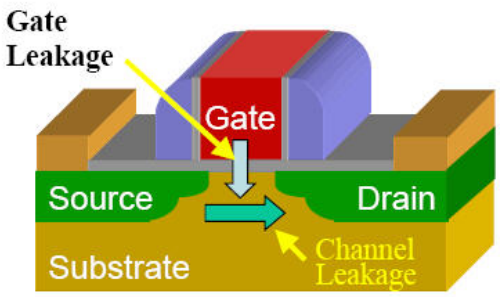
- Boltzmann showed that the entropy $S=W/T$ of an ideal gas can also be defined by the number of micromechanical states:

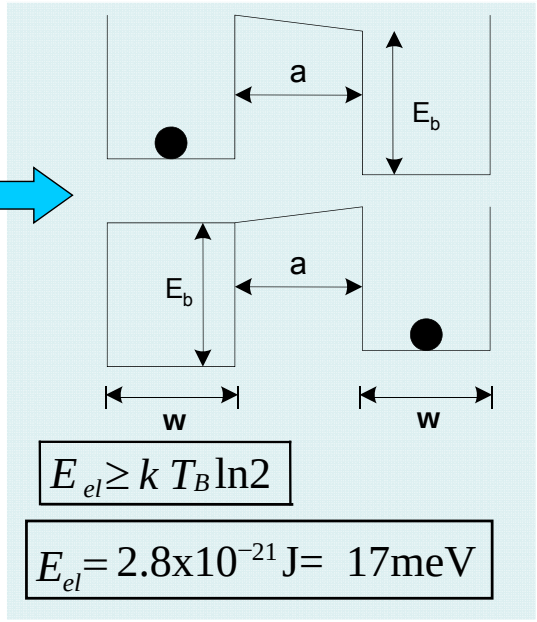
$$S = k \cdot \ln m \qquad k = 1.38 \cdot 10^{-23}\ JK^{-1}$$

➔ Minimum energy for processing one bit of information:

$$W = 2kT \ln 2$$

MOSFET is optimal electronic switch





$$E_{el} = \frac{h^2}{2ma_{el}^2} \qquad \tau_{el} = \frac{h}{E_{el}}$$

$$a_{el} = 1.5\text{nm} \qquad \tau_{el} = 40\text{fs}$$

$$E_{el} \geq k\,T_B \ln 2$$

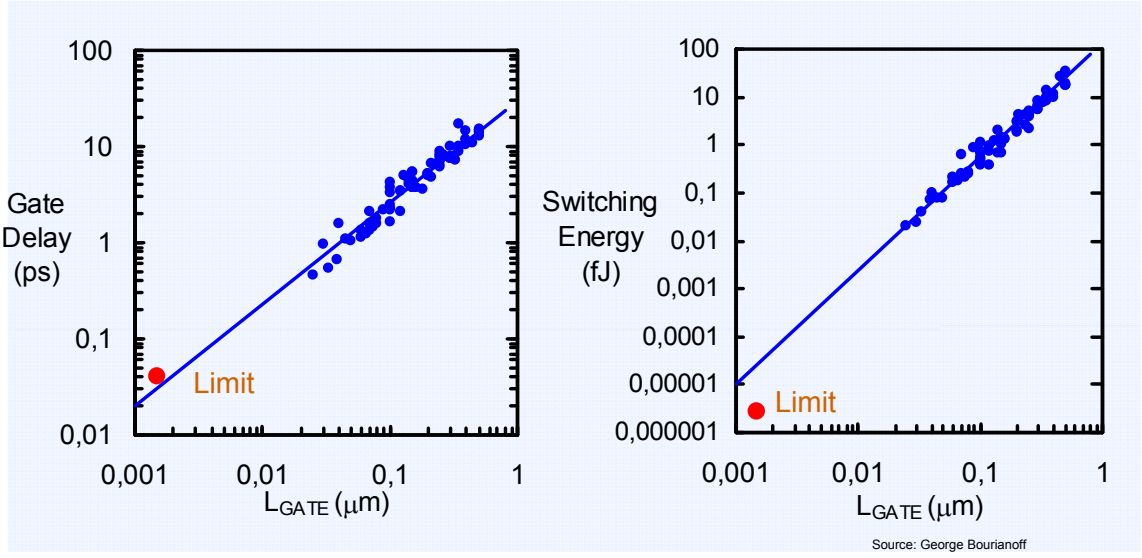
$$E_{el} = 2.8\text{x}10^{-21}\,\text{J} = 17\text{meV}$$

Source: George Bourianoff

Assumes switch at thermal equilibrium

Zhirnov et al., Proc. IEEE 91, 1934 (2003).

CMOS scaling close to optimal for electron based devices



Source: George Bourianoff

Alternative logic technologies may provide devices that provide improved power efficiency

The "next switch"
What are we looking for?

- **Required characteristics:**
  - **Energy efficiency**
  - **Scalability**
  - **Performance**
  - **Gain**
  - **Operational reliability**
  - **Room temp. operation**
- **Preferred approach:**
  - **CMOS process compatibility**
  - **CMOS architectural compatibility**

Alternative state variables (Beyond Charge State)

- **Spin state**
- **Solid state nanodomains (crystalographic, magnetic, dielectric, strain, coupled)**
- **Atomic (quantum) state**
- **Magnetic flux quanta**
- **Mechanical deformation**
- **Photons**
- **Other?**

Source: George Bourianoff

Quantum/ Molecular Electronics

What does Beyond CMOS mean?

- Evolutionary CMOS
  - continued scaling of CMOS to 32nm node-2009
  - Lithography, high k dielectrics, metal gates etc
- Revolutionary CMOS
  - 32-16nm node 2011-2019
  - monolithic integration of new technologies with CMOS –`enhanced´ CMOS. Taking it to the limit!
  - 3D architectures such as FINFET
  - 1D materials - Nanowires and carbon nanotubes
- Beyond CMOS
  - 16nm node 6nm gate length around 2020!
  - Logical state other than electron charge
  - New materials, devices, nano-architectures and system innovations
- Molecular electronics, Spintronics etc

Beyond CMOS – What do we need?

Materials

- Materials that support at least 2 stable “states”
- Passivation materials to prevent undesired state changes
- Materials that enable changes of the “state”
- Materials that enable reading the ”state” of a device
- Materials to transmit “state” information between “devices”

Function

- Integration on or with CMOS
- Scalable for several generations beyond CMOS
- High information/signal processing throughput
- Energy dissipation per operation `substantially´ less than CMOS
- Room temperature operation
- Gain!

Molecular state

Pros

- Chemical synthesis allows electronic properties of organic molecules to be tailored within a wide range
- Reproducibility between organic molecules is better than inorganic nanoclusters
- Can use self assembly
- High device densities
- Low switching energies due to small numbers of electrons

Cons

- Variability between device lots
- Contact formation requires atomic level control
- Thermal stability
- Metrology to look at atomic structure details in organic molecule embedded in contact layers

Resonant Tunnel Devices -RTD

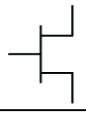
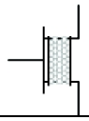
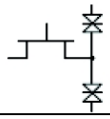
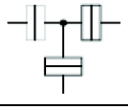
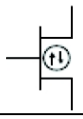
- 2 terminal devices with high switching speed and exhibit a region of negative differential resistance in their IV curves
- Connect two in series to get a device switching between 2 stable operating points. Add a third as a gate.
- Add a transistor to control peak current reduces speed by one order of magnitude.
- Ion/Ioff-Ratio is low (10), 10^5 required.

Single electron transistors -SET

- Three terminal devices which transfer electrons one-by-one
- SET channel separated from source and drain by tunnel junctions and the channel is a quantum dot
- High device density, power efficient
- Can operate at room temperature but fast operation limited to low temps due to noise immunity issues.
- Low fan out is a direct consequence of single electron operation

Ferromagnetic logic

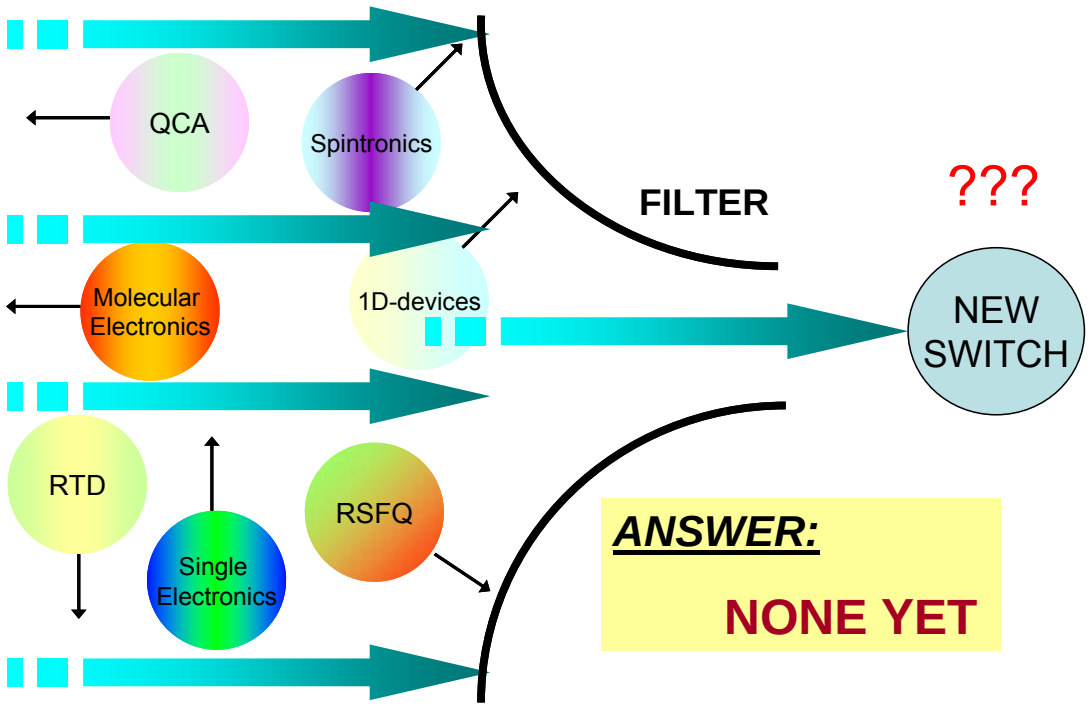
- Local ferromagnetic orientation is the state variable
- Room temperature operation
- Radiation hard
- Magnetic domain wall motion limited in switching speed

Device		 FET [B]	 1D structures	 Resonant Tunneling Devices	 SET	 Molecular	 Ferromagnetic logic	 Spin transistor
Types		Si CMOS	CNT FET NW FET NW hetero-structures Crossbar nanostructure	RTD-FET RTT	SET	Crossbar latch Molecular transistor Molecular QCA	Moving domain wall M: QCA	Spin transistor
Supported Architectures		Conventional	Conventional and Cross-bar	Conventional and CNN	CNN	Cross-bar and QCA	CNN Reconfigure logic and QCA	Conventional
Cell Size (spatial pitch)	Projected	100 nm	100 nm [C]	100 nm [C]	40 nm [L]	10 nm [Q]	140 nm [U]	100 nm [C]
	Demonstrated	590 nm	~1.5 µm [D]	3µm [H]	~700 nm [M]	~2µm [R]	250 nm [V, W]	100 µm [X]
Density (device/cm^2)	Projected	1E10	4.5E9	4.5E9	6E10	1E12	5E9	4.5E9
	Demonstrated	2.8E8	4E7	1E7	2E8	2E7	1.6E9	1E4
Switch Speed	Projected	12 THz	6.3 THz [E]	16 THz [I]	10 THz [M]	1 THz [S]	1 GHz [U]	40 GHz [Y]
	Demonstrated	1 THz	200 MHz [F]	700 GHz [J]	2 THz [N]	100 Hz [R]	30 Hz [V, W]	Not known
Circuit Speed	Projected	61 GHz	61 GHz [C]	61 GHz [C]	1 GHz [L]	1 GHz [Q]	10 MHz [U]	Not known
	Demonstrated	5.6 GHz	220 Hz [G]	10 GHz [Z]	1 MHz [F]	100 Hz [R]	30 Hz [V]	Not known
Switching Energy, J	Projected	3E-18	3E-18	>3E-18	1×10^{-18} [L] [$>1.5\times10^{-17}$] [O]	5E-17 [T]	~1E-17 [V]	3E-18
	Demonstrated	1E-16	1E-11 [G]	1E-13 [K]	8×10^{-17} [P] [$>1.3\times10^{-14}$] [O]	3E-7 [R]	6E-18 [W]	Not known
Binary Throughput, GBit/ns/cm^2	Projected	238	238 [C]	238 [C]	10	1000	5E-2	Not known
	Demonstrated	1.6	1E-8	0.1	2E-4	2E-9	5E-8	Not known
Operational Temperature		RT	RT	4.2 – 300 K	20 K [L]	RT	RT	RT
Materials System		Si	CNT, Si, Ge, III-V, In_2O_3, ZnO, TiO_2, SiC,	III-V Si-Ge	III-V Si	Organic molecules	Ferromagnetic alloys	Si, III-V, complex metals oxides
Research activity [A]			171	88	65	204	25	102

Technology Performance Evaluation

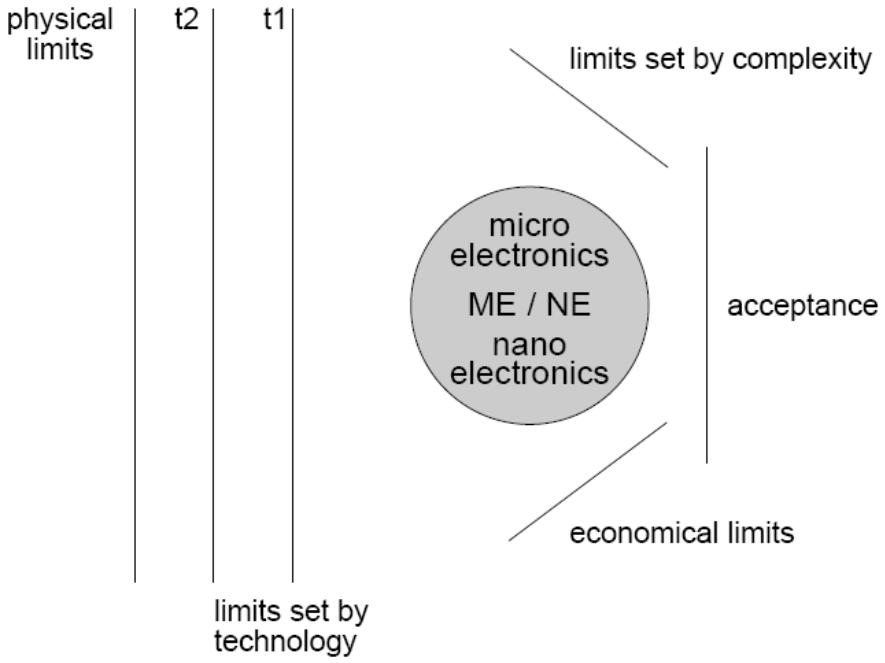
Logic Device Technologies	Scalability	Perform-ance	Energy Efficiency	Gain	Operational Reliability	Room Temp. Operation ***	CMOS Compatibility **	CMOS Architectural Compatibility *
1D Structures (CNTs and NWs)	2.4	2.5	2.3	2.2	2.1	2.8	2.5	2.7
Resonant Tunneling Devices	1.6	2.1	2.1	1.7	1.8	2.5	2.0	1.9
Single Electron Transistors	2.0	1.3	2.5	1.4	1.2	1.9	2.1	2.0
Spin Transistor	1.6	1.7	2.1	1.4	1.6	2.3	1.7	1.6
Molecular Devices	2.3	1.2	2.4	1.1	1.2	2.5	1.5	1.6
Ferromagnetic Devices	1.4	1.2	1.8	1.5	2.0	2.5	1.6	1.5

Which of Current Nanoelectronic Concepts Will Become the NEW SWITCH?

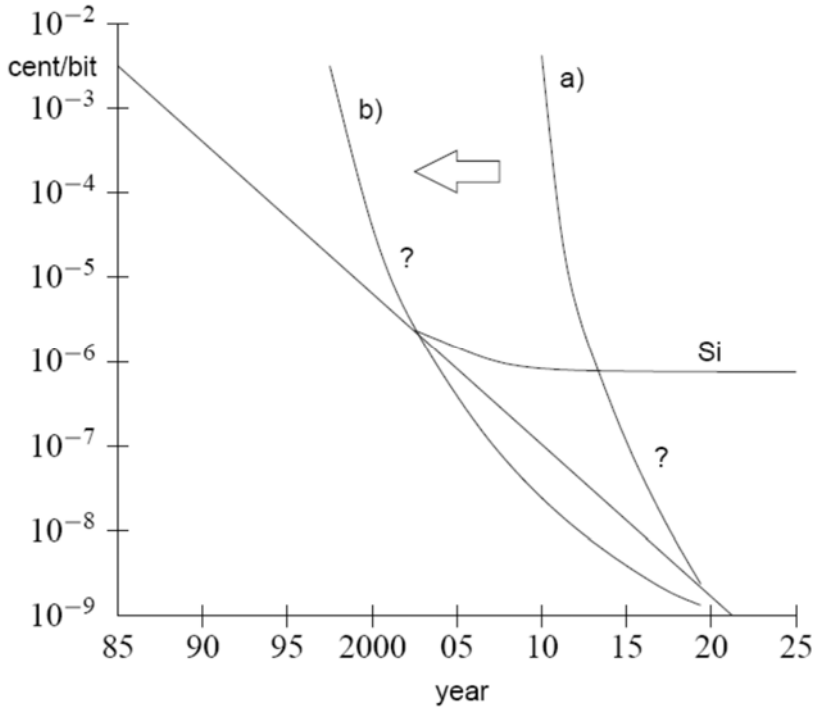


Limits of Integrated Electronics

Limits of Micro- and Nanoelectronics



Replacement of Technologies



Conclusions

- Scaling of CMOS will continue for at least another 15 years
- Consensus has emerged on promising research directions in beyond charge state computing
- Viable technology concept for beyond charge based computing is not yet visible