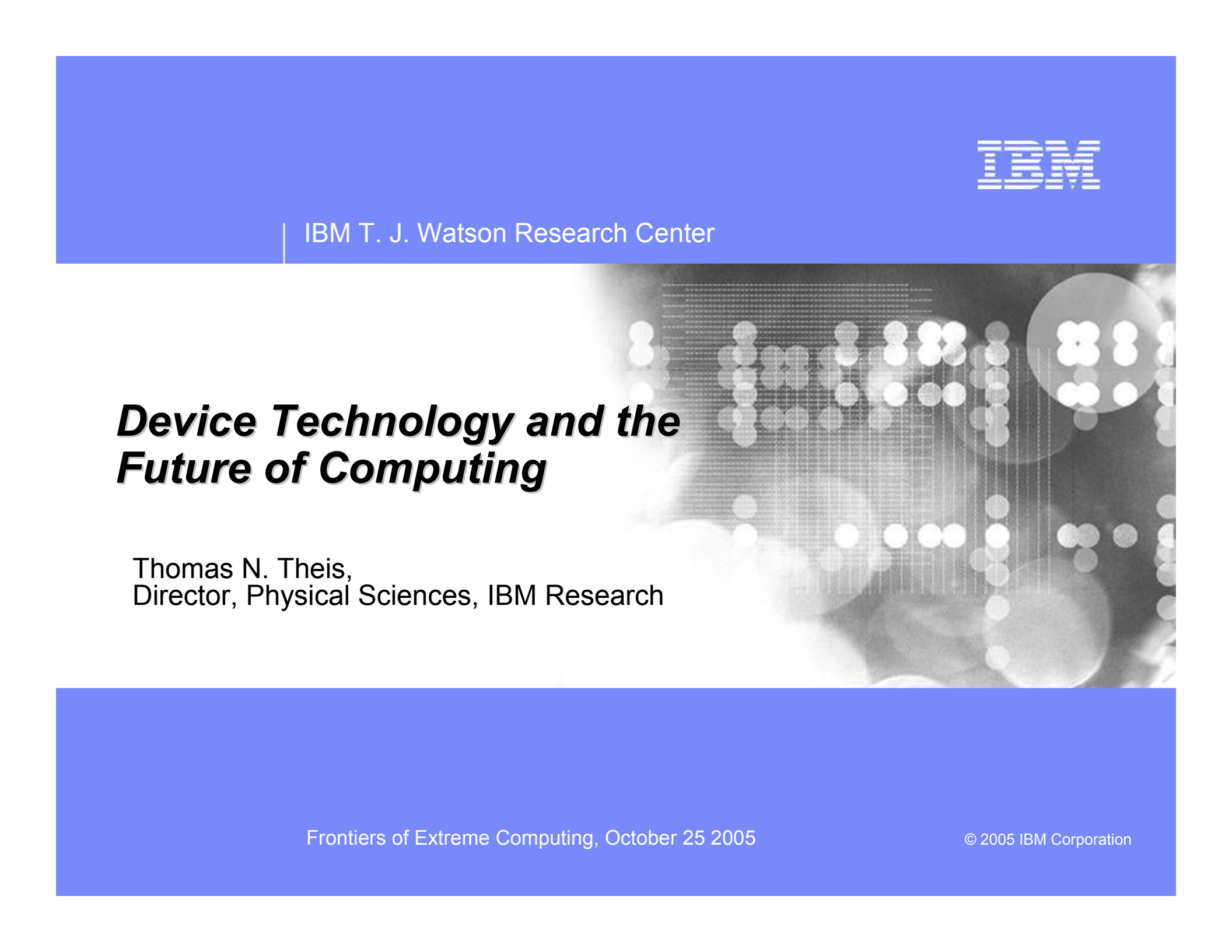


IBM T. J. Watson Research Center

Device Technology and the Future of Computing

Thomas N. Theis,
Director, Physical Sciences, IBM Research

Frontiers of Extreme Computing, October 25 2005

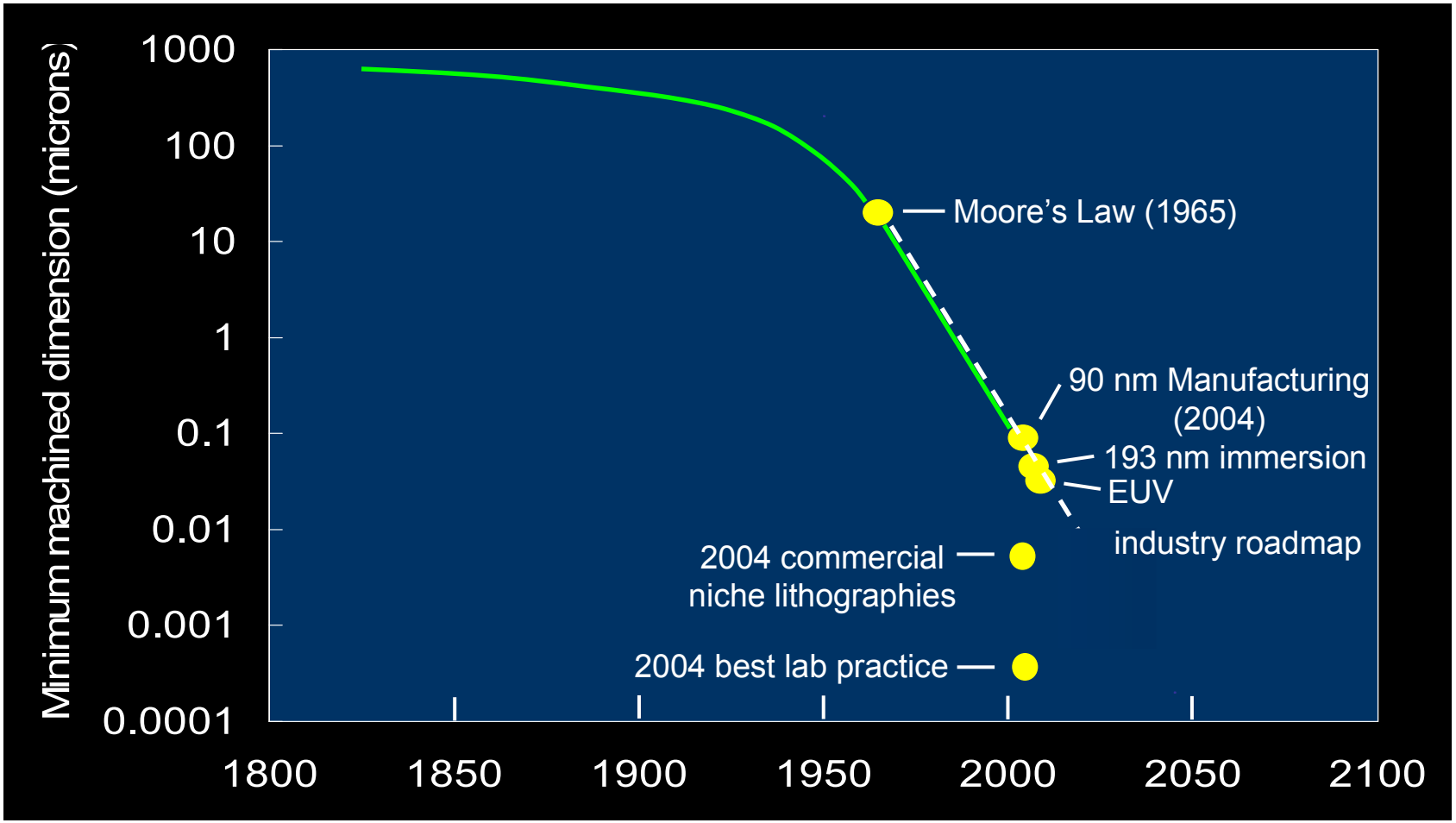
My view…

- Moore's Law is an observation about the number of devices that can be placed on a chip at a given time.
- There is no fundamental physical reason why device densities cannot follow the Moore's Law trend for many more decades.
  - Lithographic dimensions will eventually be shrunk to atomic scale and lithographic processes can seamlessly combine with synthesis.
  - New devices can circumvent the scaling limits of FETs.

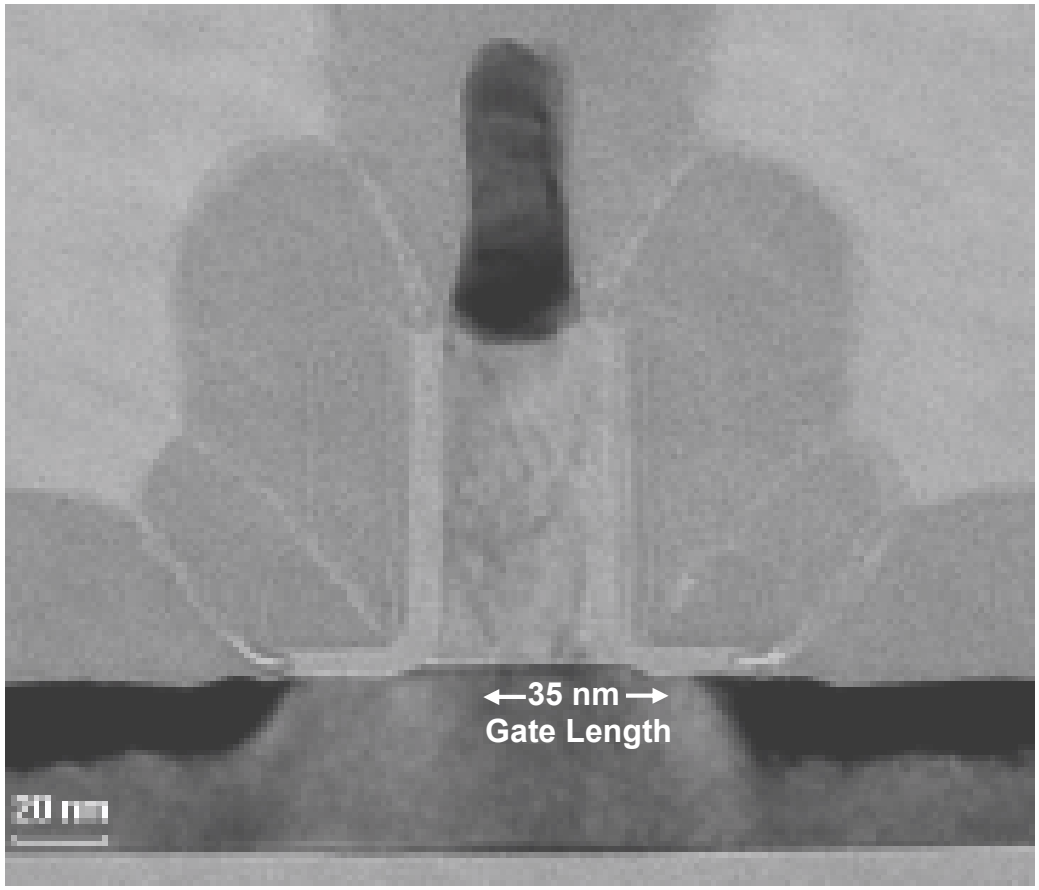
Topics

- Making devices smaller is not the problem!
- Silicon CMOS Technology will be extended for at least a decade
- The “ultimate” FET may not be made of silicon
- Non-silicon memory devices will scale to very small dimensions.
- New devices may enable adiabatic computing and reversible logic.

A Brief History of Miniaturization

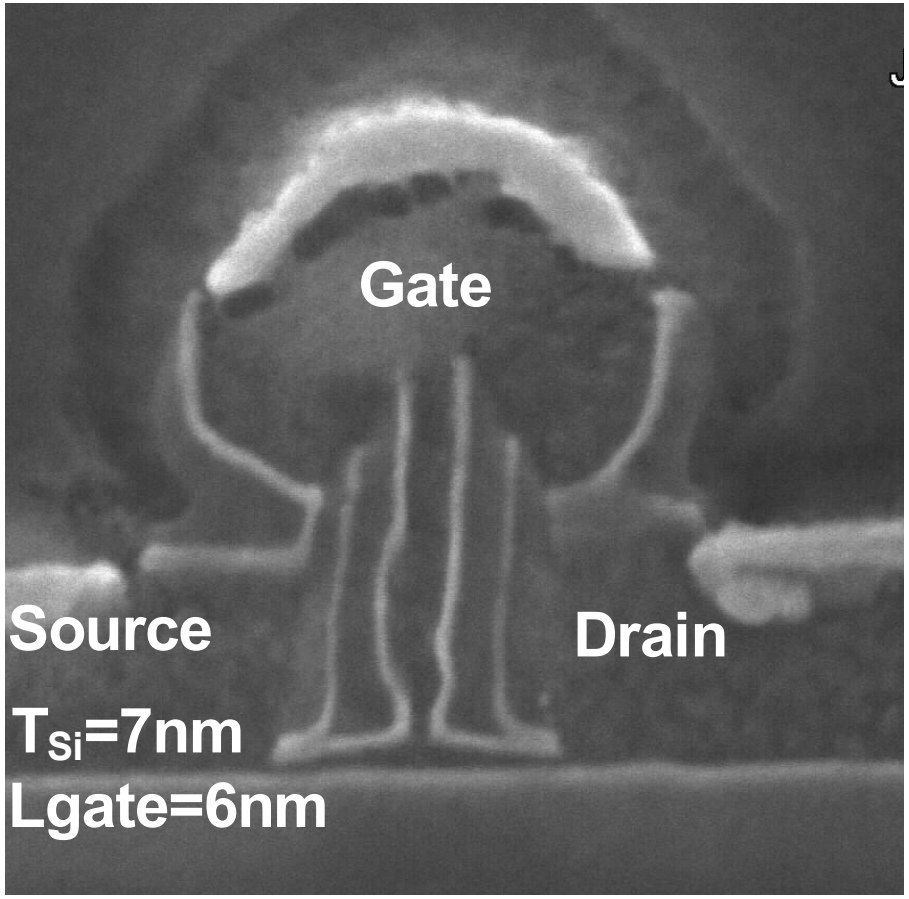


The silicon transistor in manufacturing …



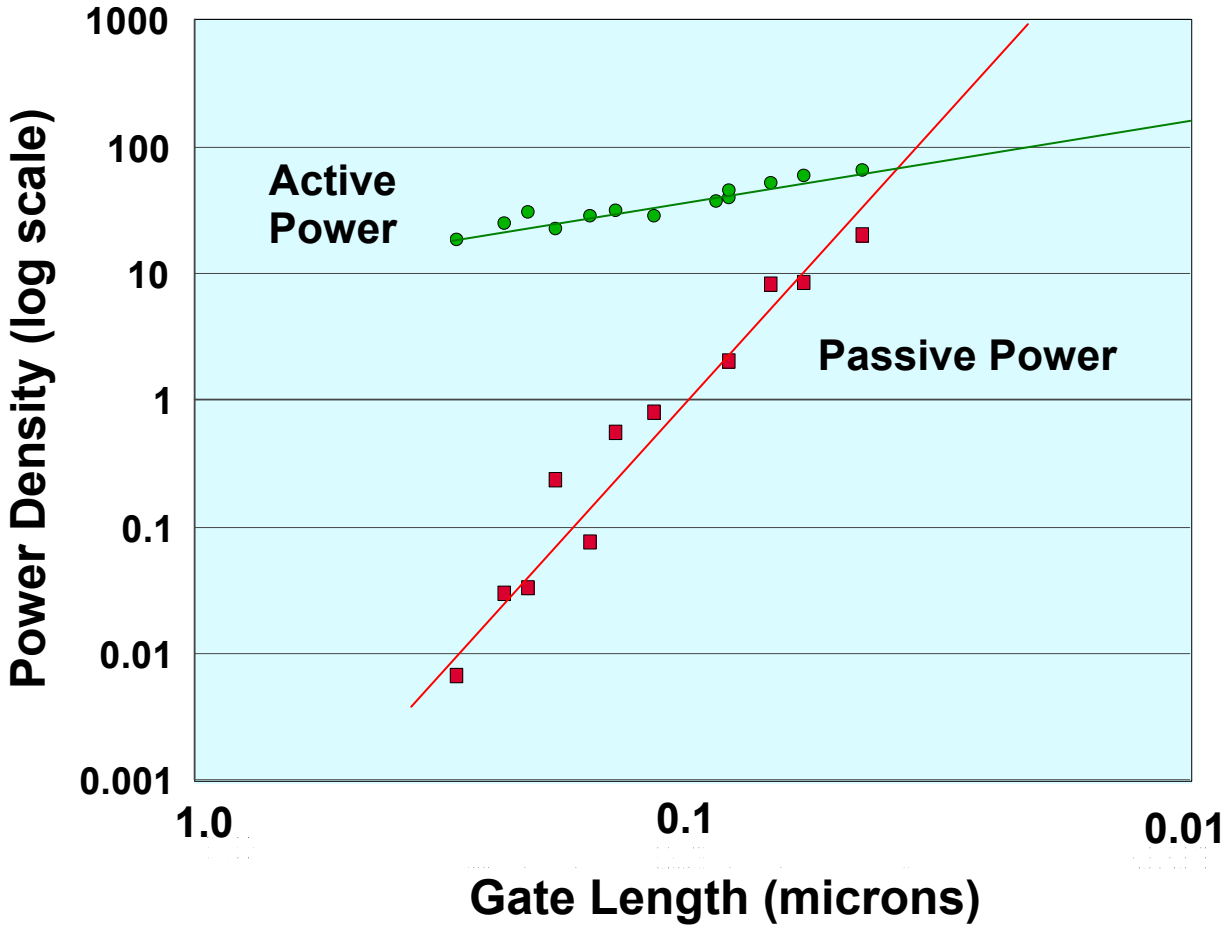
90 nm technology generation

… and in the lab.



B. Doris et al., *IEDM* , 2002

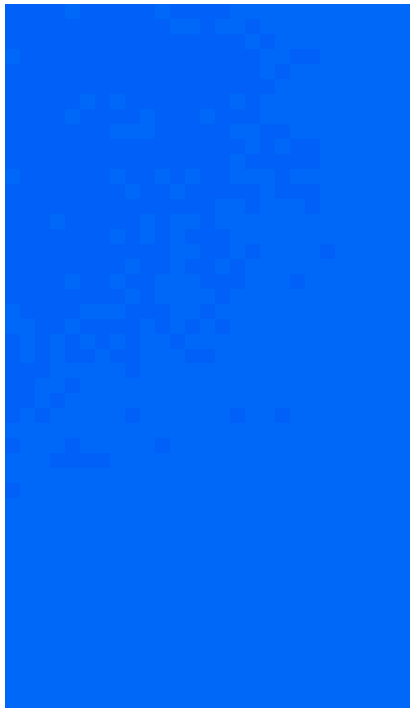
Still, we are approaching some limits.



Active Power Can Better Managed!

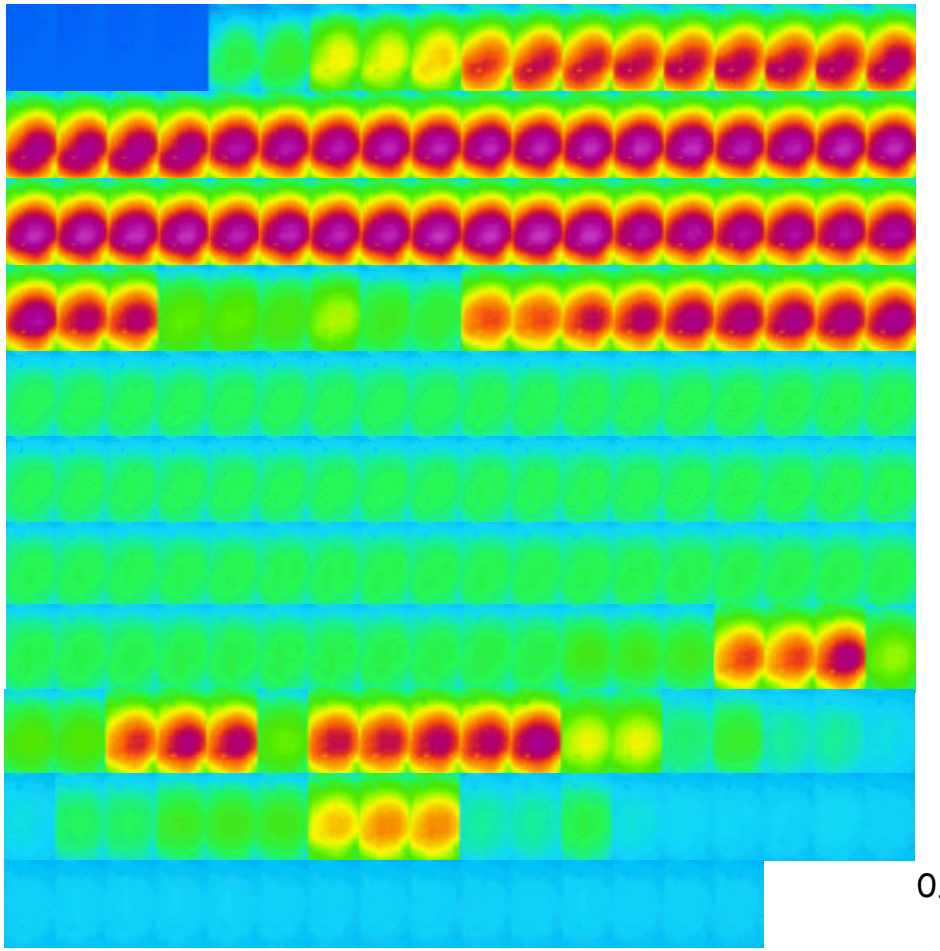
Thermal Mapping of Fully Operating IBM Microprocessor (2 GHz, 1.4 V) in a System Environment

Thermal images during bootup:



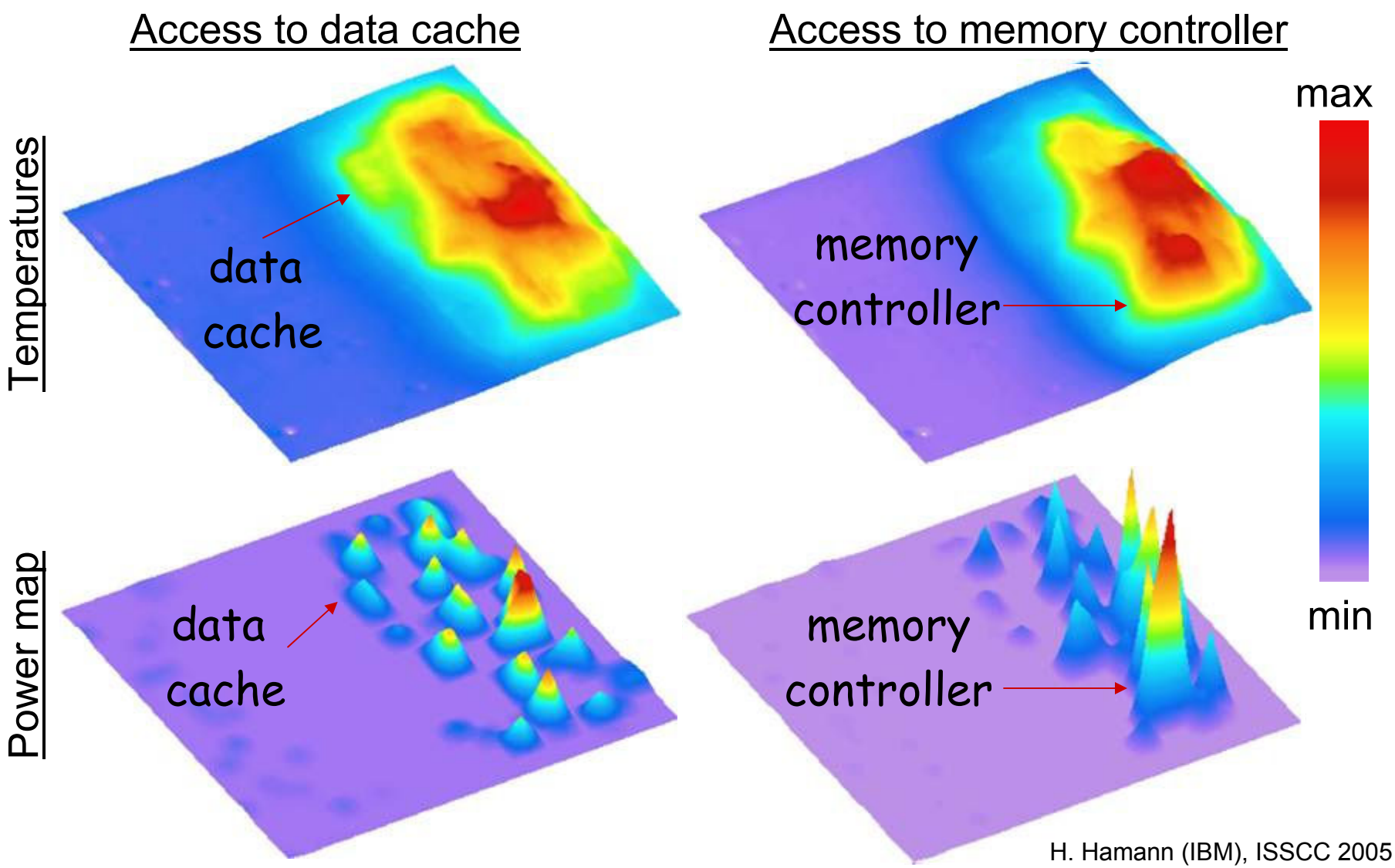
H. Hamann (IBM), ISSCC 2005

Filmstrip during bootup:



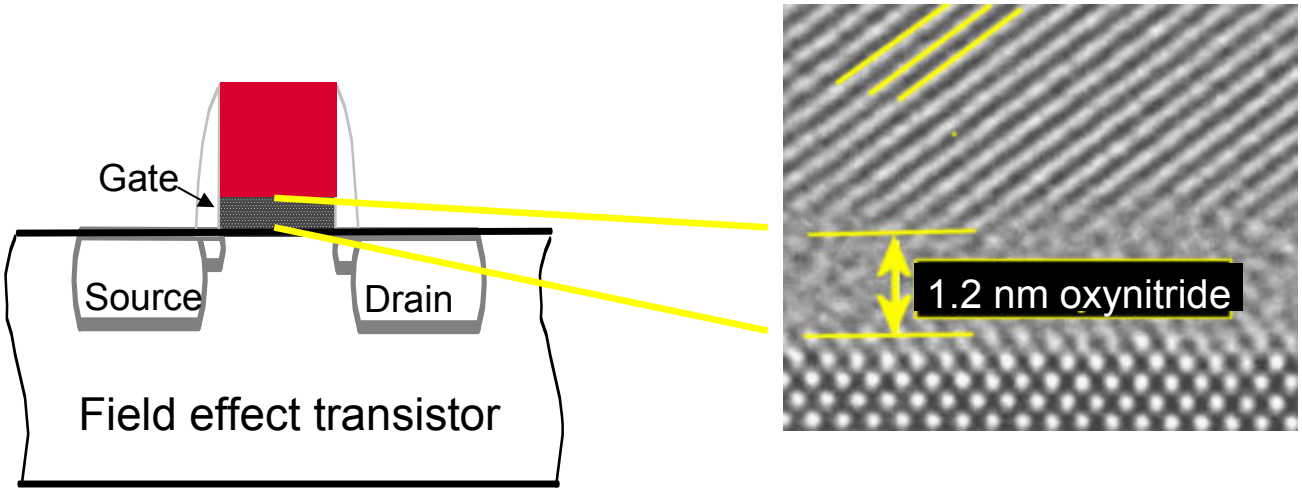
0.4 s per frame
Total ~ 80s

Better Information for Layout and High-level Design



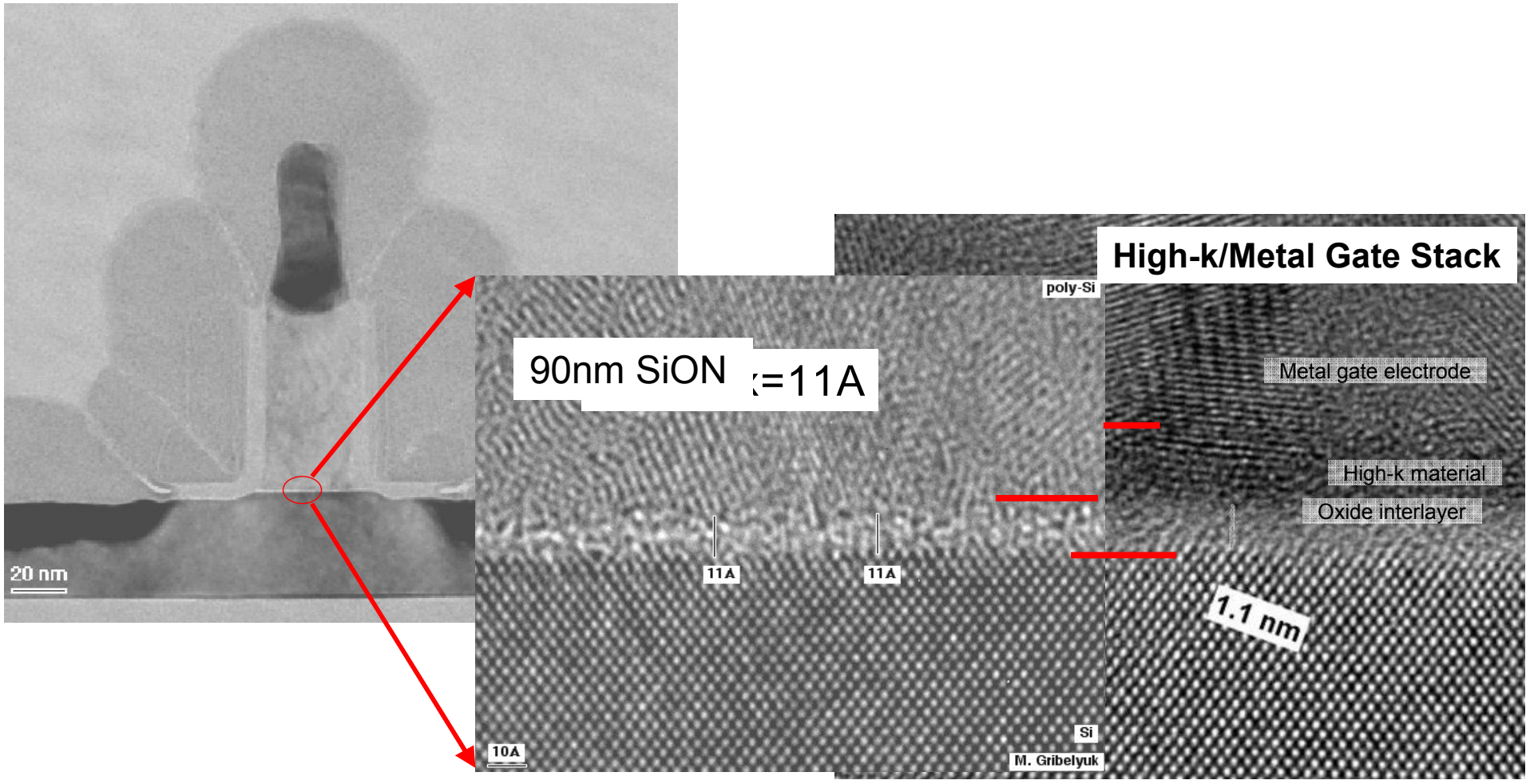
H. Hamann (IBM), ISSCC 2005

The Problem with Passive Power Dissipation: You Can't Scale Atoms!



- Direct tunneling through the gate insulator will be the dominant cause of static power dissipation.
- Single atom defects can cause local leakage currents 10 – 100x higher than the average current, impacting reliability and generating unwanted variation between devices.

The Work-Around: High-k Insulator / Metal Gate Stack

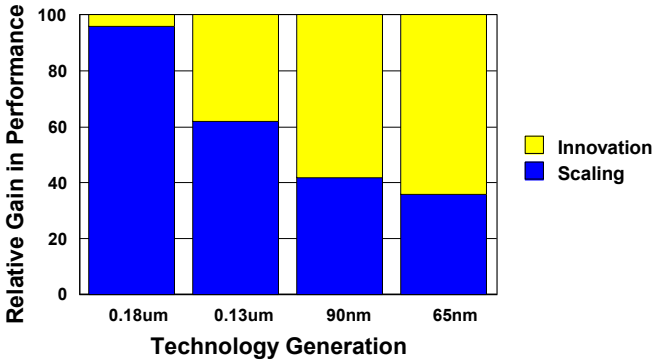


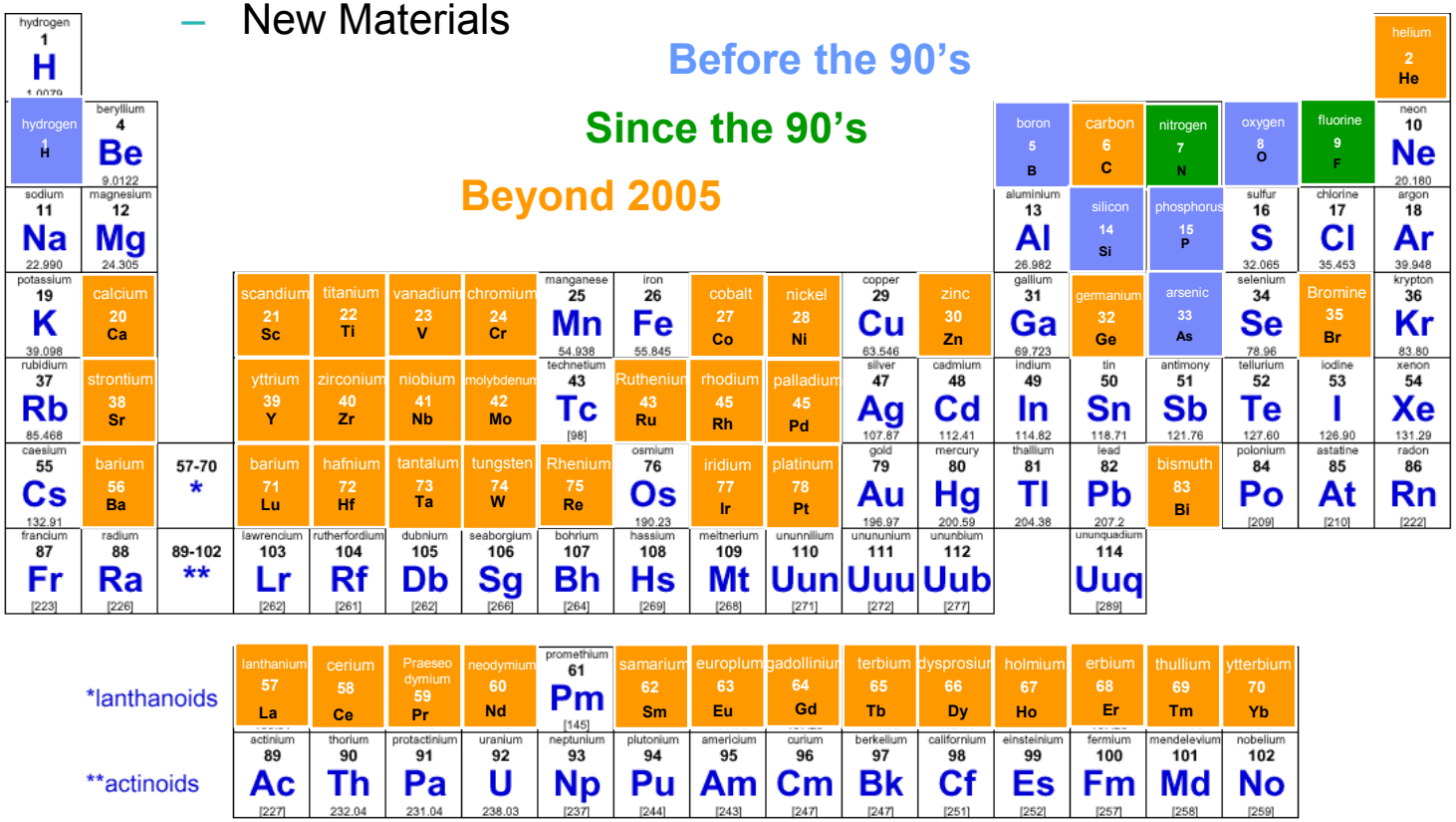
90nm Gate Dielectric:
Tinv = 19A
ToxGL = 11A

High-k/Metal Gate Stack:
Tinv = 14.5A
ToxGL = 16A

Improving Performance

- No longer possible by scaling alone
  - New Device Structures
  - New Device Design point
  - New Materials

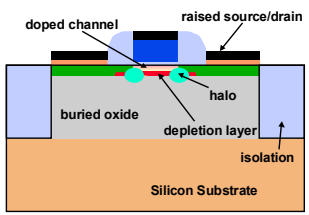




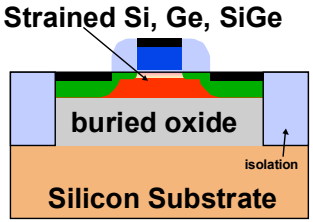
Innovation Will Continue: Transistor Roadmap Options

	2004	2007	2010	2013	2016	2020
Physical Gate	37 nm	25 nm	18 nm	13 nm	9 nm	6 nm

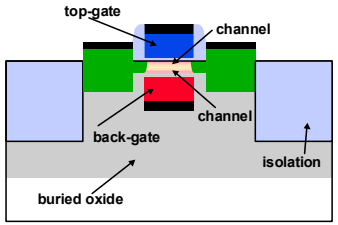
Ultrathin SOI



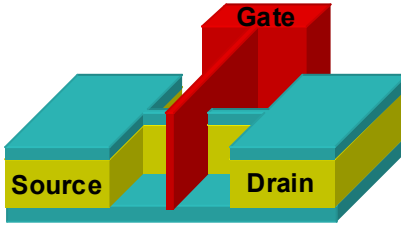
High *k* gate dielectric



Double-Gate CMOS

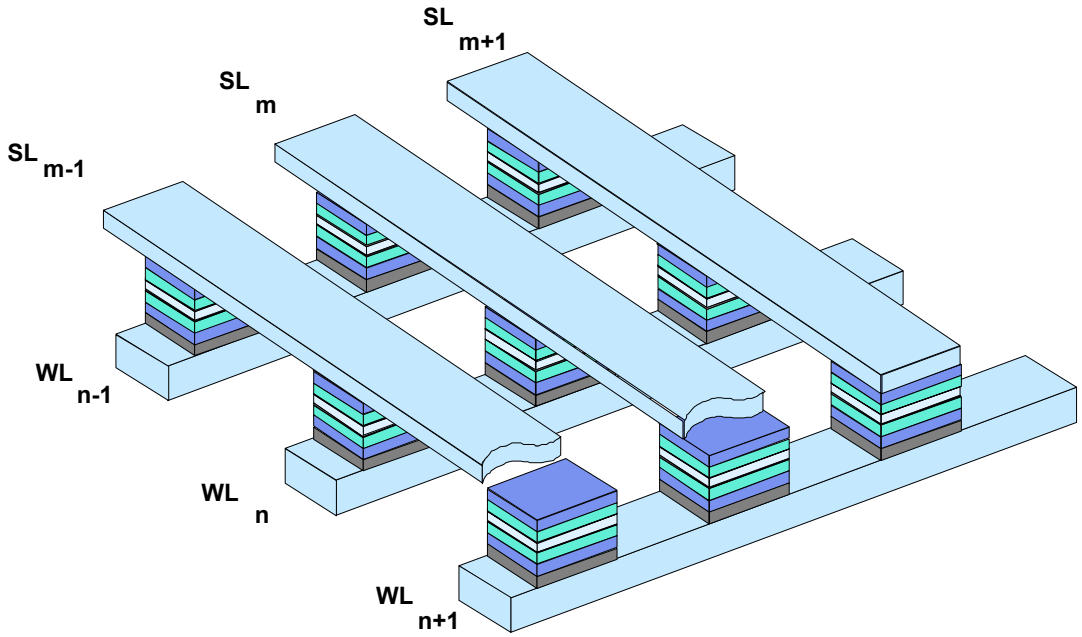


FinFET



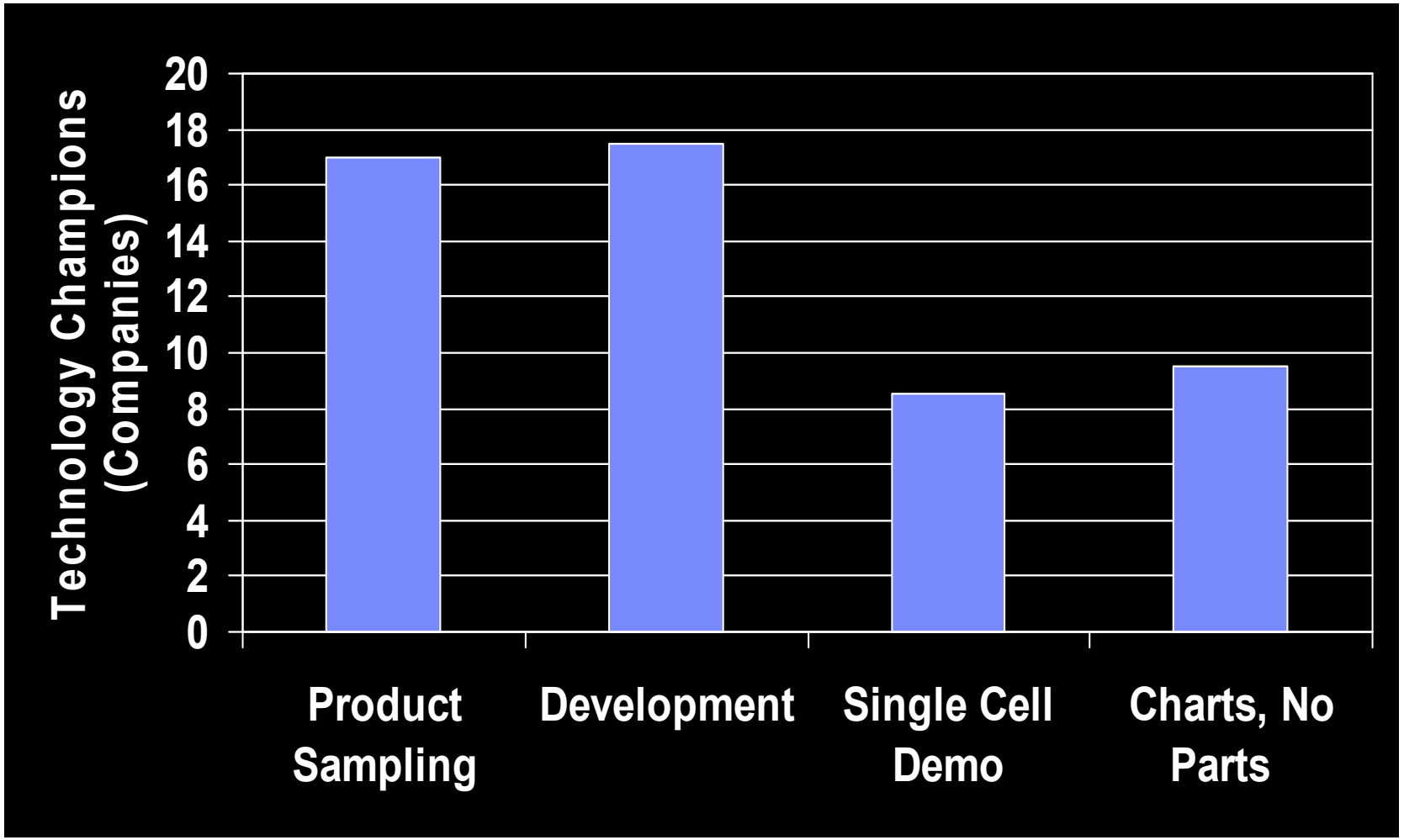
In general, growing power dissipation and increasing process variability will be addressed by introduction of new materials and device structures, and by design innovations in circuits and system architecture.

Memory may be easier to shrink than logic.



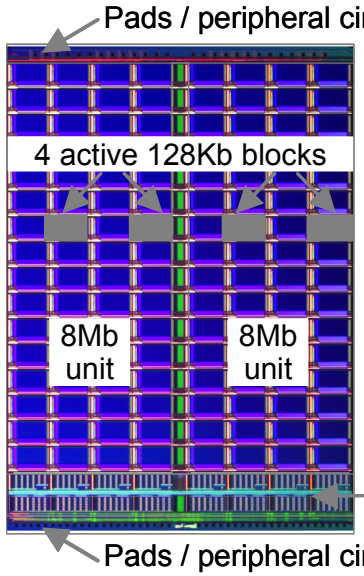
- Everyone is looking for a dense (cheap) cross-point memory.
- It is relatively easy to identify materials that show bistable hysteretic behavior (easily distinguishable, stable on/off states).

Relative Maturity of Nonvolatile Memory Technologies



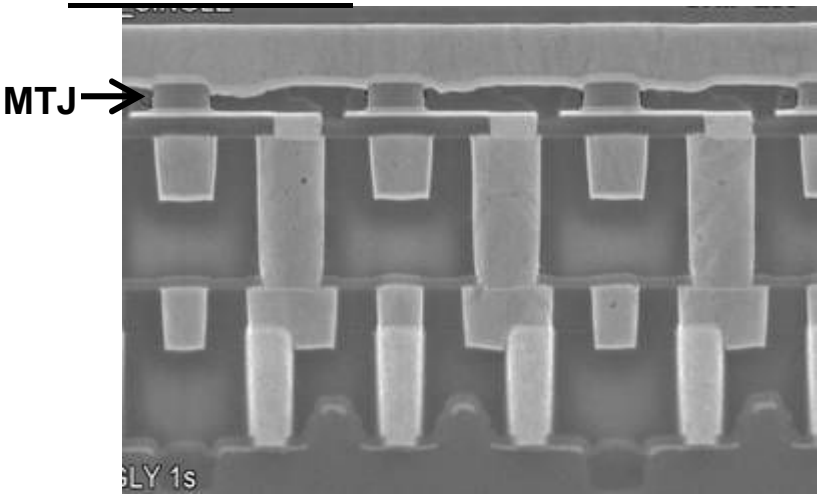
16Mbit Magnetic Random Access Memory (MRAM) Demo

Chip image:



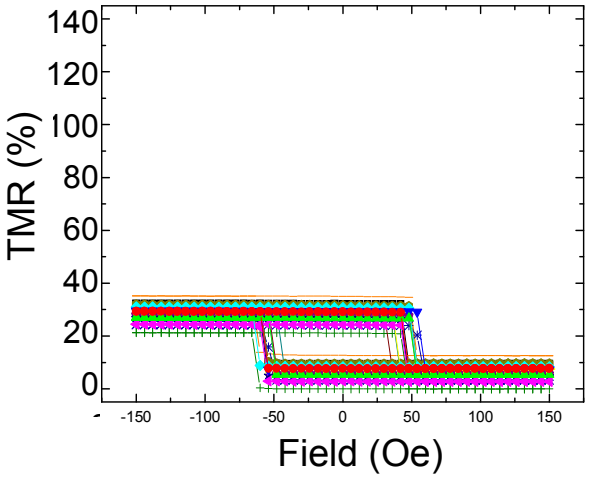
- 4 active bits / block
- Good power supply distribution
- Write currents trimmed by block

Cross-section:

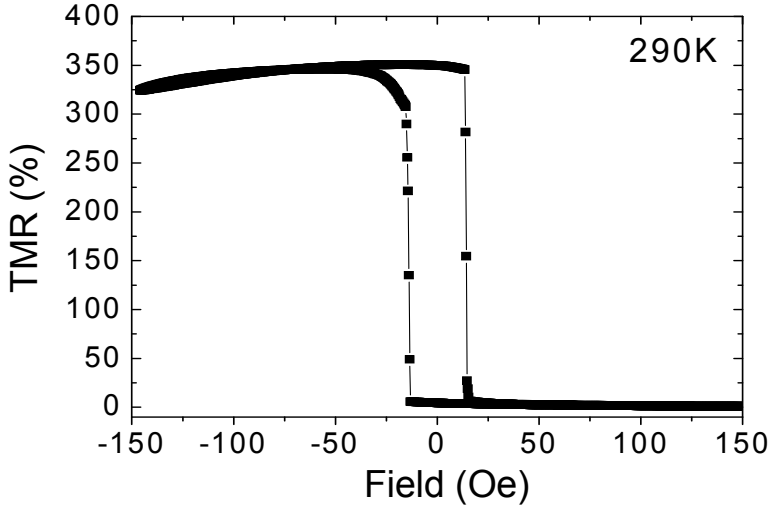


Materials Advances

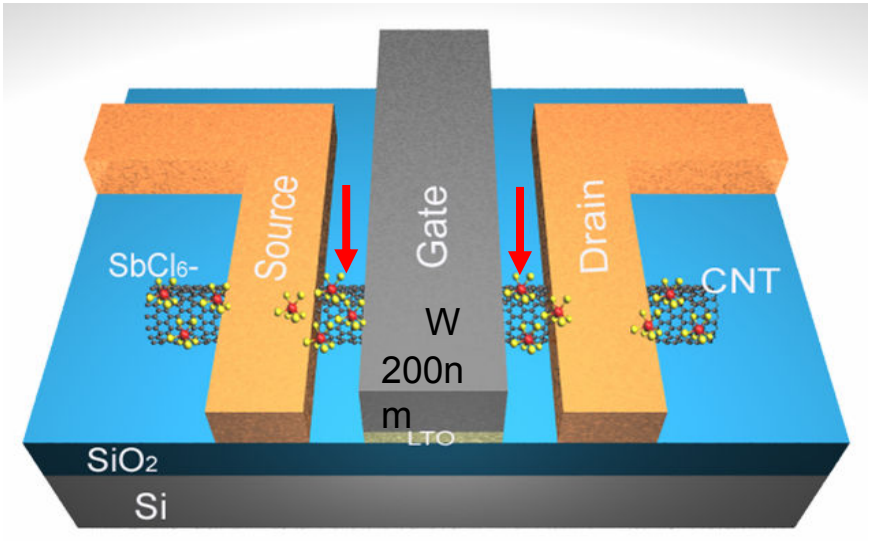
1995



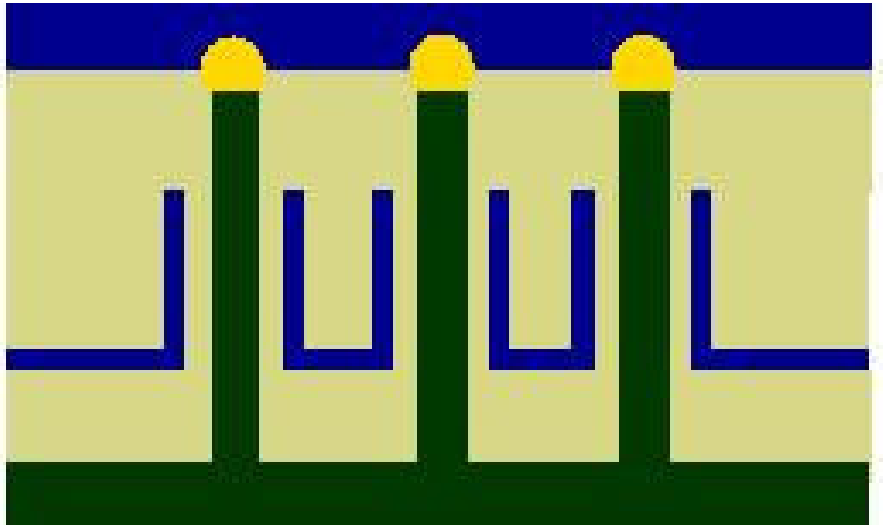
2005



Post-Silicon CMOS: The Quest for the Ultimate FET



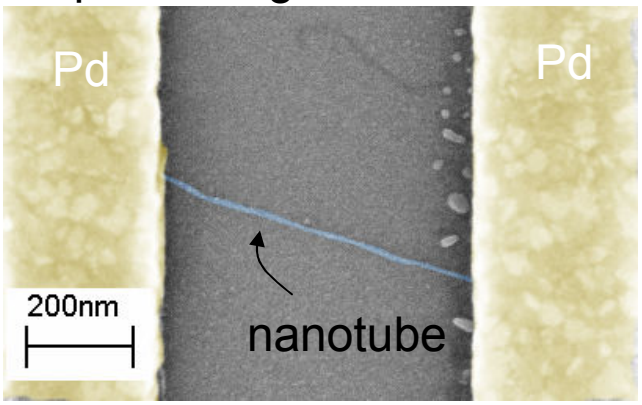
Self-Aligned Carbon Nanotube FET:
Extension Contacts Based on
Charge-Transfer Chemical Doping



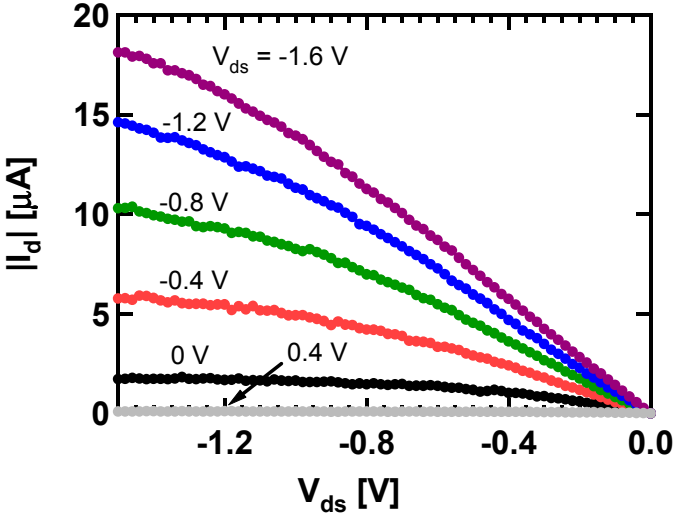
Vertical Transistor
Based on Semiconductor Nanowires

Intrinsic Performance of Cabon Nanotube FETs

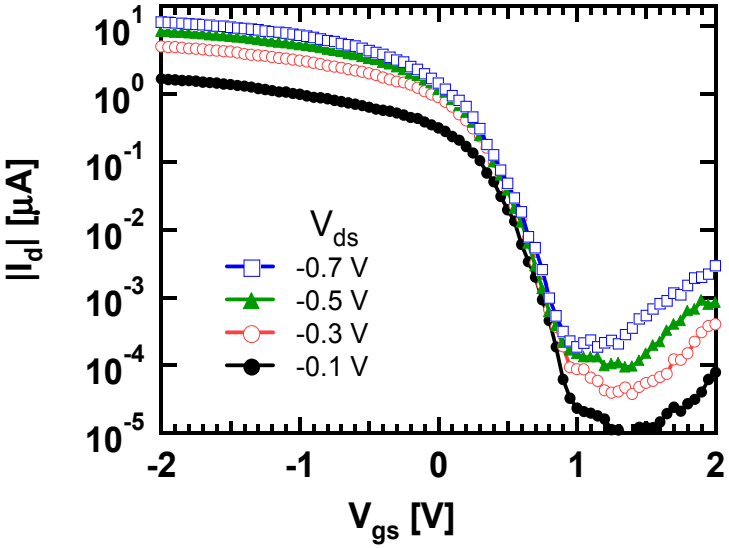
Simple back-gated CNTFET



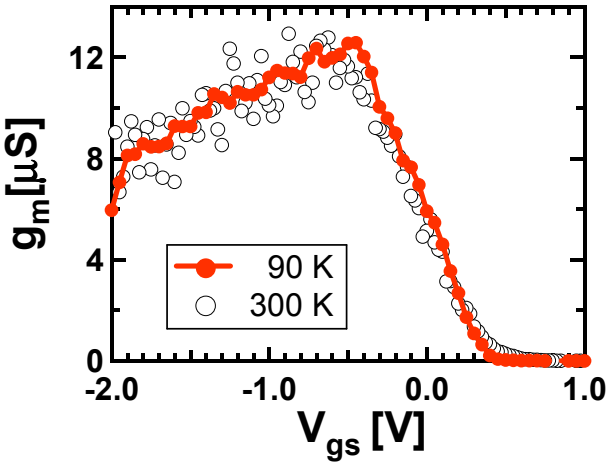
Output Characteristics



Subthreshold Characteristics



Temperature dependence



Yu-Ming Lin *et al.* (IBM), EDL 2005

Intrinsic Switching Speed of CNFETs

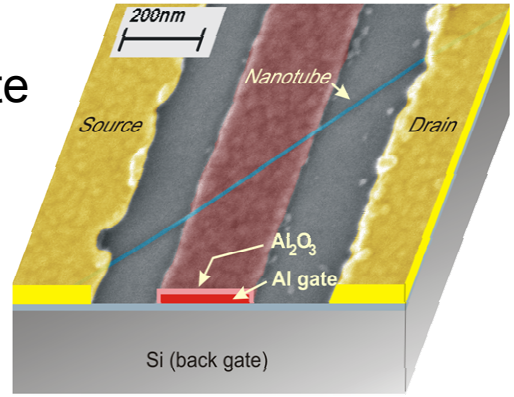
Cut-off Frequency $f_T = \frac{g_m}{2\pi C_g}$ C_g : gate capacitance

	Lin et al. (IBM)	**Javey et al. (Stanford)**	**Seidel et al. (Infineon)**
Diameter	~ 1.8 nm	~ 1.7 nm	~ 1.1 nm
Gate Dielectric	10-nm SiO_2	8-nm HfO_2	12-nm SiO_2
Maximum g_m	12.5 μS	27 μS	3.5 μS
C_g/L	38 pF/m	120 pF/m	32 pF/m
f_T @ L_g = 65 nm	**800 GHz**	550 GHz	260 GHz

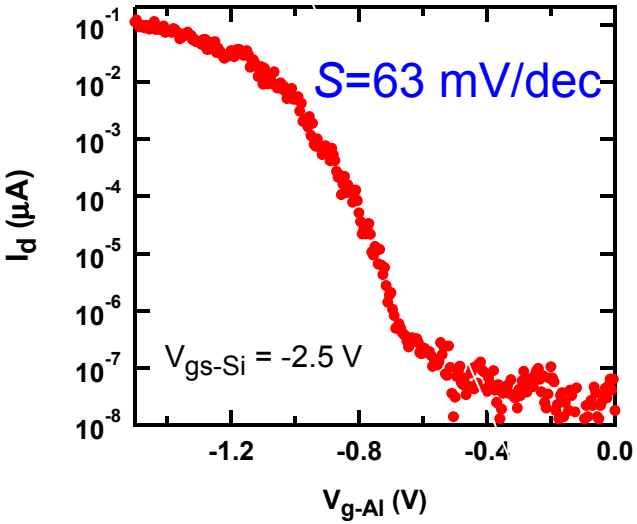
Yu-Ming Lin *et al.* (IBM), EDL 2005

Steep Sub-threshold Slopes for Low-Voltage Operation

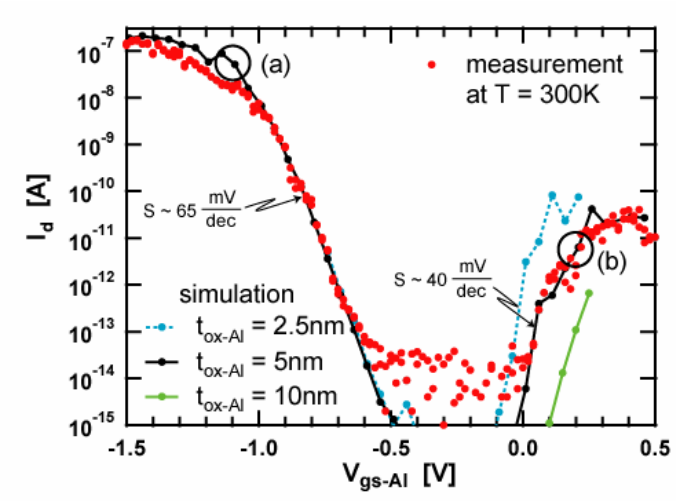
Dual-Gate CNTFET

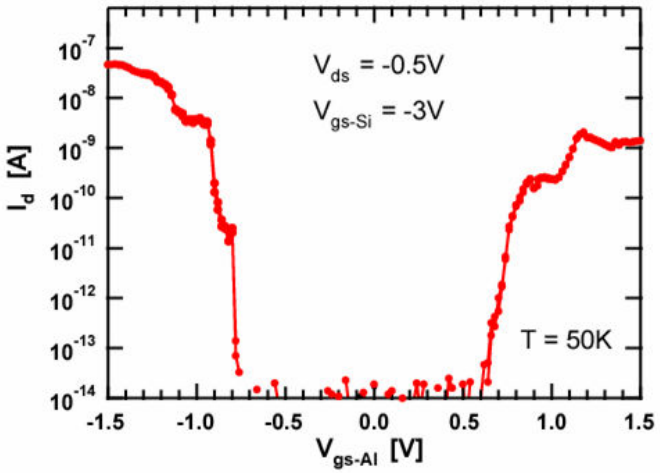


Bulk switching



Sub-thermal *S* through band-to-band tunneling





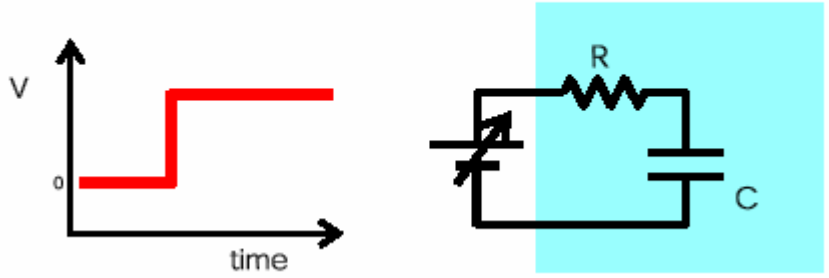
Adiabatic Computing and Reversible Logic

Can we operate FETs at or below the “kT” limit?

Adiabatic Computing

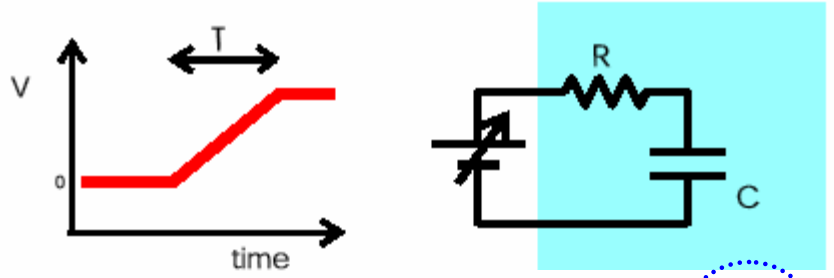
How much energy must be dissipated to charge a capacitor?

Abrupt method



$$E = \frac{1}{2} CV^2$$

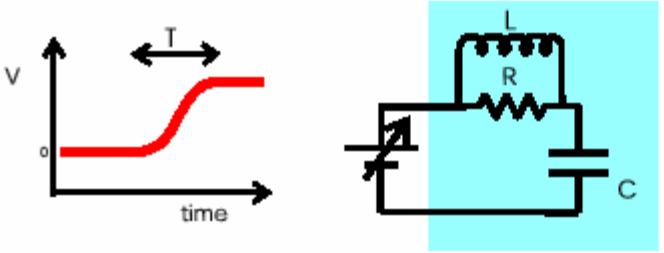
Quasi-static Charging (aka 'adiaba



$$E = \frac{1}{2} CV^2 \left(\frac{2RC}{T}\right)$$

This assumes T >> RC.

Quasi-static Charging + Superconductivity



Charging through a super-conductor, which behaves as an inductor and resistor in parallel.

$$E = \frac{\pi^4}{8} CV^2 \frac{RC(L/R)^2}{T^3}$$

This assumes T >> RC and T >> L/R.

(To implement logic in this system would require superconducting FETs. Such FETs are possible, but there have been very few experimental results.)

Trade-off depends strongly on device physics!

Adiabatic Switching

To take advantage of quasi-static charging in logic, there are 2 steps:

First, close switch ($V_{CLK} = V_{CAP}$)

Then, apply clock power (slowly)

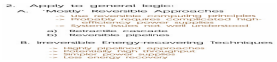
Rule 1: never close a switch (turn on an FET) while there is voltage across it.

Rule 2: don't ramp the voltage too quickly.

Applications of Adiabatic Charging

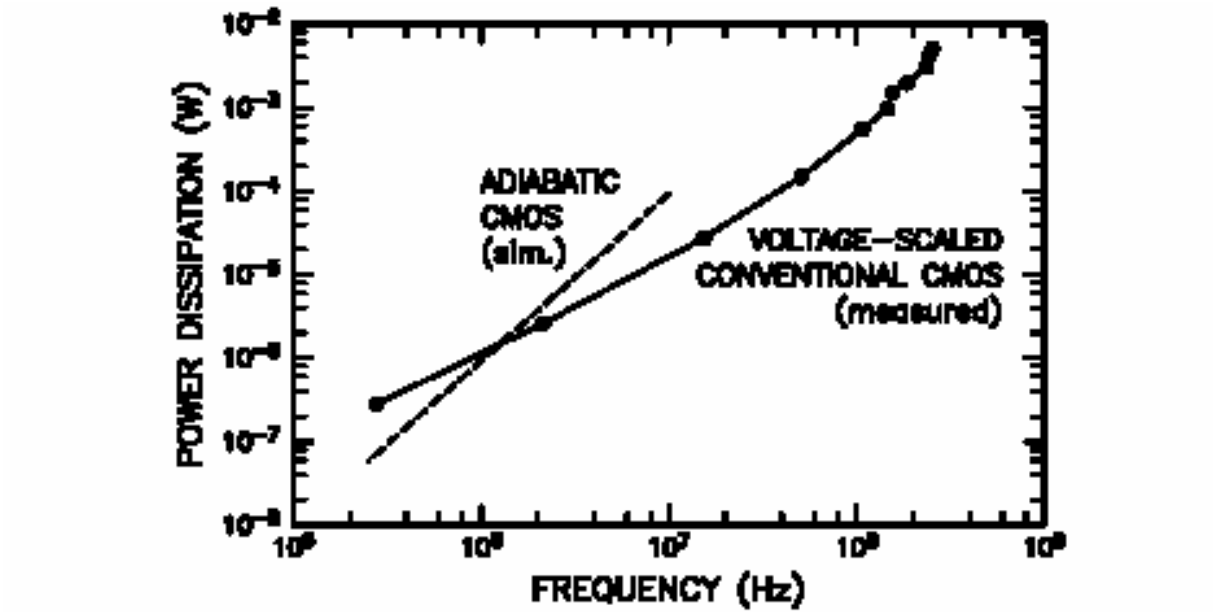
- Drive specific capacitances which cause large dissipation.
  - Power supplies
  - Energy conserving data bus drivers

- Broadly implement reversible logic.
  - Retractile cascade, reversible pipelines
  - High-efficiency regenerative power supply (very complex)



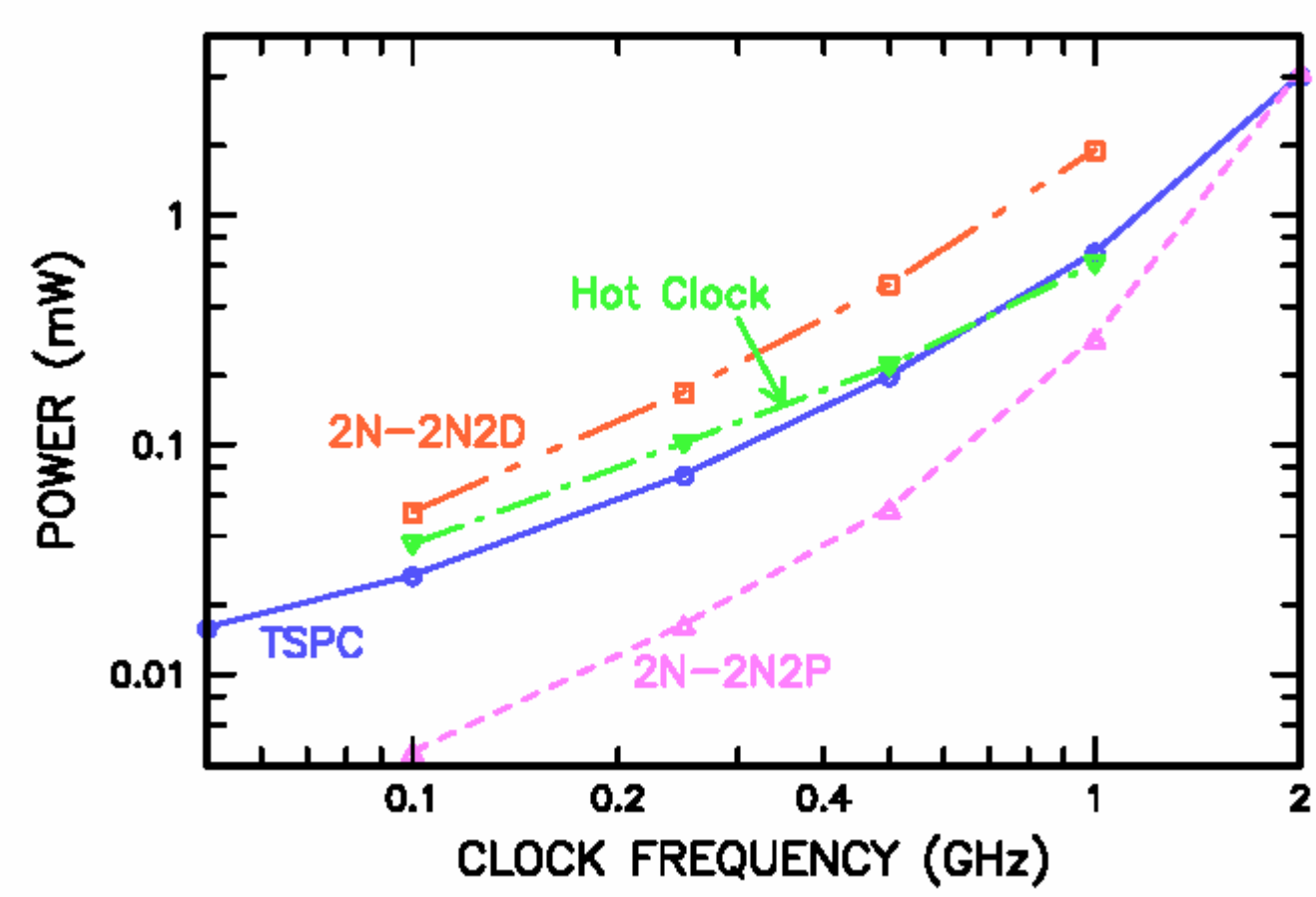
Reversible Logic: Implementation with FETs

- It is conceptually possible to build general purpose reversible computers with energy dissipation per operation going asymptotically to zero as frequency goes to zero.
- But, frequency must be reduced by about 1/1000 to achieve benefits with respect to conventional approaches to CMOS logic.



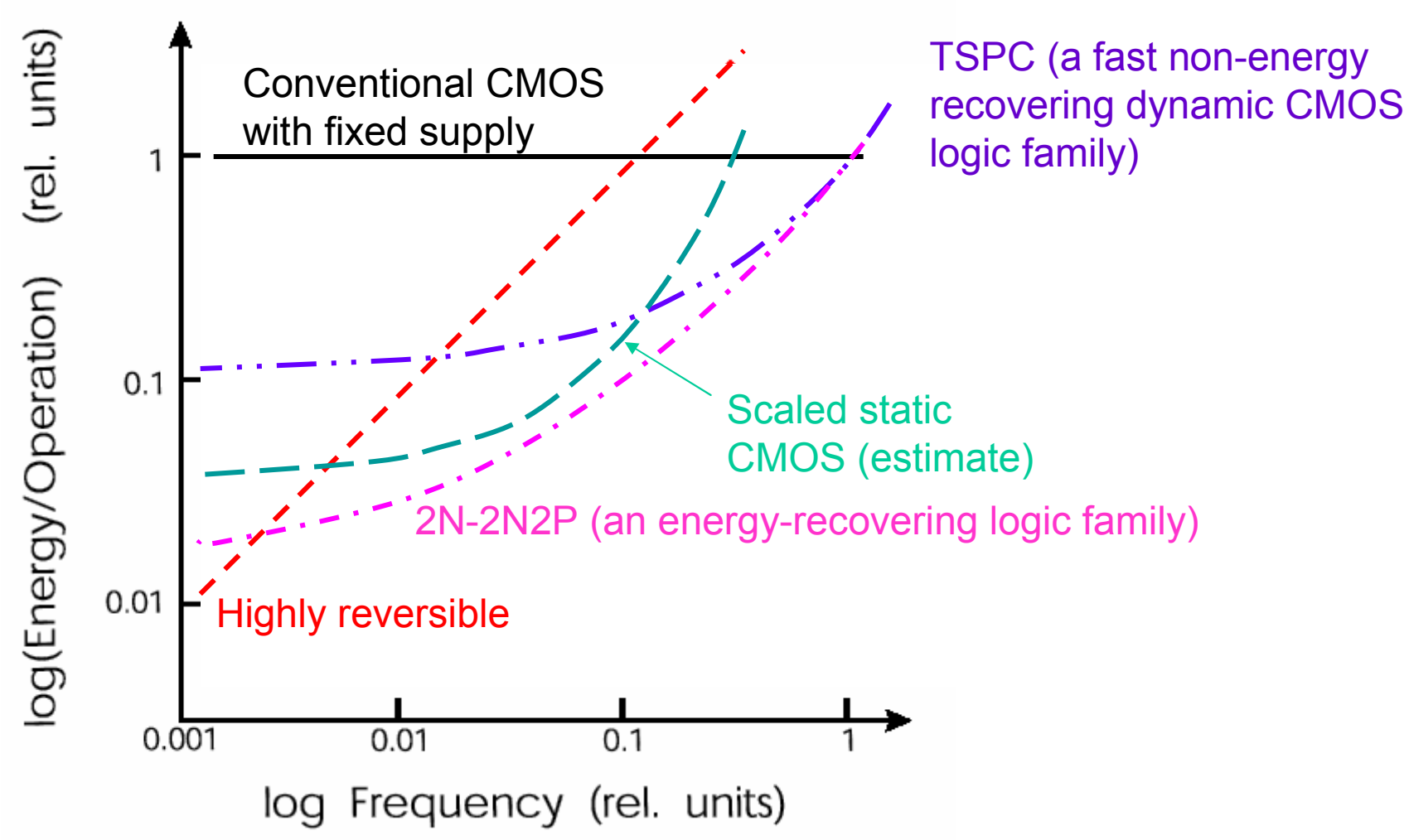
Dissipation of 4 bit ripple counter (D. J. Frank, 1995)

Recent Optimized Power vs Frequency Assessment



DJ Frank, MIT Workshop on Reversible Computation, February 14, 2005

Qualitative Overall Comparison

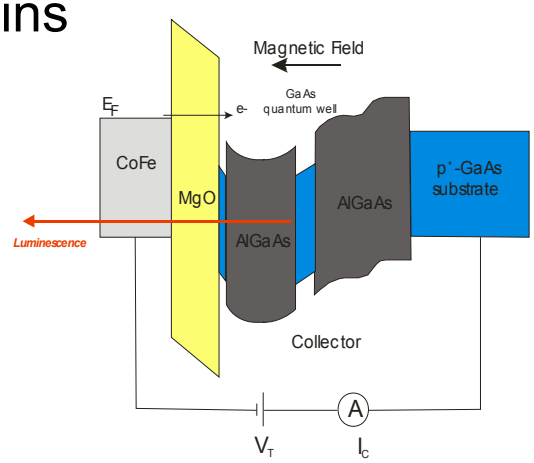


DJ Frank, MIT Workshop on Reversible Computation, February 14, 2005

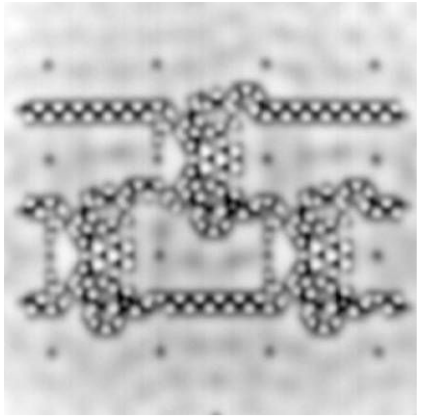
Are there devices that are better suited than FETs for the implementation of reversible logic?

Beyond Charged-Based Logic?

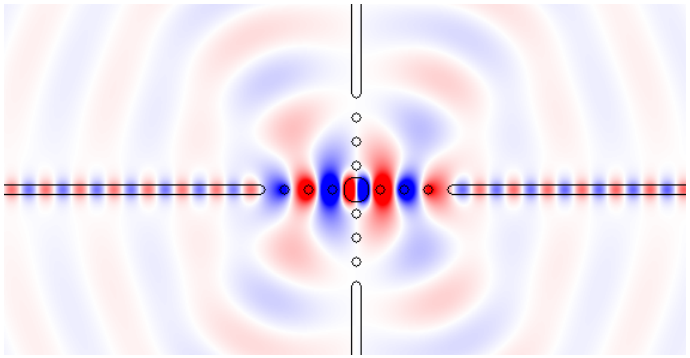
- Spins



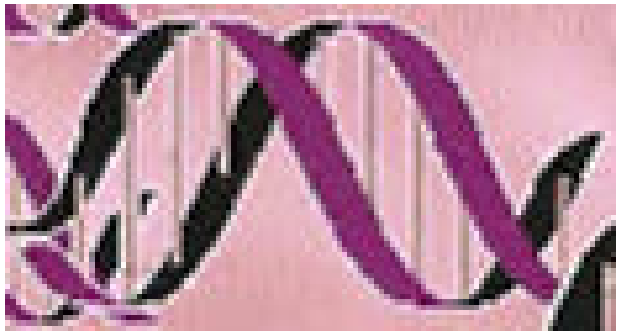
- Nanomechanics



- Photons



- DNA Chemistry



Nanoelectronics Research Initiative (NRI)

- AMD, Freescale, Micron, TI, IBM, Intel
  → Joint Industry funding of University Research
- Leveraging existing NSF and new (joint NSF/NRI) funding
- Promoting both
  - Invention / Discovery (distributed research, “let many flowers bloom”)
  - Proof of Concept (focused university consortia with outstanding facilities)
- “Extend the historical cost/function reduction, along with increased performance and density … orders of magnitude beyond the limits of CMOS”
  - Non-charge-based logic
  - Non-equilibrium systems
  - Novel energy transfer mechanisms for device interconnection
  - Phonon engineering
  - Directed self-assembly of complex structures

Conclusions

- Silicon CMOS logic will be extended at least another 10 years.
  - New materials and transistor structures
  - Cooperative circuit and device technology co-design
- The “ultimate FET” may not contain silicon.
- New, dense, cheap memory devices are on the way.
- Reversible logic is possible, but may be practical only with devices that are “beyond the FET”.
- The search for logic devices “beyond the FET” is underway, (but there should be more focus on devices suitable for adiabatic computation and reversible logic)

Thanks!

Non-Si FET

- The potential of carbon nanotubes (CNT)
  - Metallic & semiconducting nanotubes: interconnects (up to ~10^9A/cm^2) & switches.
  - 1D transport (low elastic and inelastic scattering low energies, ballistic or quasi-ballistic transport, $f_T \leq 1$THz.
  - Low energy dissipation (power density problem, no electromigration)
  - Good control of electrostatics (“thin body devices”, no mobility degradation).
  - Inert (no dangling bonds to passivate, wide choice of gate insulators – high k materials)
  - Direct band-gap materials (integrate electronic & opto-electronic devices using the same material).