

Trends and Challenges in High-Performance Microprocessor Design

Stefan Rusu

Senior Principal Engineer
Intel Corporation

stefan.rusu@intel.com

Agenda

- **Microprocessor Design Trends**
- **Process Technology Directions**
- **Active Power Management**
- **Leakage Reduction Techniques**
- **Thermal Modeling**
- **Call to Action**
- **Summary**

Microprocessor Evolution



	4004 Processor	Itanium [®] 2 Processor
Year	1971	2004
Transistors	2300	592 M
Process	10 um	0.13 um
Die size	12 mm ²	432 mm ²
Frequency	108 kHz	1.7 GHz

Moore's Law - 1965

The experts look ahead

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

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wrist-watch needs only a display to be feasible today.

But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Computers will be more powerful, and will be organized in completely different ways. For example, memories built of integrated electronics can be distributed throughout the

machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

Present and future

By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional ones that result in electronics functions supplied to the user as irreducible units. These technologies were first investigated in the late 1950's. The object was to miniaturize electronics equipment to include increasingly complex electronic functions in limited space with minimum weight. Several approaches evolved, including microassembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and converged so that each borrowed techniques from another. Many researchers believe the way of the future to be a combination of the various approaches.

The advocates of semiconductor integrated circuitry are already using the improved characteristics of thin-film resistors by applying such films directly to an active semiconductor substrate. Those advocating a technology based upon films are developing sophisticated techniques for the attachment of active semiconductor devices to the passive film arrays.

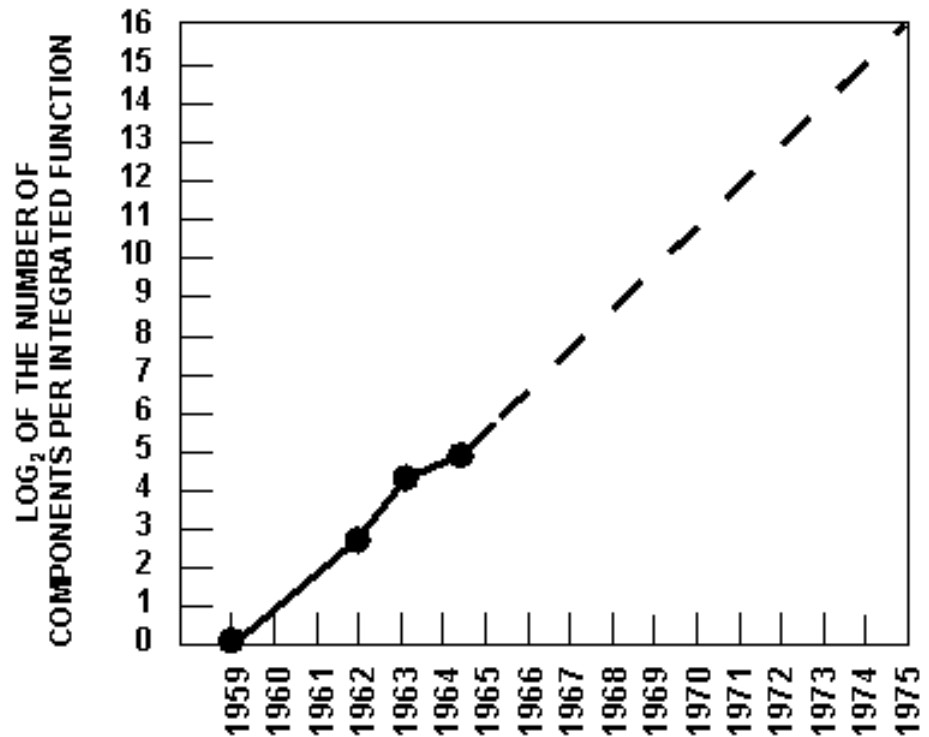
Both approaches have worked well and are being used in equipment today.

The author



Dr. Gordon E. Moore is one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics. He earned a B.S. degree in chemistry from the University of California and a Ph.D. degree in physical chemistry from the California Institute of Technology. He was one of the founders of Fairchild Semiconductor and has been director of the research and development laboratories since 1955.

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



Electronics, April 1965

Past Forecasts

“Heavier-than air flying machines are not possible”

Lord Kelvin, 1895

“I think there is a world market for maybe five computers”

IBM Chairman Thomas Watson, 1943

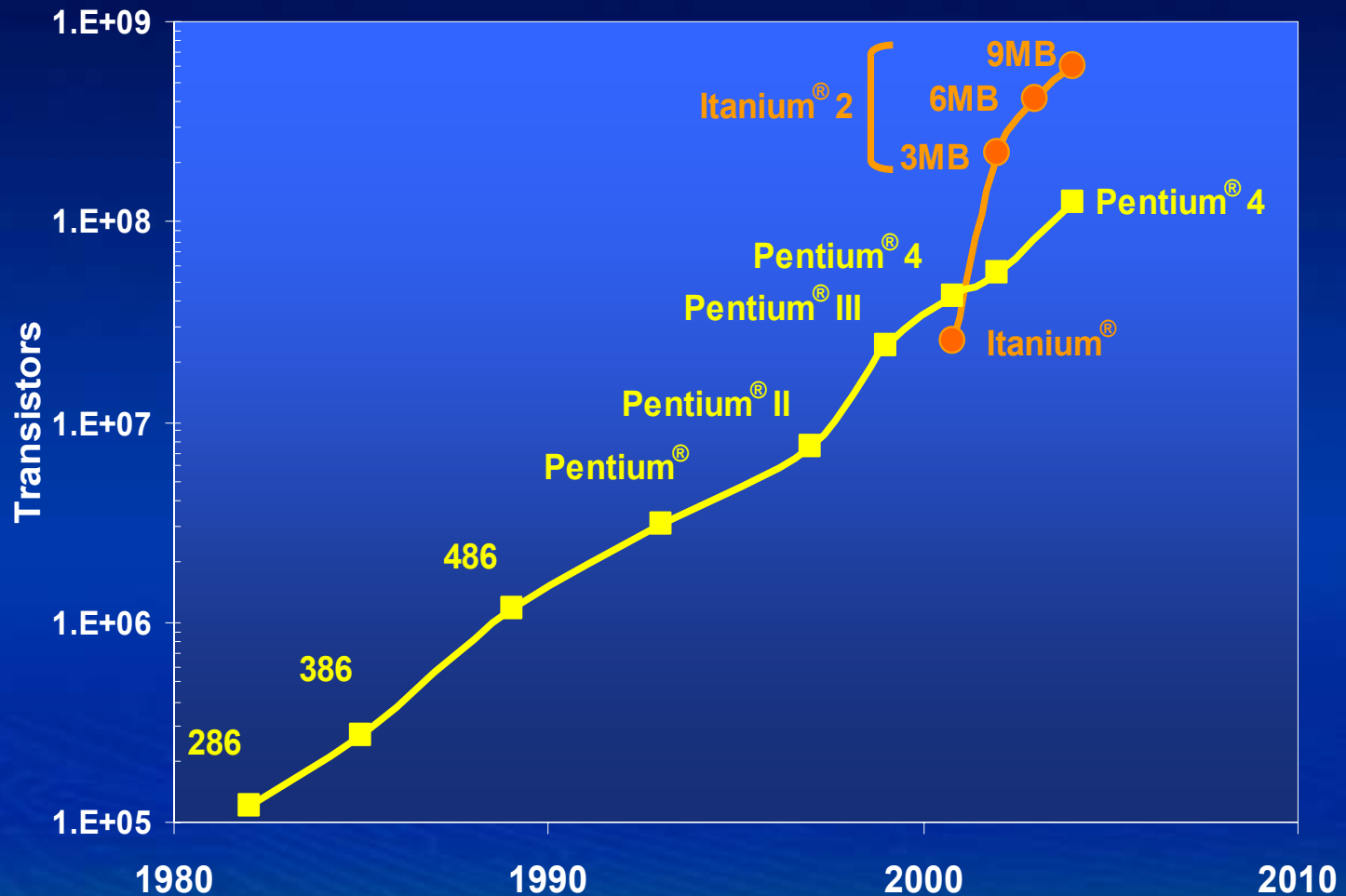
“640,000 bytes of memory ought to be enough for anybody”

Bill Gates, 1981

“The Internet will catastrophically collapse in 1996”

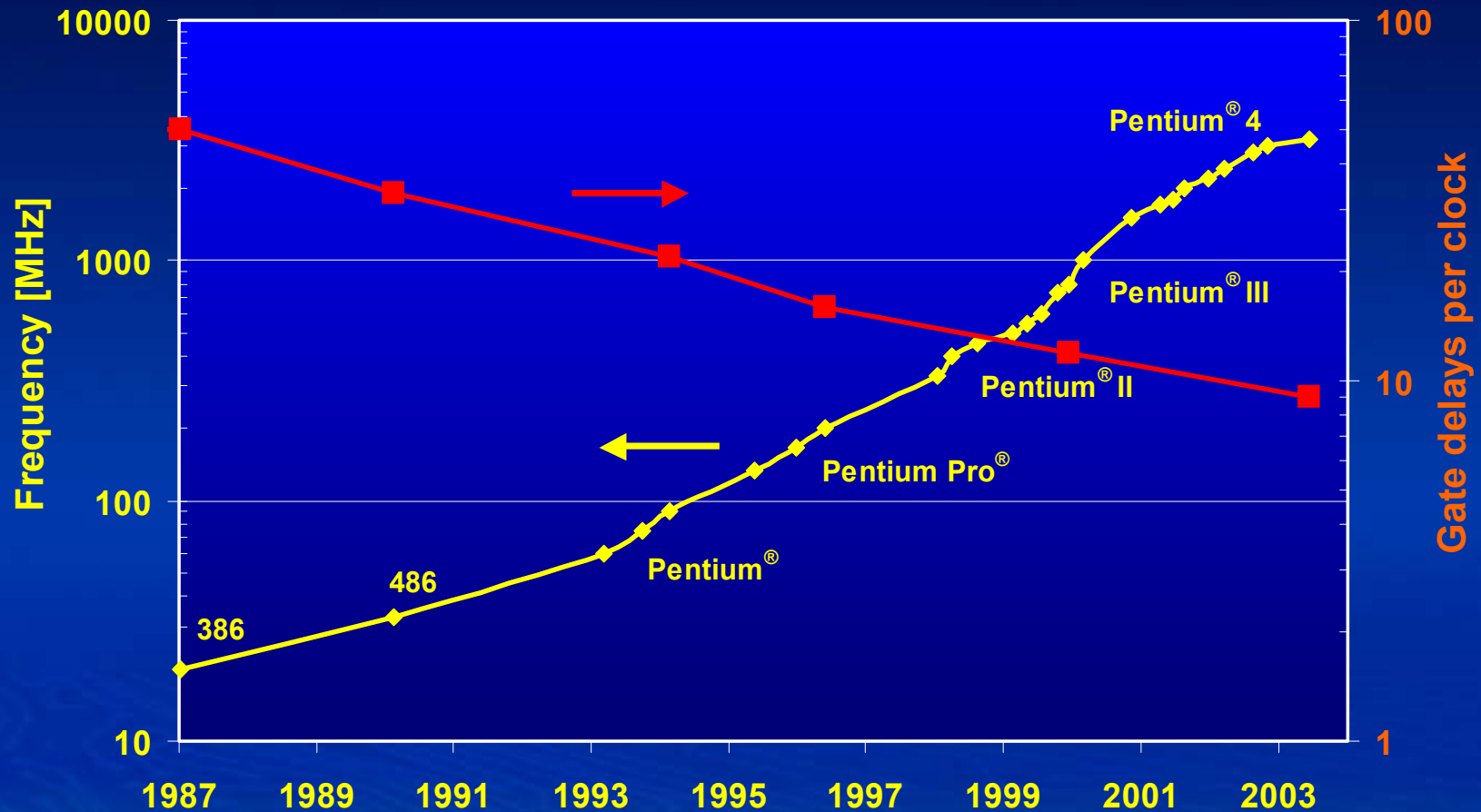
Robert Metcalfe

Moore's Law Continues



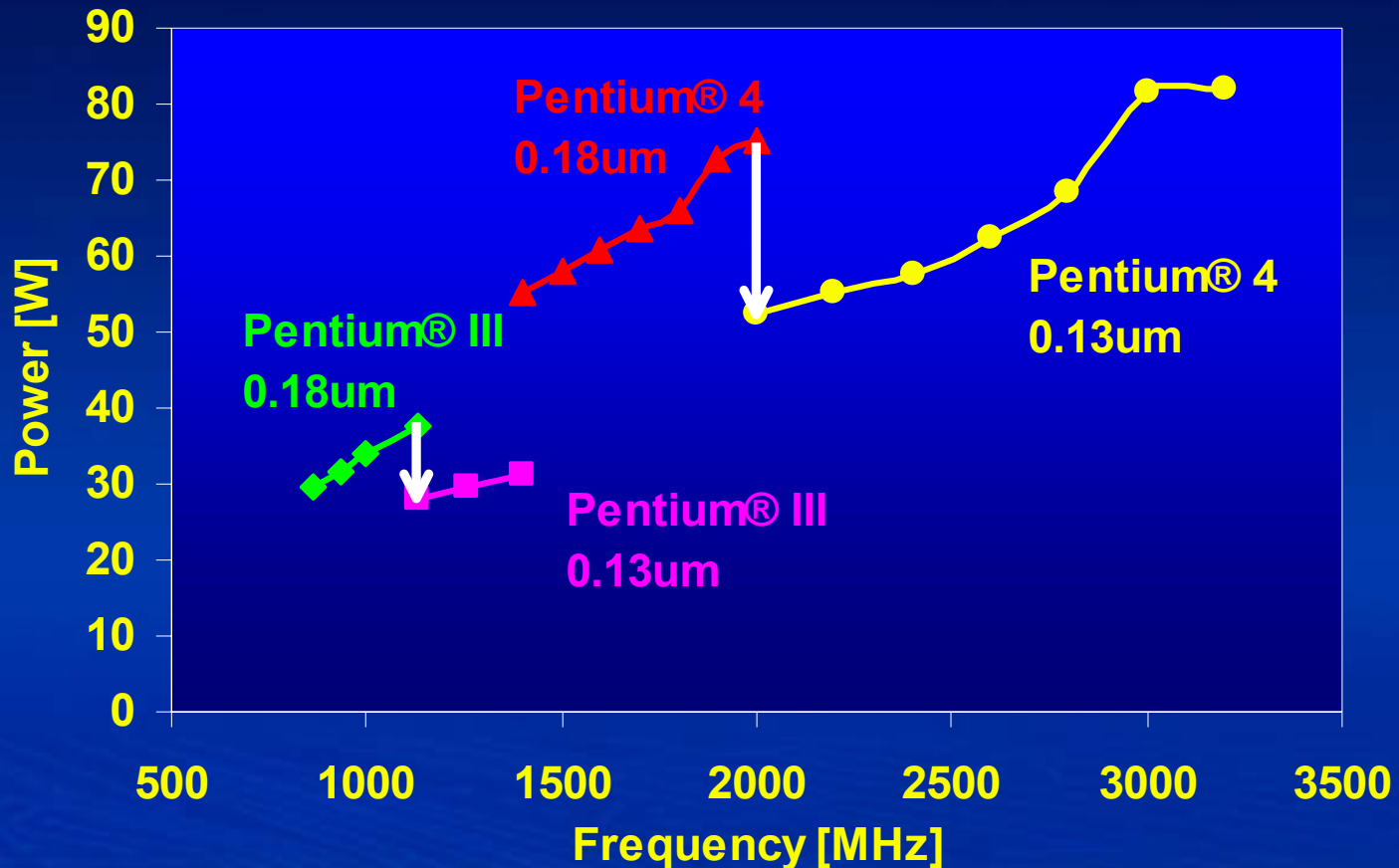
Heading Toward 1 Billion Transistors By 2005

Processor Frequency Trend



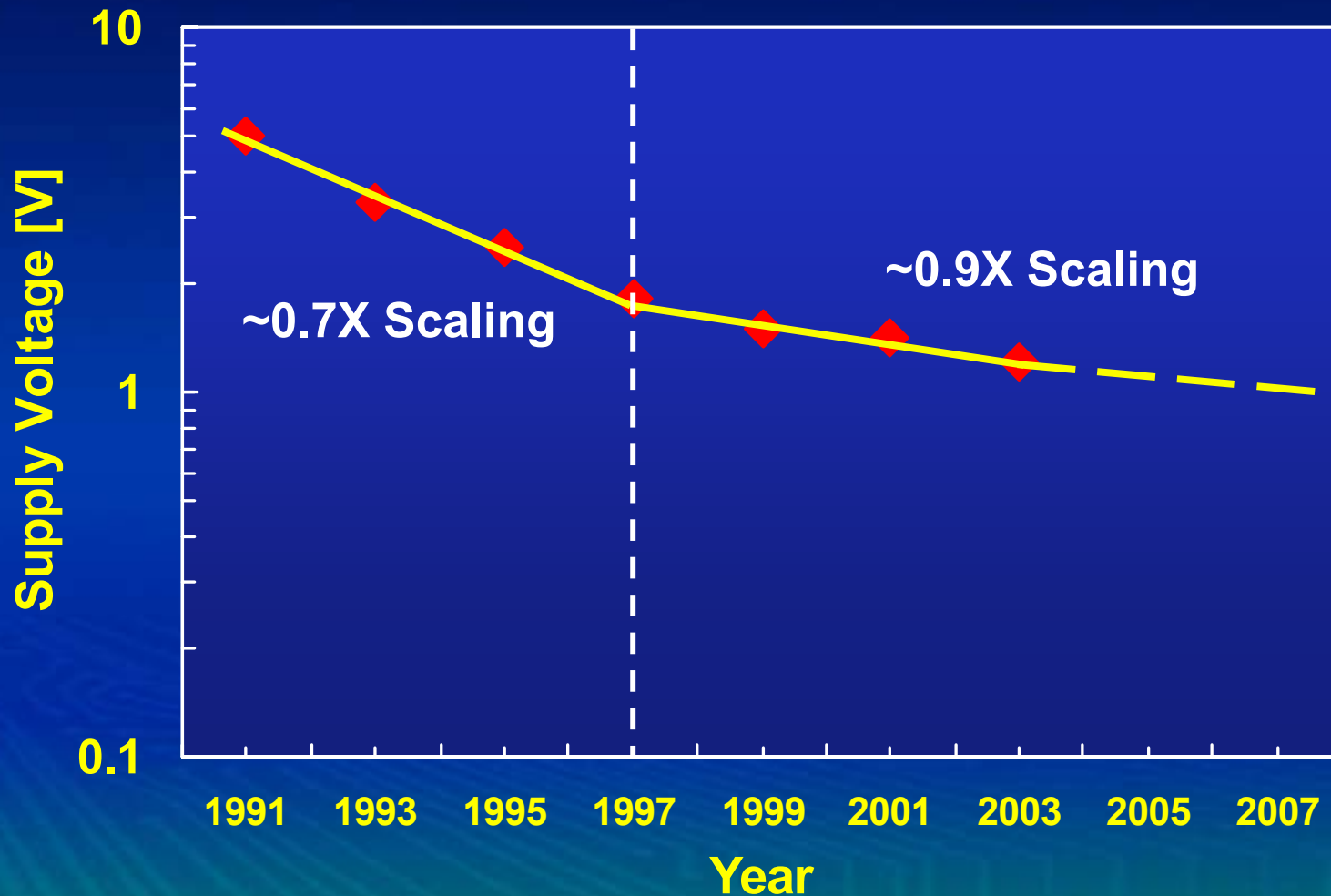
- Frequency doubles each generation
- Number of gates per clock reduces by 25%

Processor Power Trend

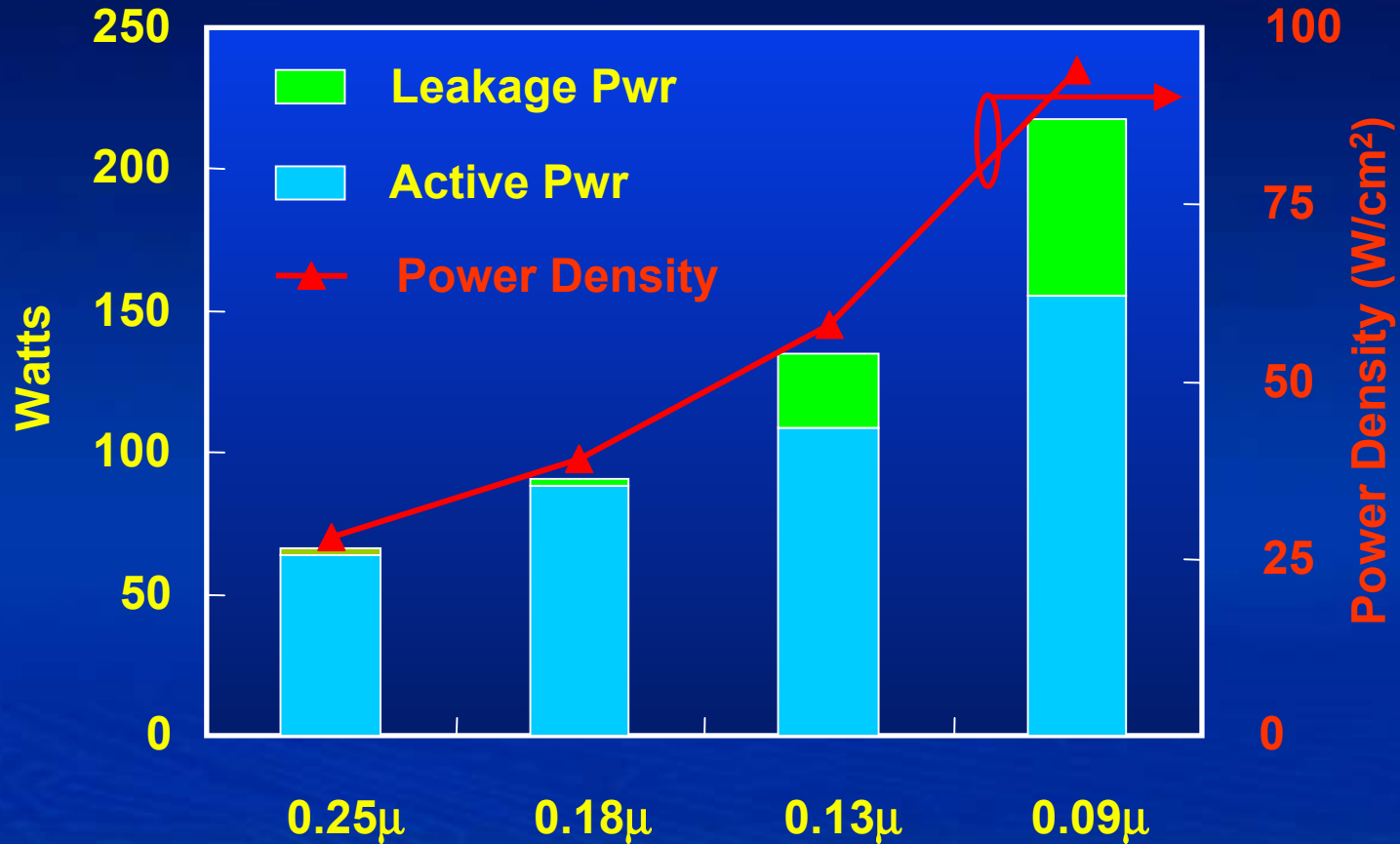


- Lead processor power increases every generation
- Process scaling provides higher performance at lower power

Voltage Scaling Is Slowing Down

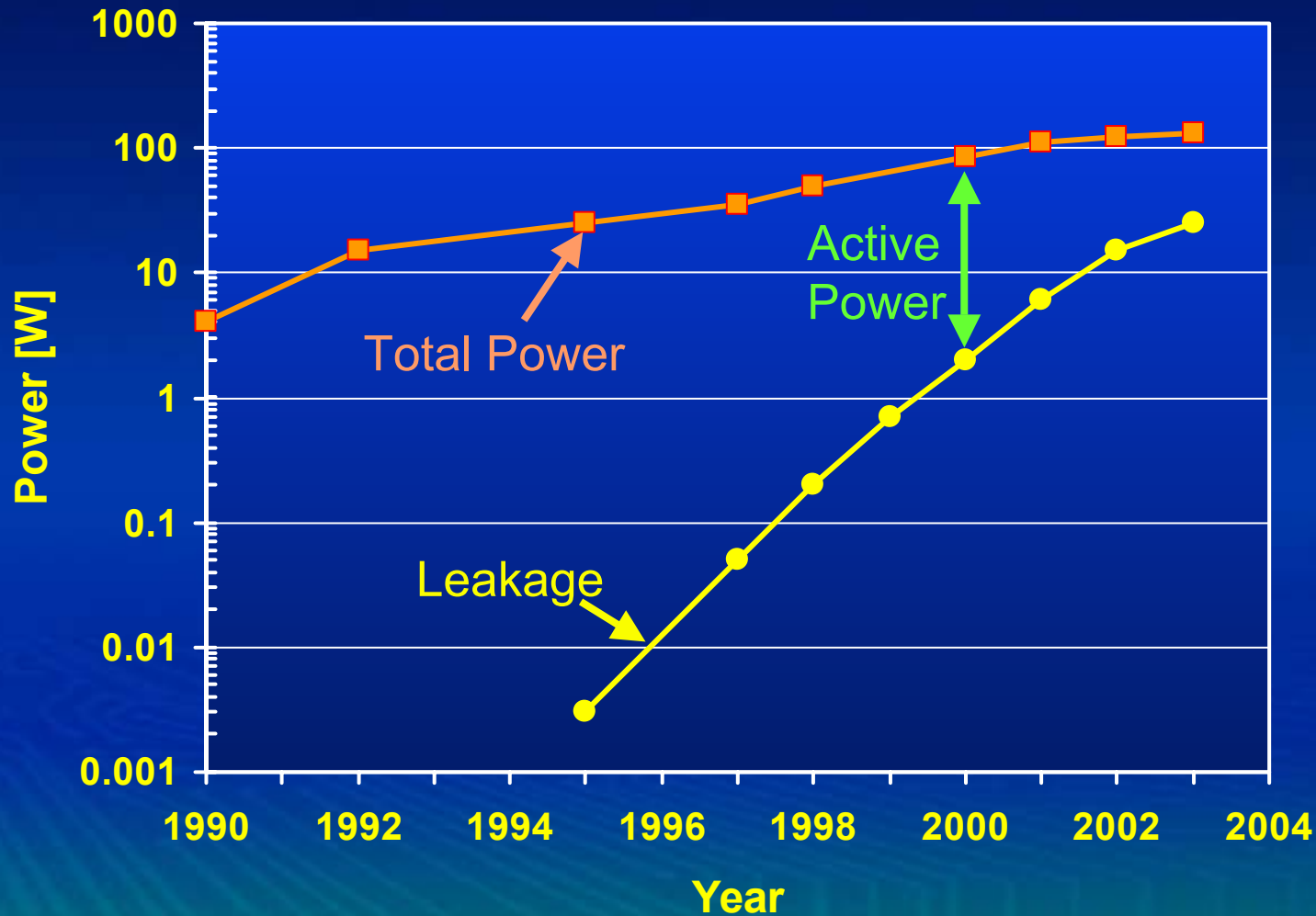


Power Density Trend

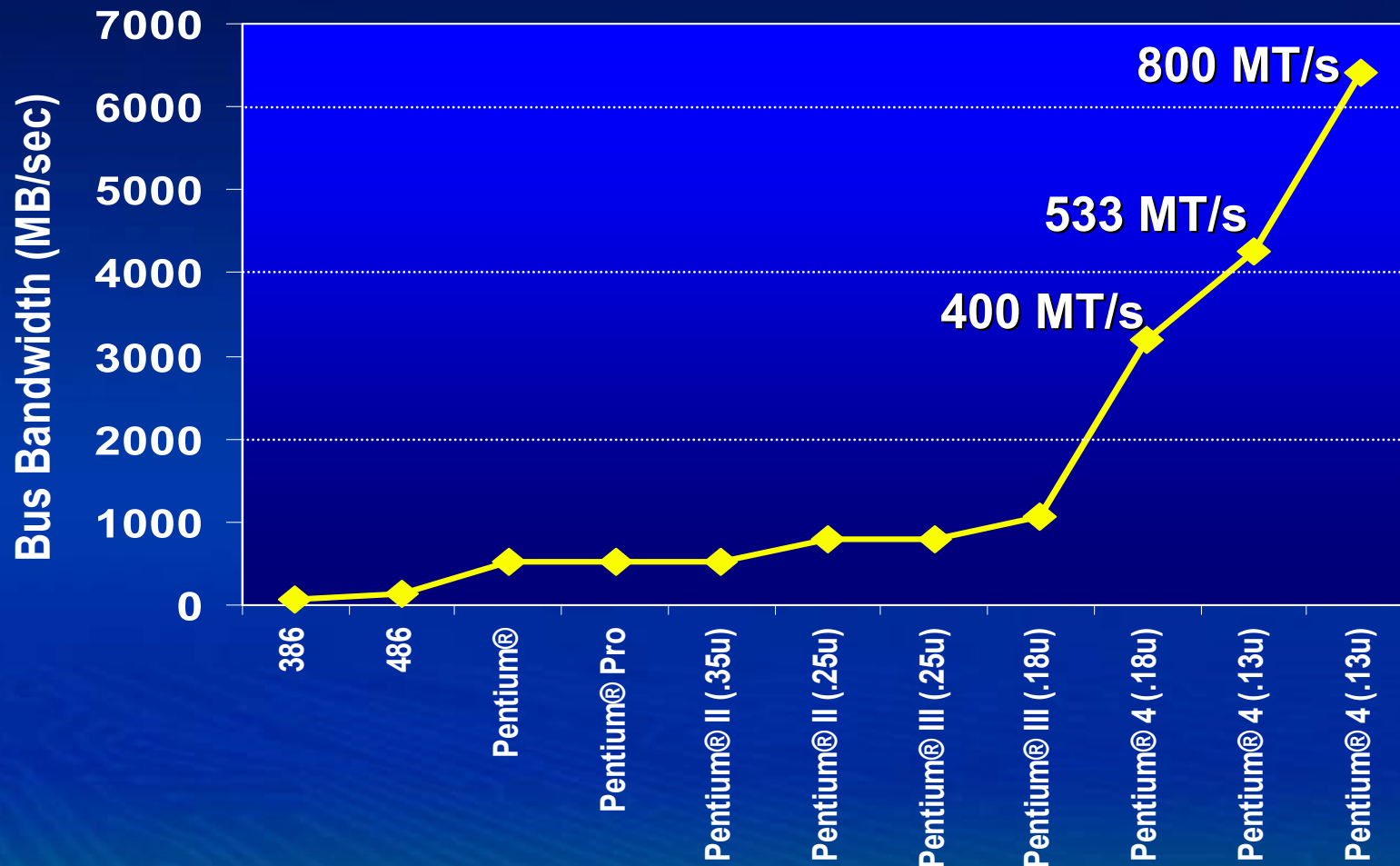


- Assumptions: 15mm die, 1.5x frequency increase per generation

Active and Leakage Power Trends



Bus Bandwidth Trend



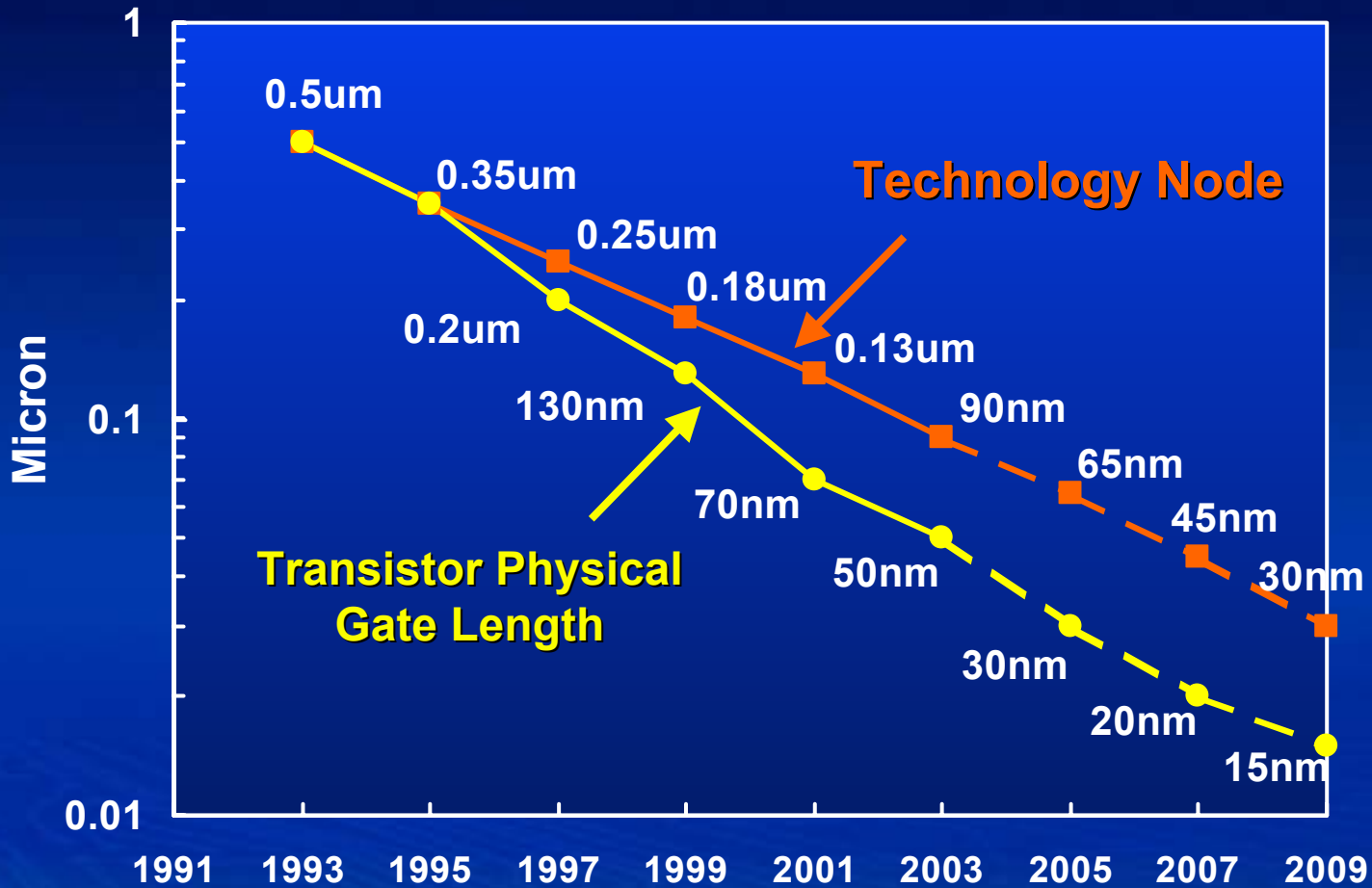
Memory and I/O Bandwidth Are Essential For High Performance



Agenda

- **Microprocessor Design Trends**
- **Process Technology Directions**
- **Active Power Management**
- **Leakage Reduction Techniques**
- **Thermal Modeling**
- **Call to Action**
- **Summary**

Transistor Physical Gate Length

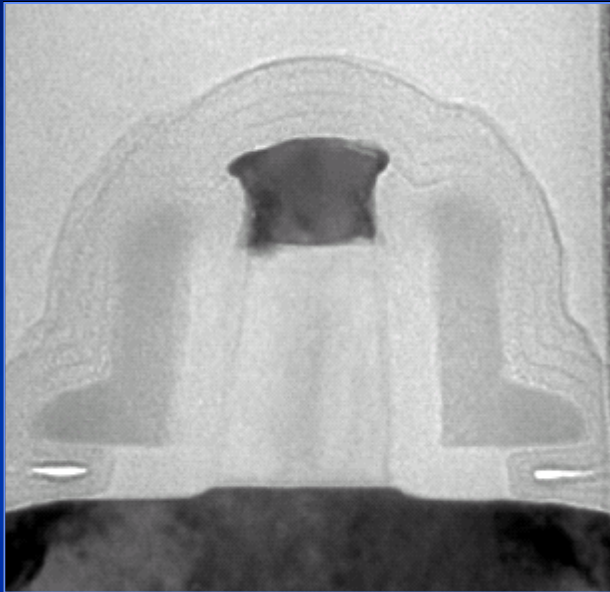


New Process Generation Every 2 Years

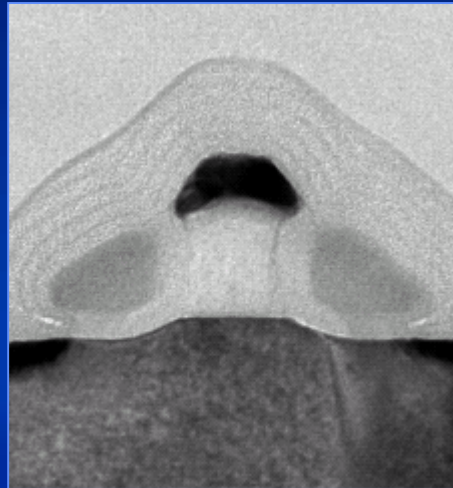


Planar CMOS Transistor Scaling

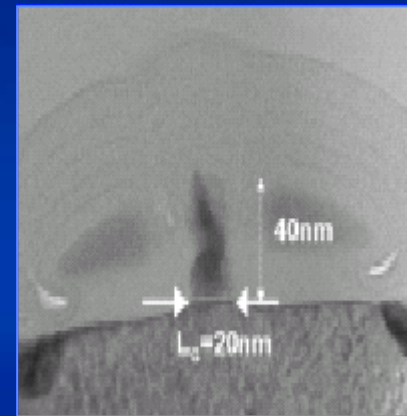
90nm Process
2003



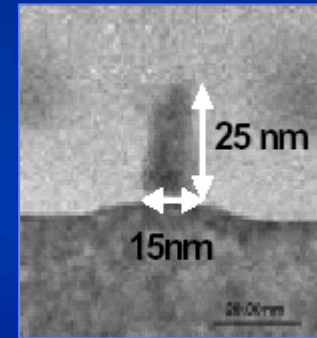
65nm Process
2005



45nm Process
2007



32nm Process
2009



$L_{gate} \rightarrow$ 50nm
Production

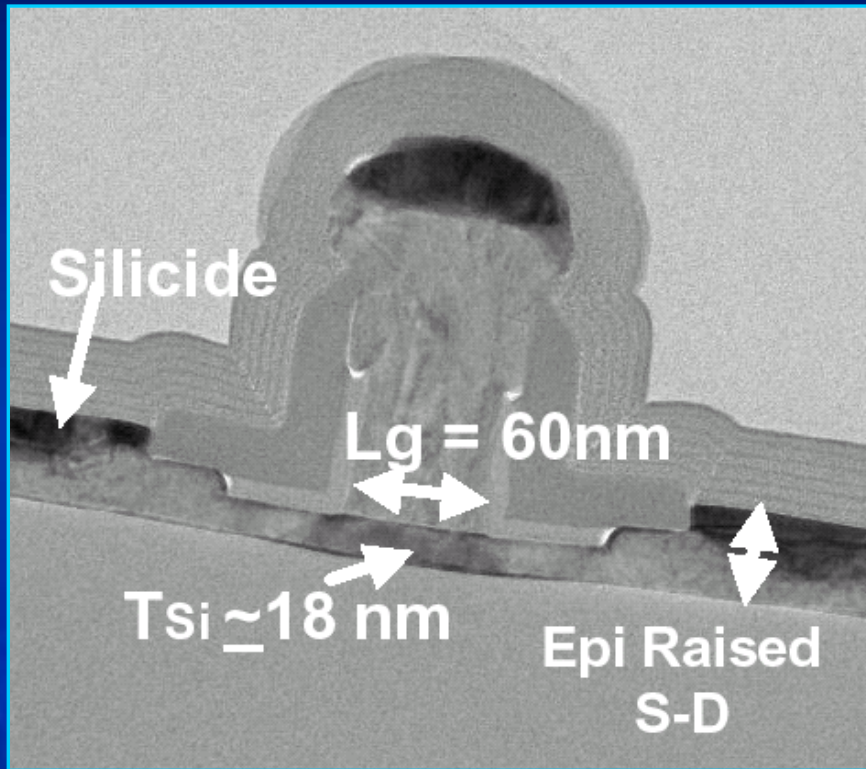
30nm
Prototype

20nm
Prototype

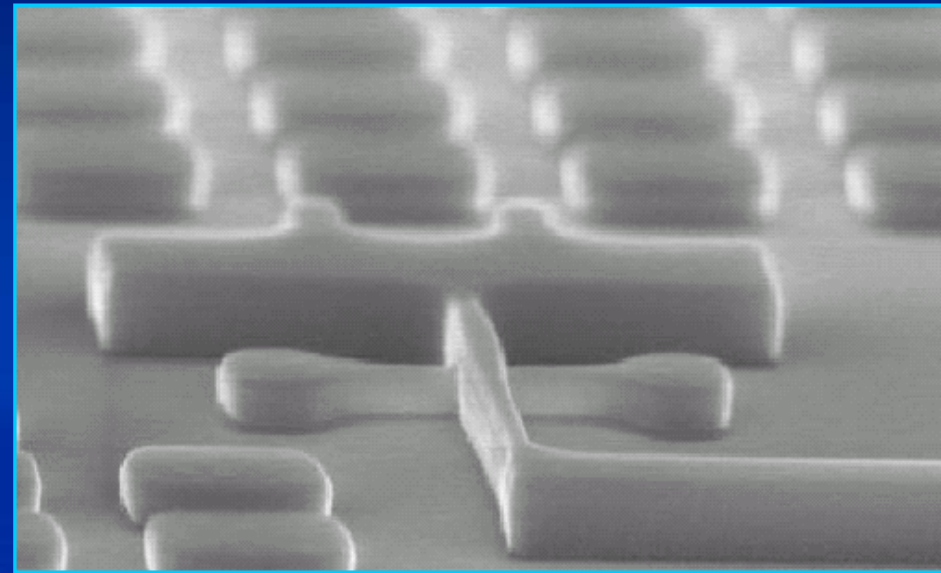
15nm
Prototype

- Intel R&D groups are exploring aggressive scaling of conventional planar CMOS transistors

Depleted Substrate Transistor

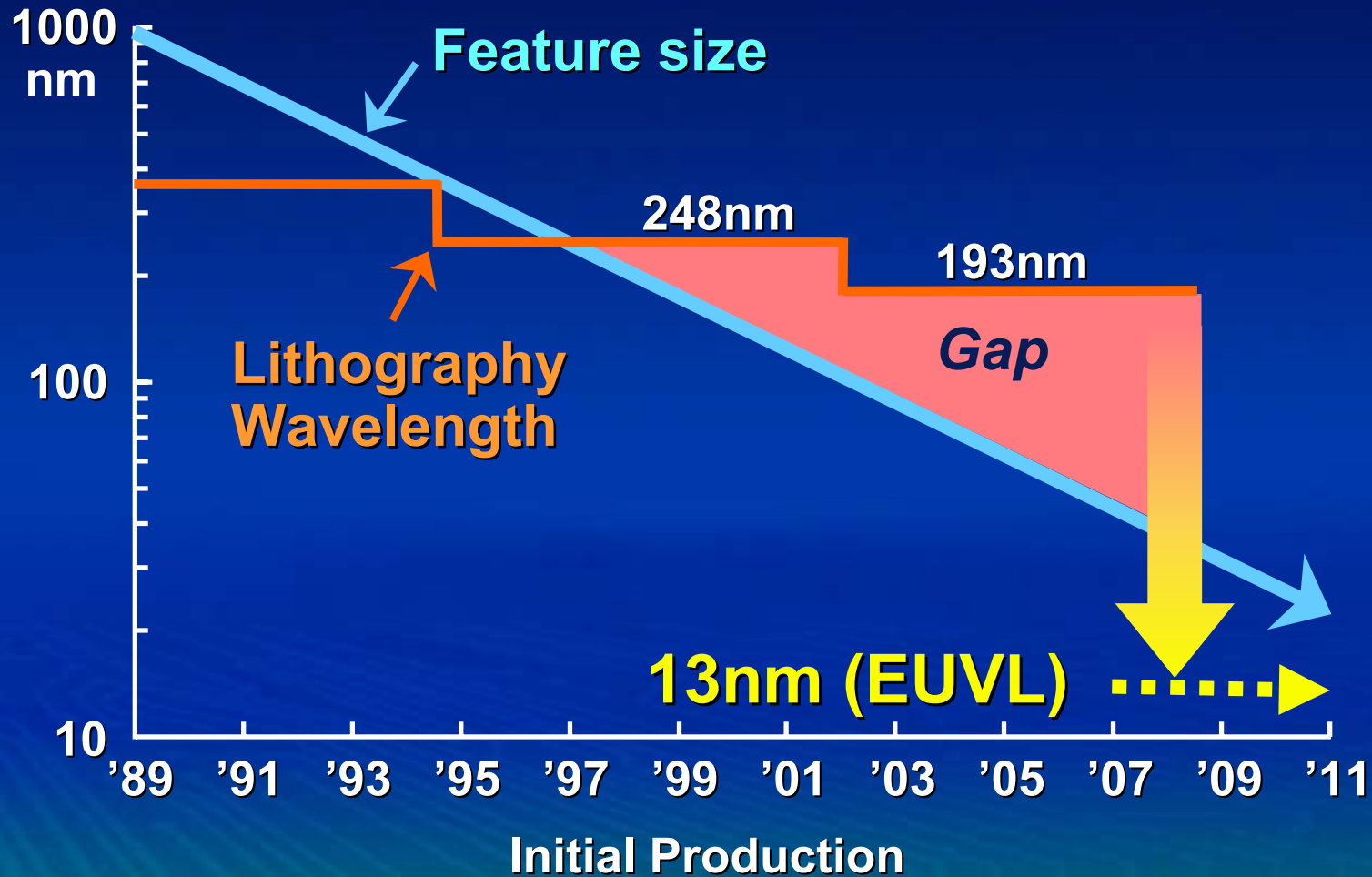


Single-gate DST



Tri-gate DST

Lithography Challenges

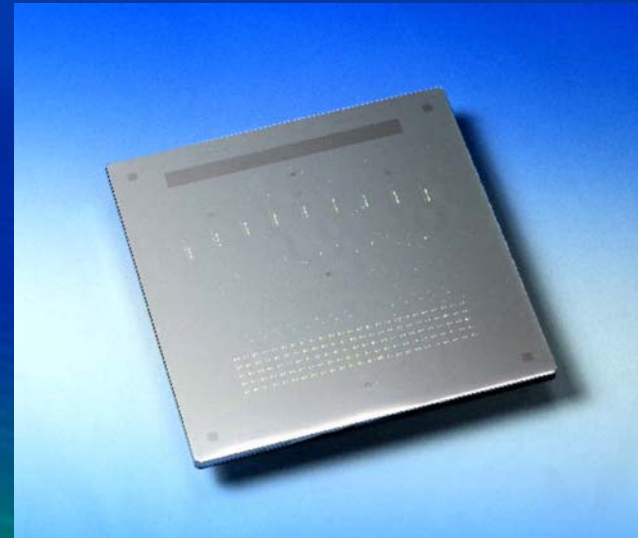


Extreme Ultraviolet Lithography

- EUV lithography uses extremely short wavelength light (20x shorter than today's lithography processes)
 - *Visible light* – 400 to 700 nm
 - *DUV lithography* – 193 and 248 nm
 - *EUV lithography* – 13 nm

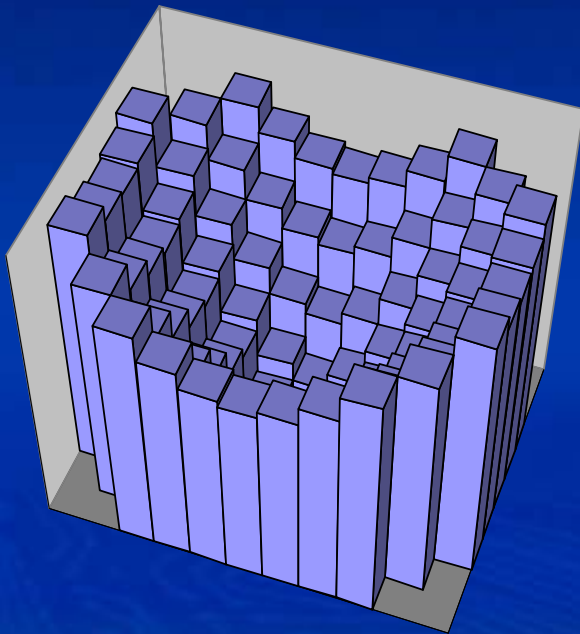


World's First 6-inch EUV ETS Mask



Process Fluctuations

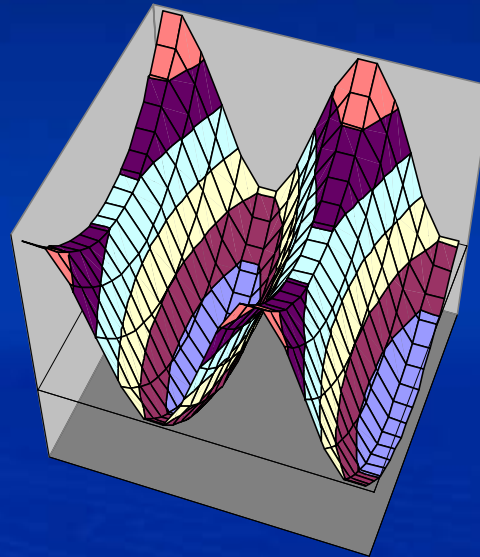
Die-to-Die Fluctuations



Resist Thickness

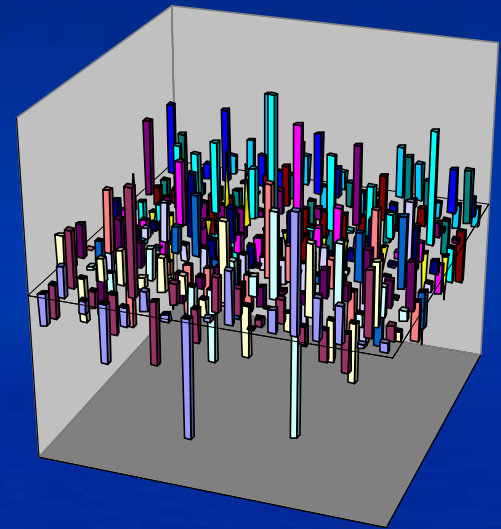
Within-Die Fluctuations

Systematic



Lens Aberrations

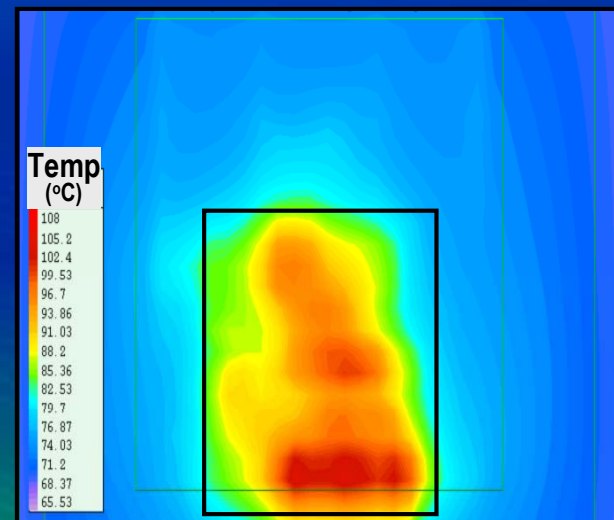
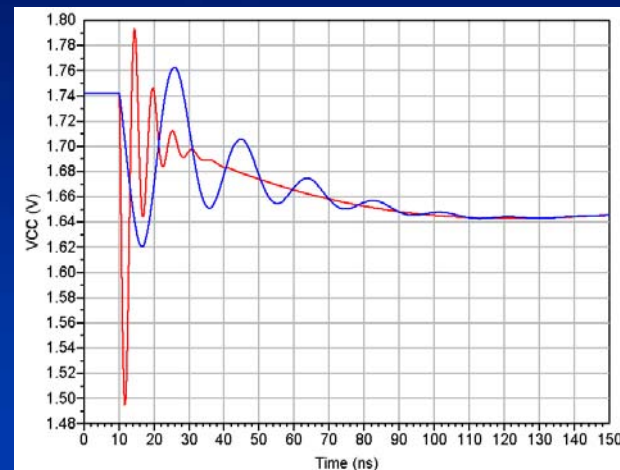
Random



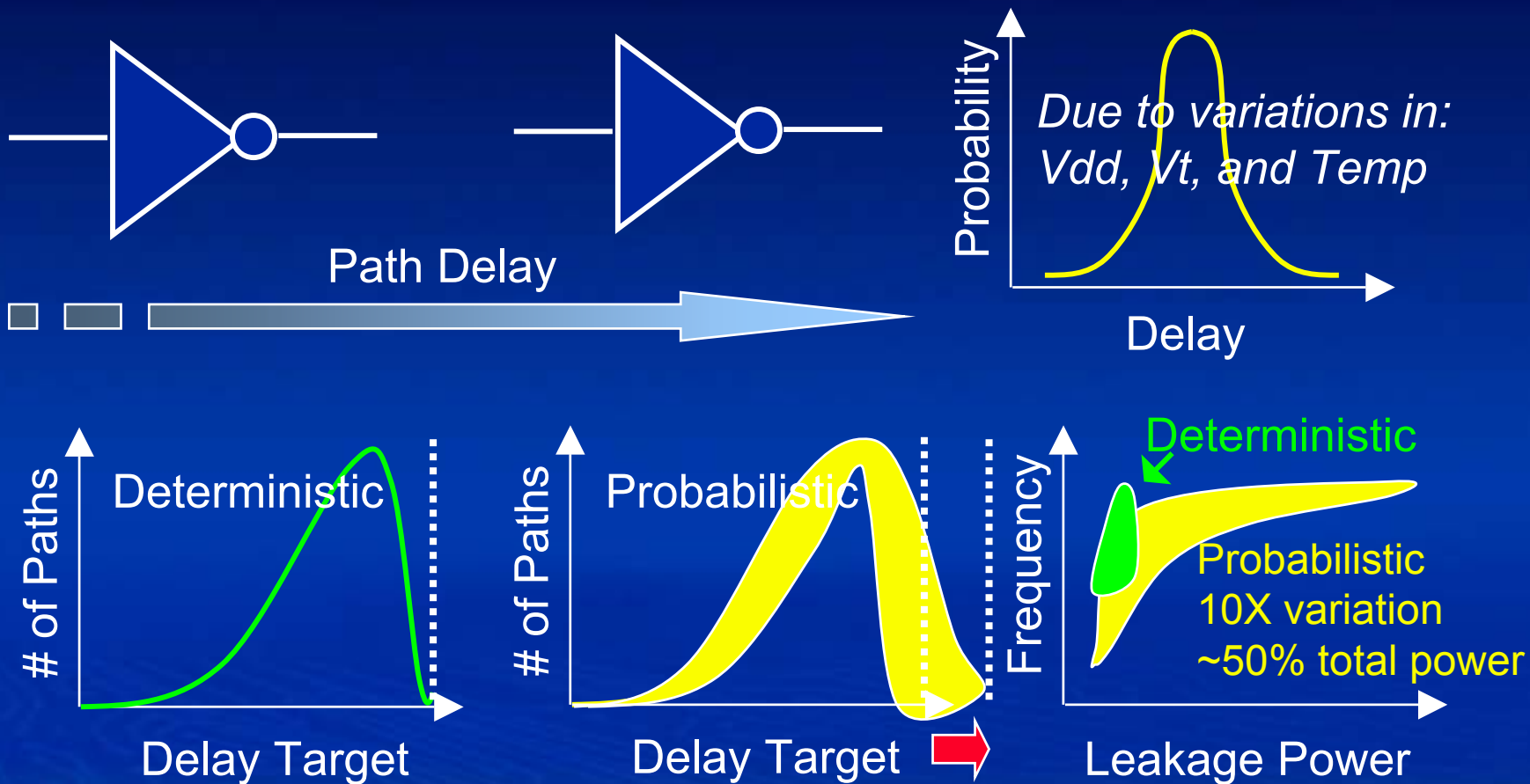
Random Placement
of Dopant Atoms

P, V, T Variations

- **P**rocess
 - Die-to-die variation
 - Within-die variation
 - Static for each die
- **V**oltage
 - Chip activity change
 - Current delivery—RLC
 - Dynamic: ns to 10-100us
 - Within-die variation
- **T**emperature
 - Activity & ambient change
 - Dynamic: 100-1000us
 - Within-die variation

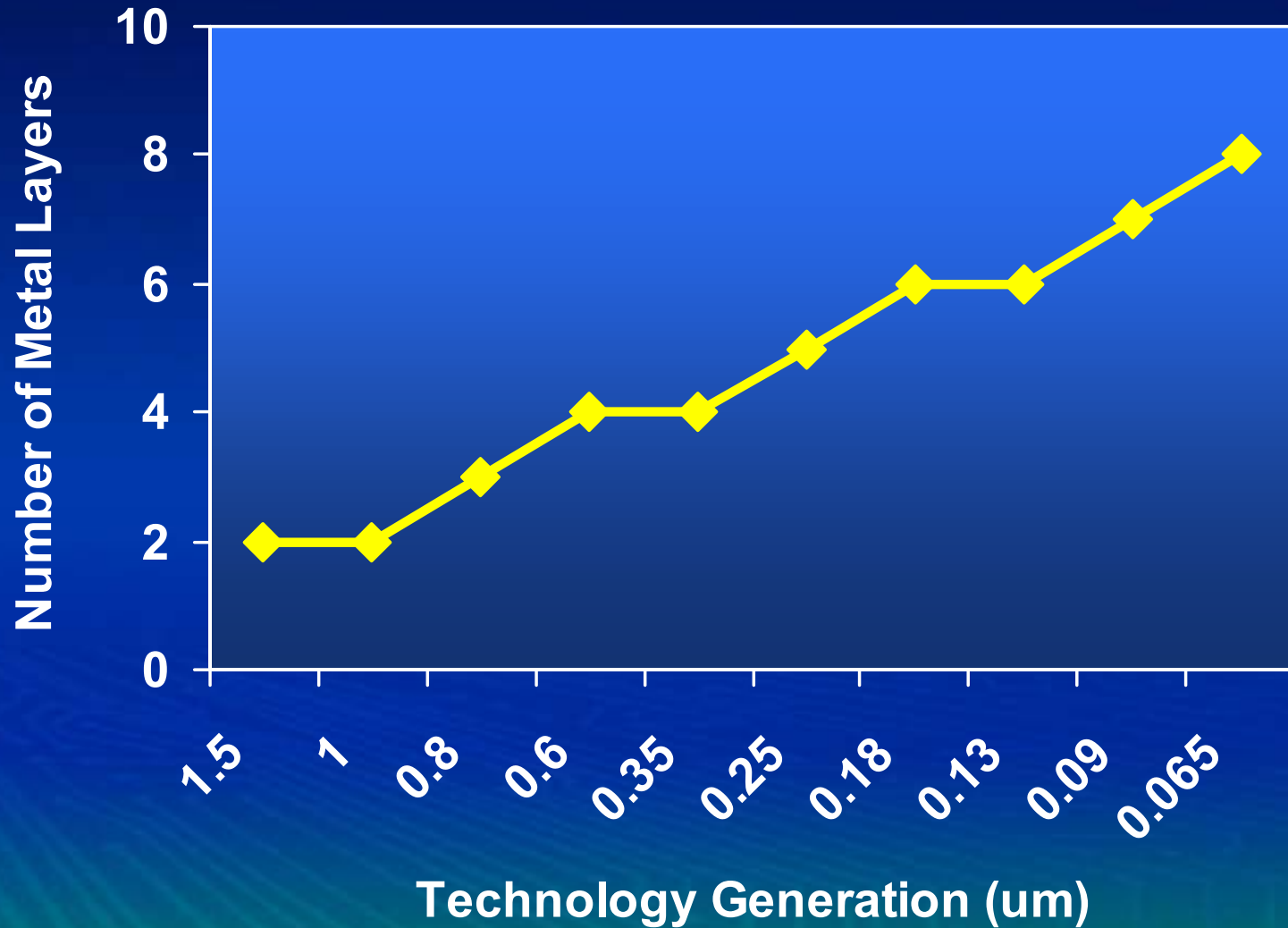


Impact on Design Methodology



- Major paradigm shift from deterministic design to probabilistic / statistical design

Metal Layers



90nm Generation Interconnects

M7

M6

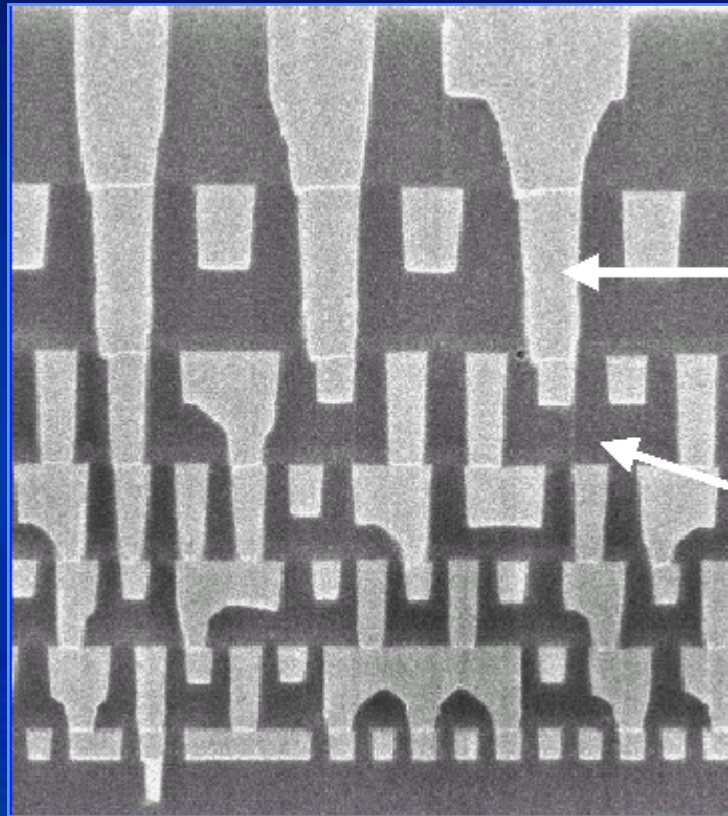
M5

M4

M3

M2

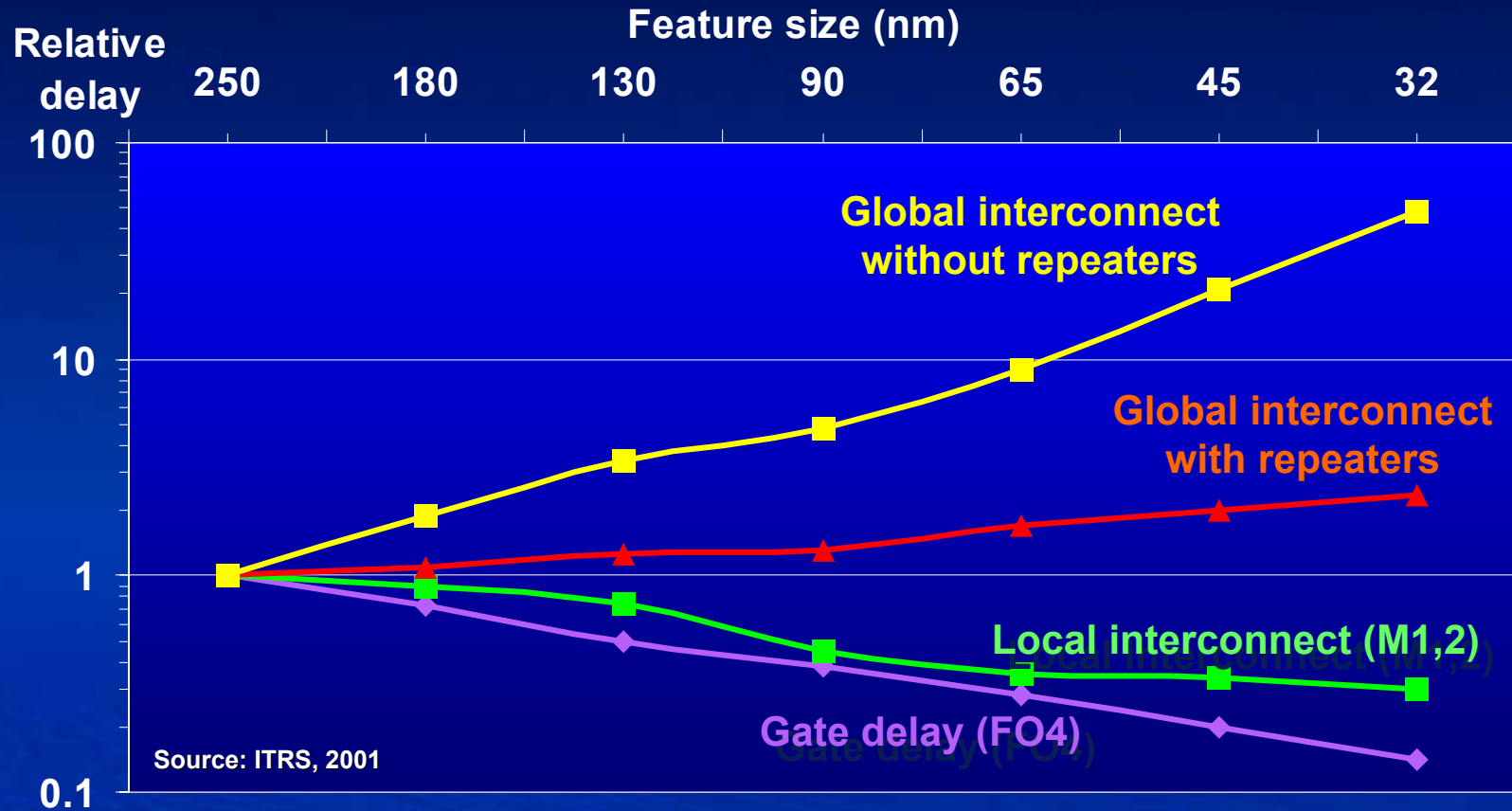
M1



Copper
Interconnects

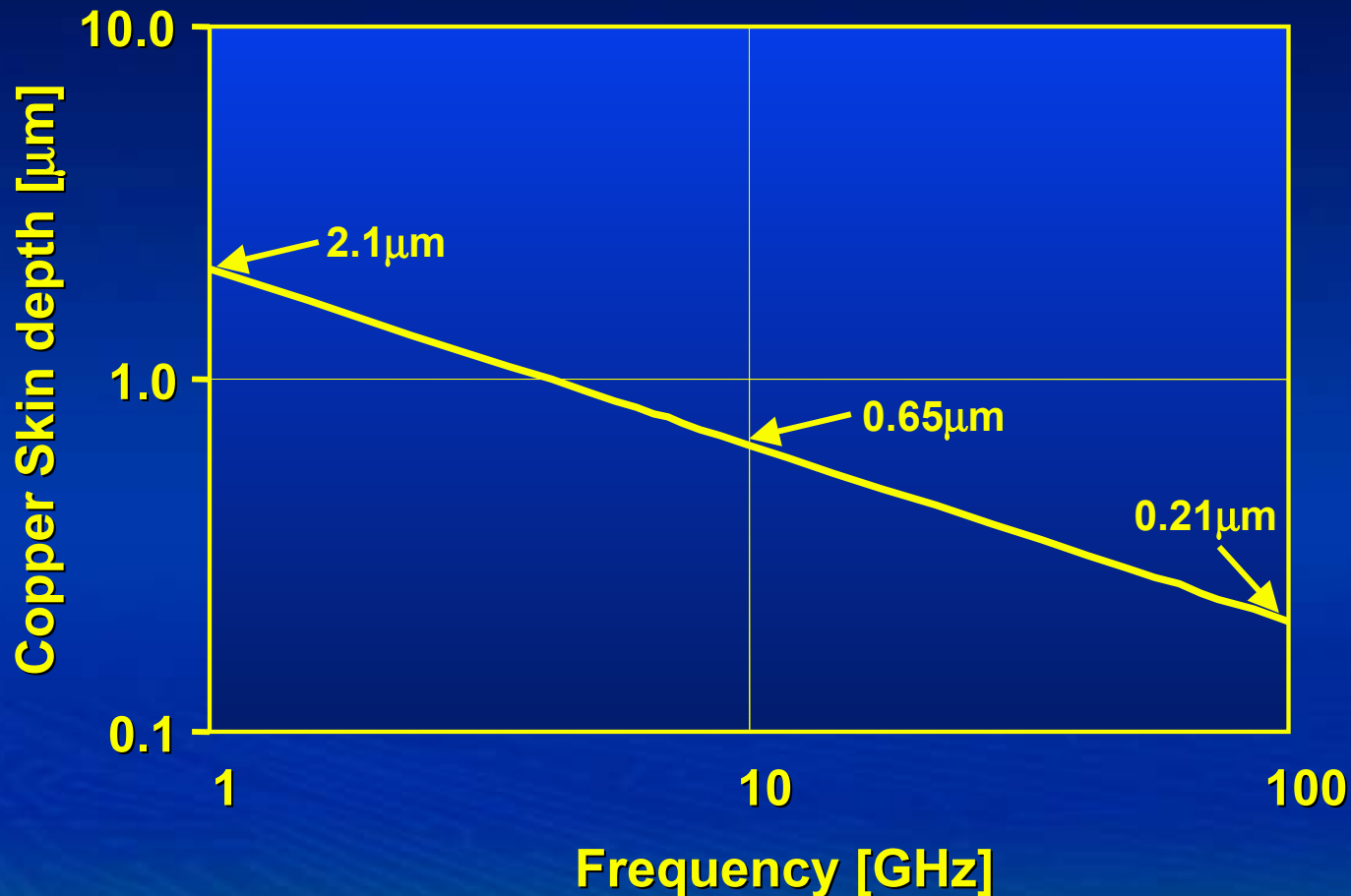
Low-k CDO
Dielectric

On-chip Interconnect Trend



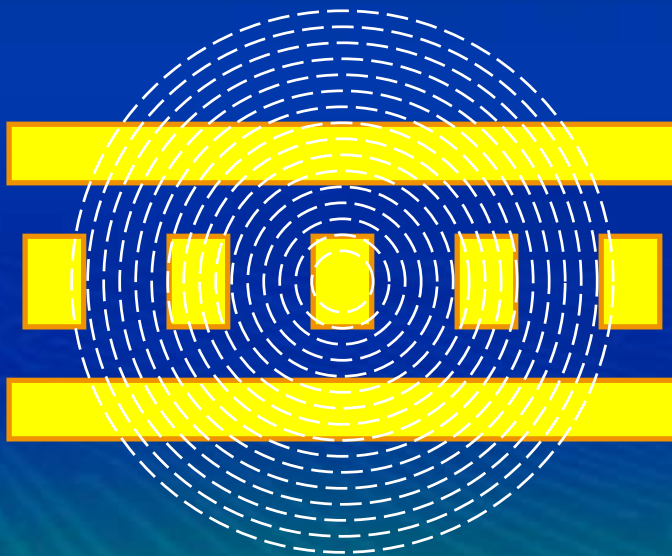
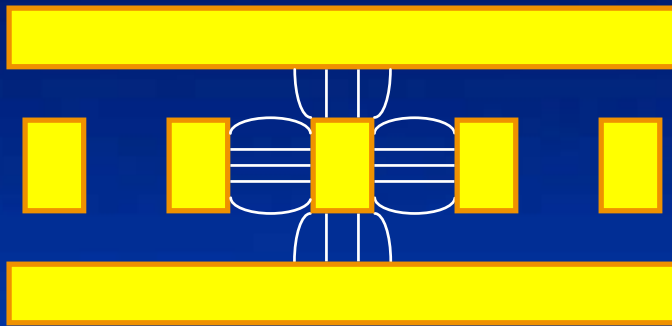
- Local interconnects scale with gate delay
- Intermediate interconnects benefit from low-k material
- Global interconnects do not scale

Skin Effect



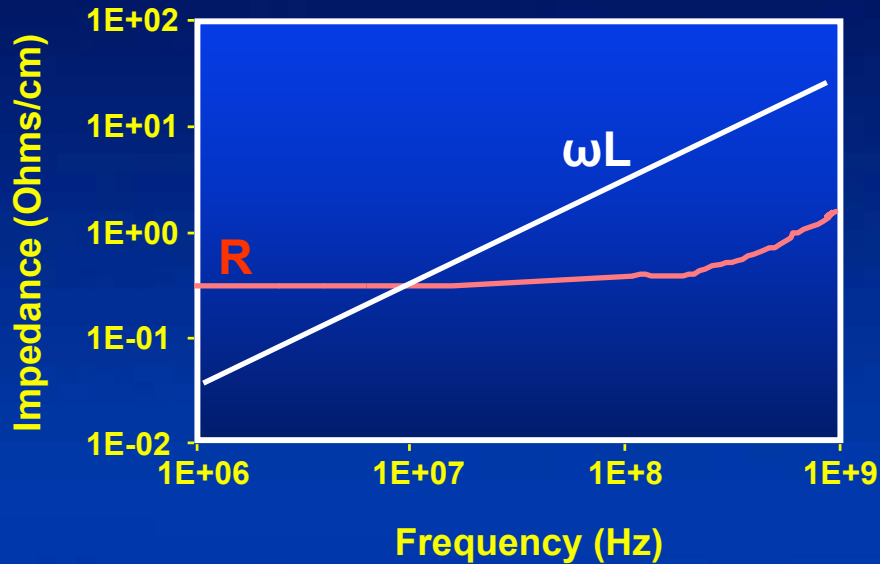
- Edge frequency is 5-9x the clock frequency

Capacitive vs. Inductive Coupling

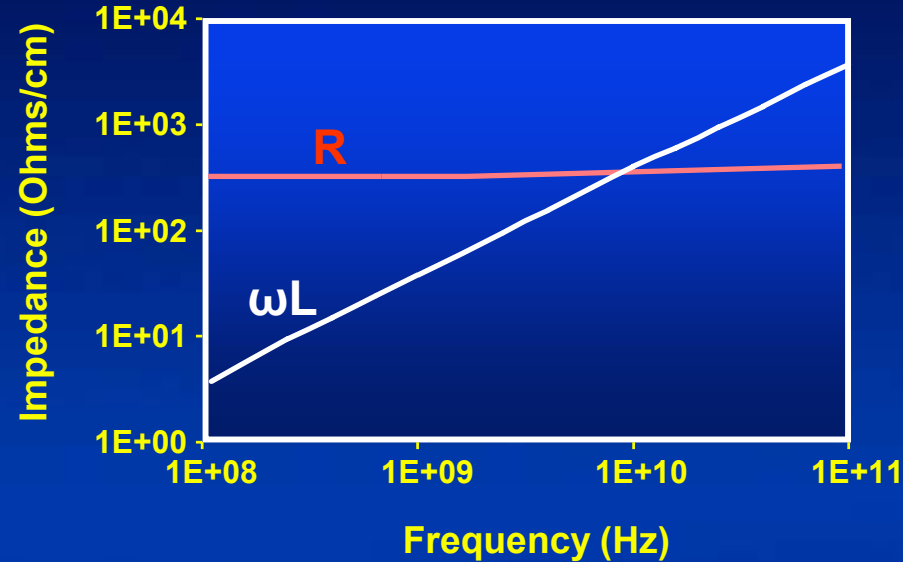


- **Capacitive Coupling**
 - Due to electric field
 - “Near” field effect
 - Measures resistance to a voltage change
- **Inductive Coupling**
 - Due to magnetic field
 - “Far” field effect
 - Measures resistance to a current change
 - Frequency dependent

Inductive Noise



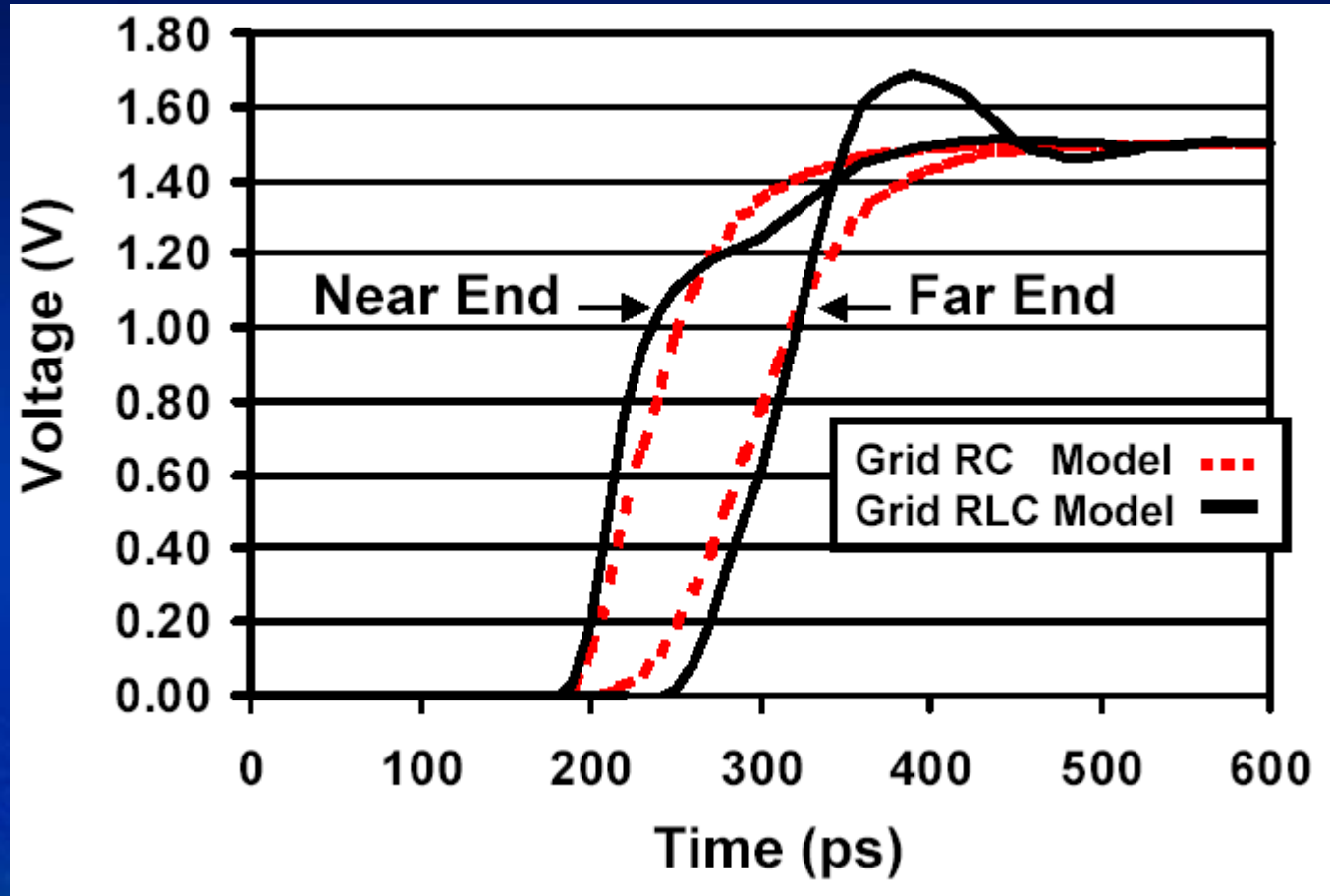
PCB (FR4) Signal Trace



VLSI Metal Line

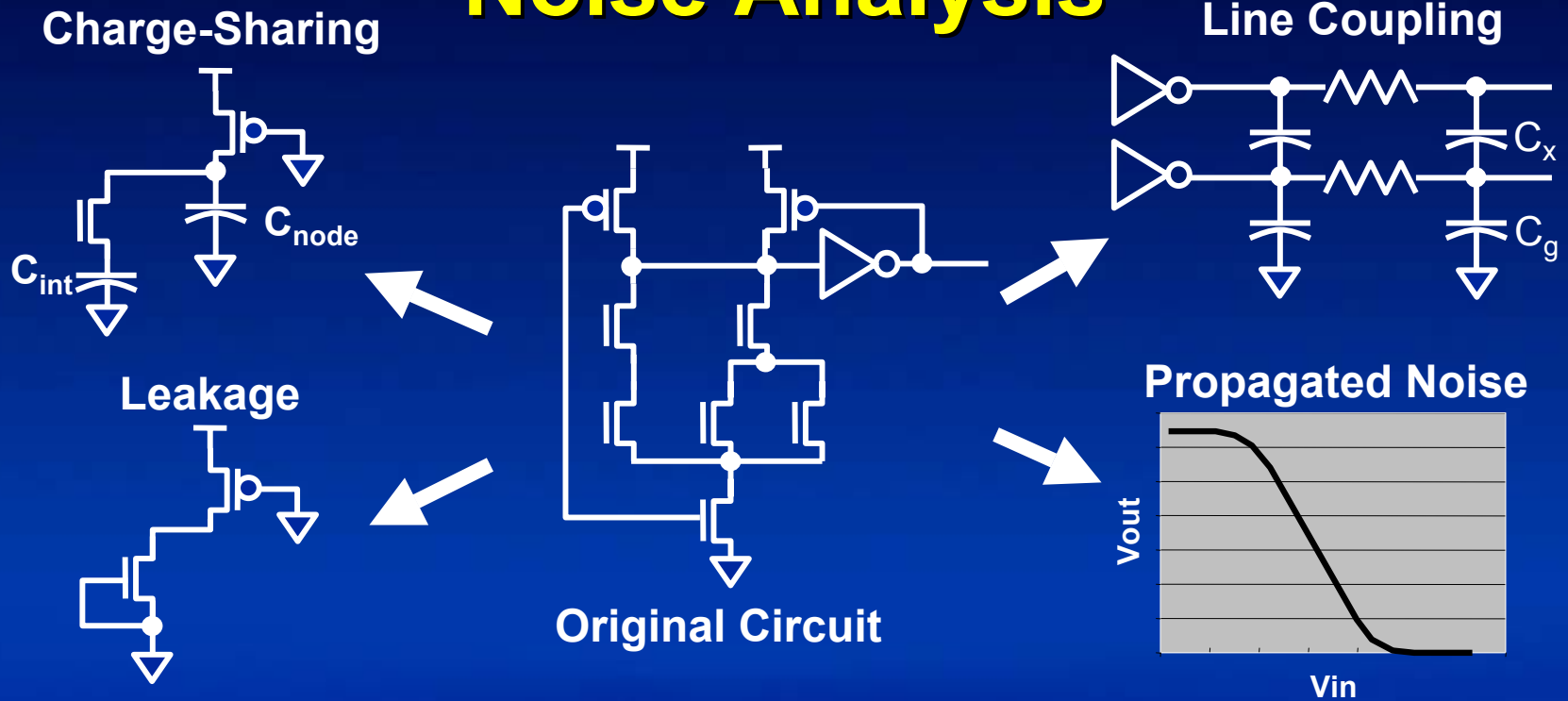
- Inductance of VLSI metal lines is becoming important at operating frequencies above 1GHz
- Need accurate R,L,C extraction tools

Inductance Effect



Xanthopoulos, ISSCC-2001

Noise Analysis



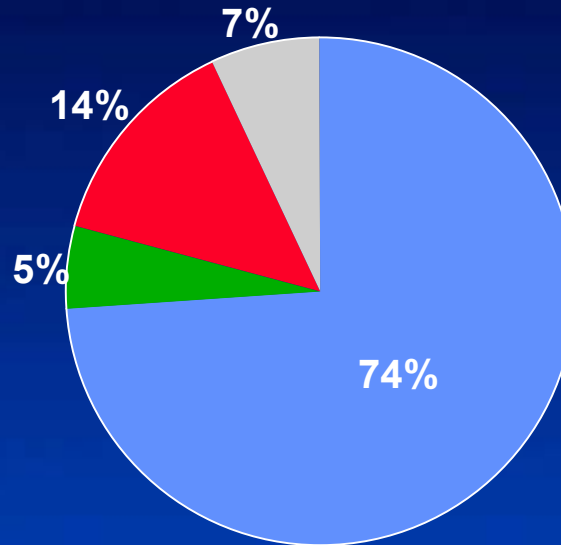
- Each circuit broken into equivalent models for various noise sources
 - Models charge-sharing, coupling, leakage, supply noise, contention
 - Calculated noise propagated to next stage to model amplification
- Macro-block results rolled up to fullchip analysis
 - Abstracted noise from block level are collected for the full chip and combined with coupling analysis

Agenda

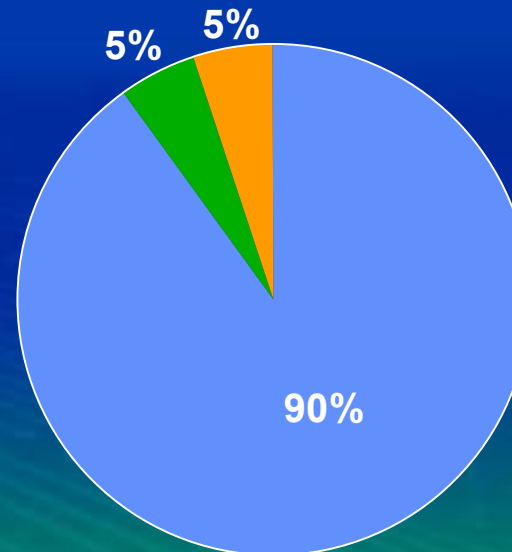
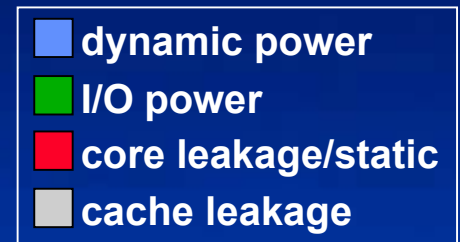
- Microprocessor Design Trends
- Process Technology Directions
- **Active Power Management**
- Leakage Reduction Techniques
- Thermal Modeling
- Call to Action
- Summary

Itanium[®] 2 Processor Power Charts

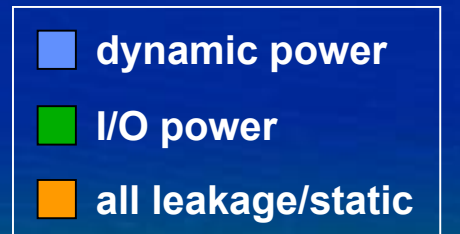
- Maintain the same 130W power envelope
 - 50% frequency increase
 - 2X larger L3 cache
 - Leakage increased 3.5X
- Aggressive management of dynamic power
 - Reduced clock loading
 - Reduced contention power
 - L3 cache power management



130nm Itanium[®] 2 Processor 6M 1.5GHz



180nm Itanium[®] 2 Processor 3M 1.0GHz



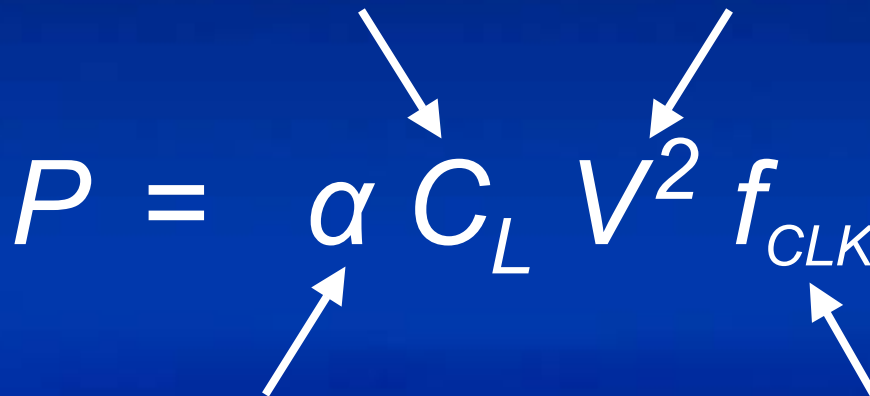
Active Power Reduction

Reduce switched capacitance:

- Minimize loading from diffusion, wire, gate
- Use more efficient layout techniques

Technology scaling:

- Dynamic voltage scaling
- Supply voltage scaling is slowing down
- Thresholds don't scale

$$P = \alpha C_L V^2 f_{CLK}$$


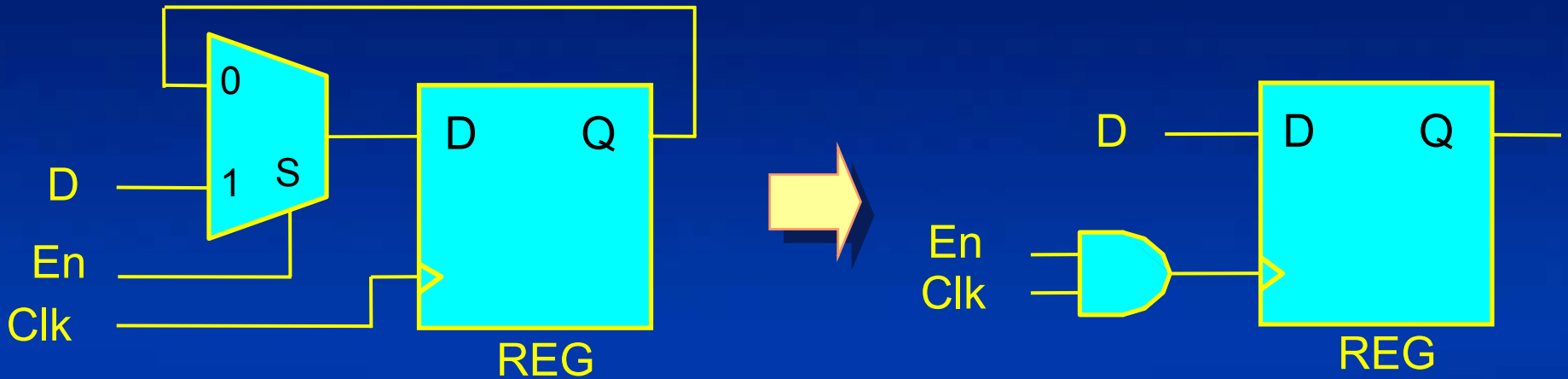
Reduce switching activity:

- Conditional execution
- Conditional clocking
- Conditional precharge
- Turn off inactive blocks

Reduce clock frequency:

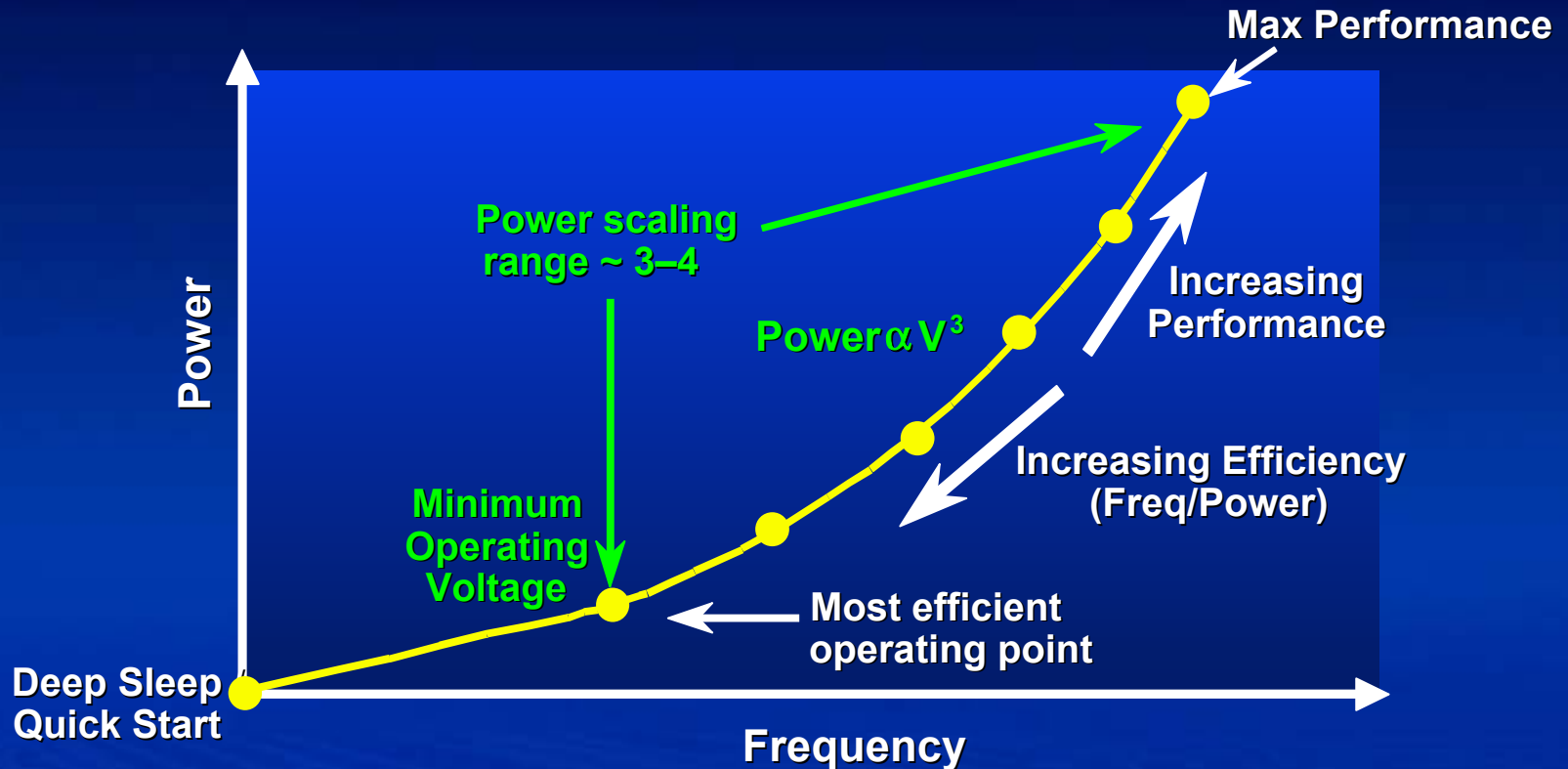
- Use parallelism
- Less pipeline stages
- Use double-edge flip-flops

Clock Gating



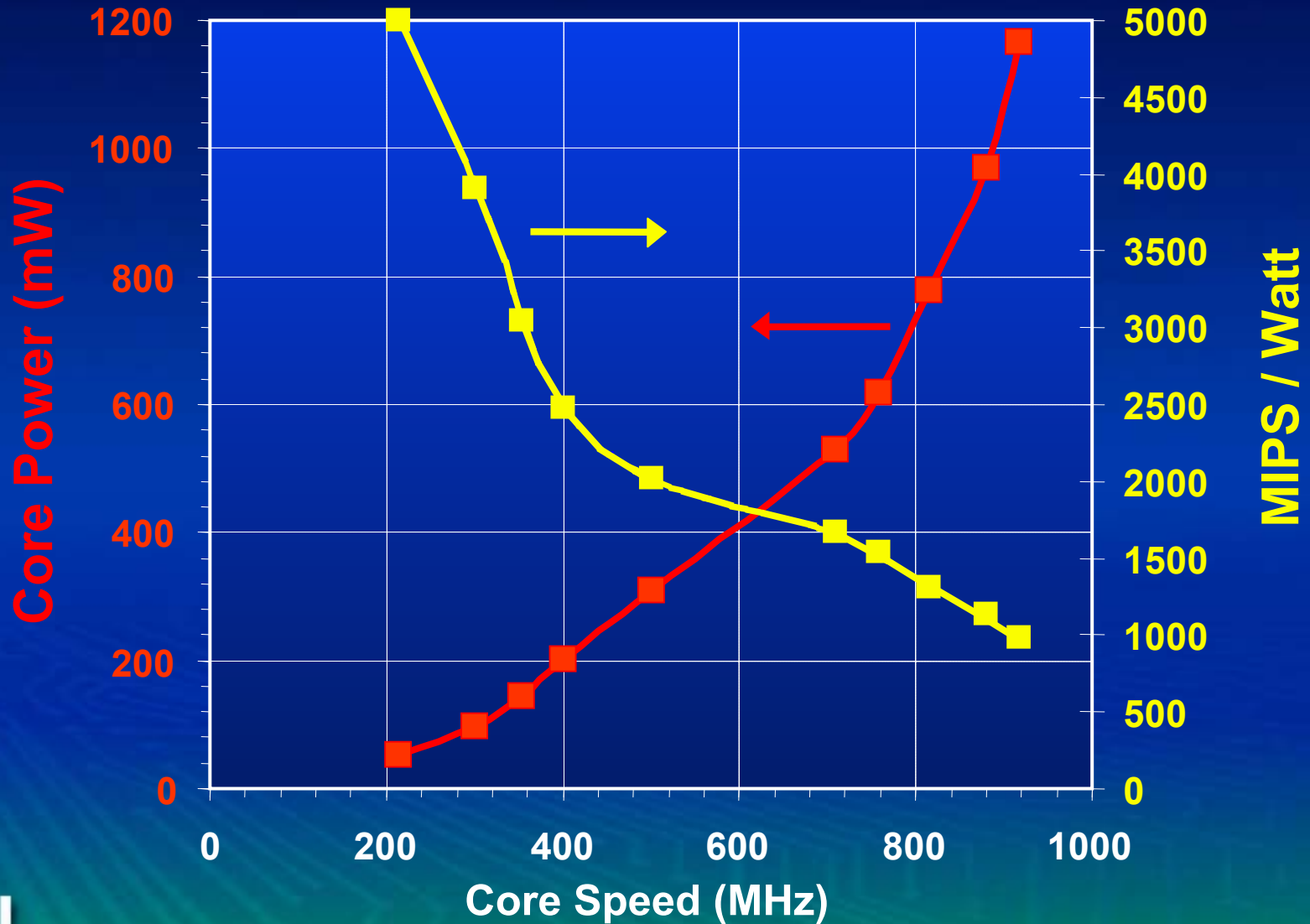
- Save power by gating the clock when data activity is low
- Requires detailed logic validation

Active Power Management

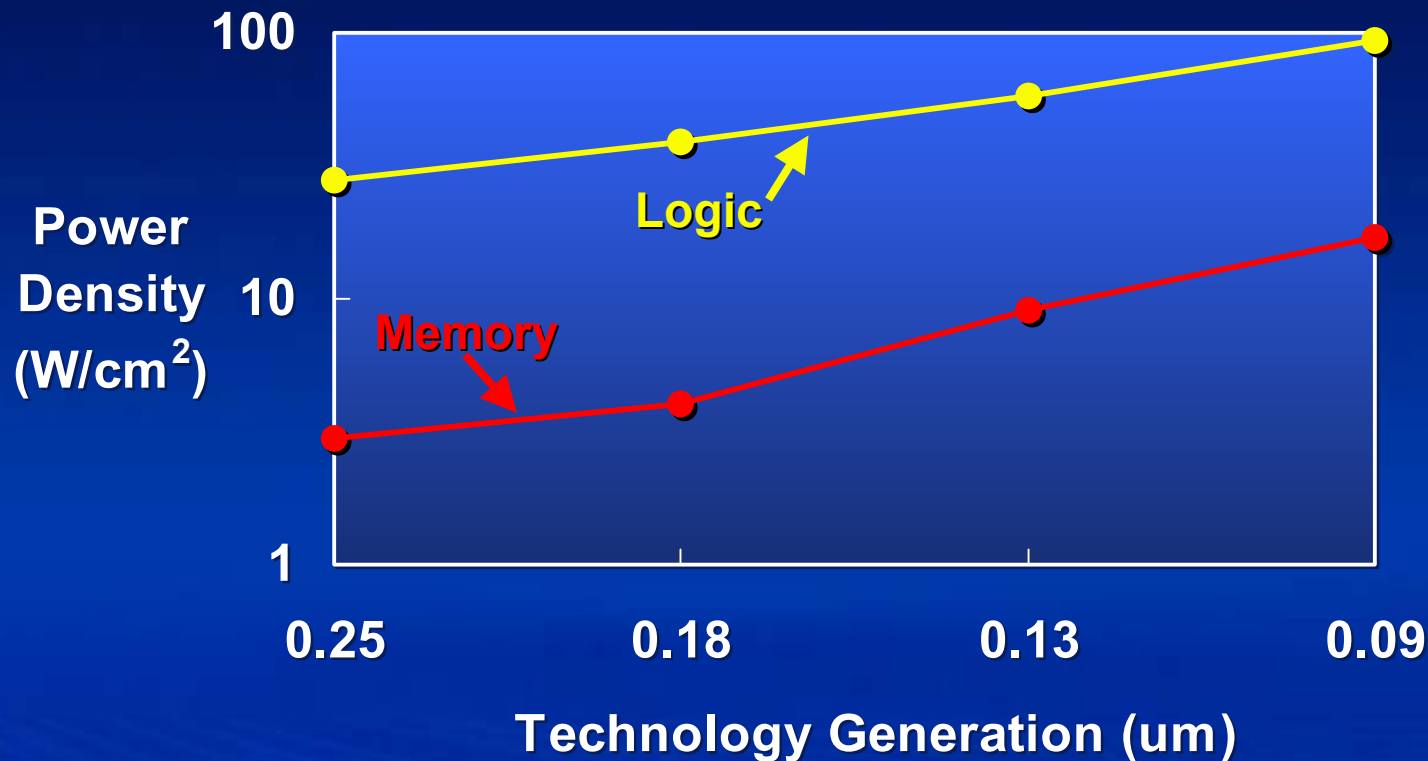


- Voltage-frequency scaling with active thermal feedback
- Multi-operating states from high performance to deep sleep
- Power management reduces average and peak power

Xscale V/F Adjustment

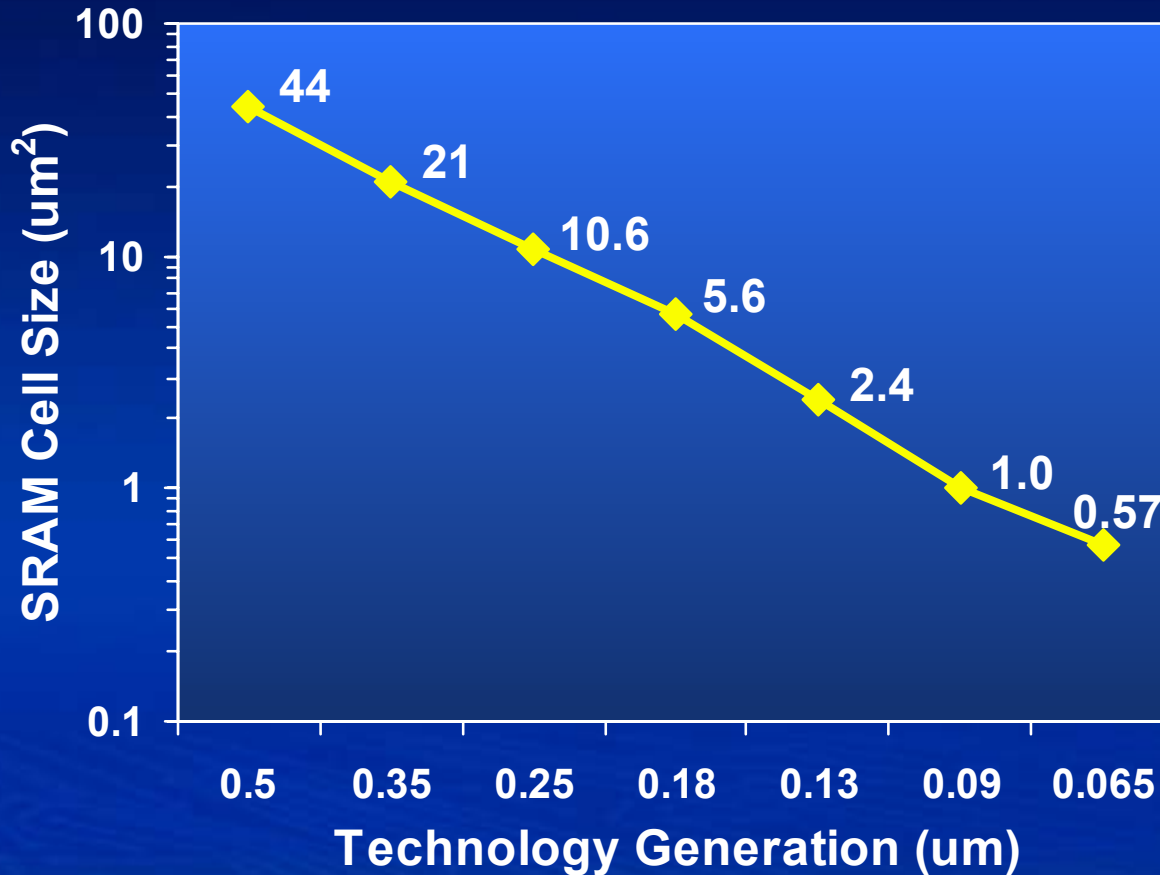


Exploit Memory Power Efficiency



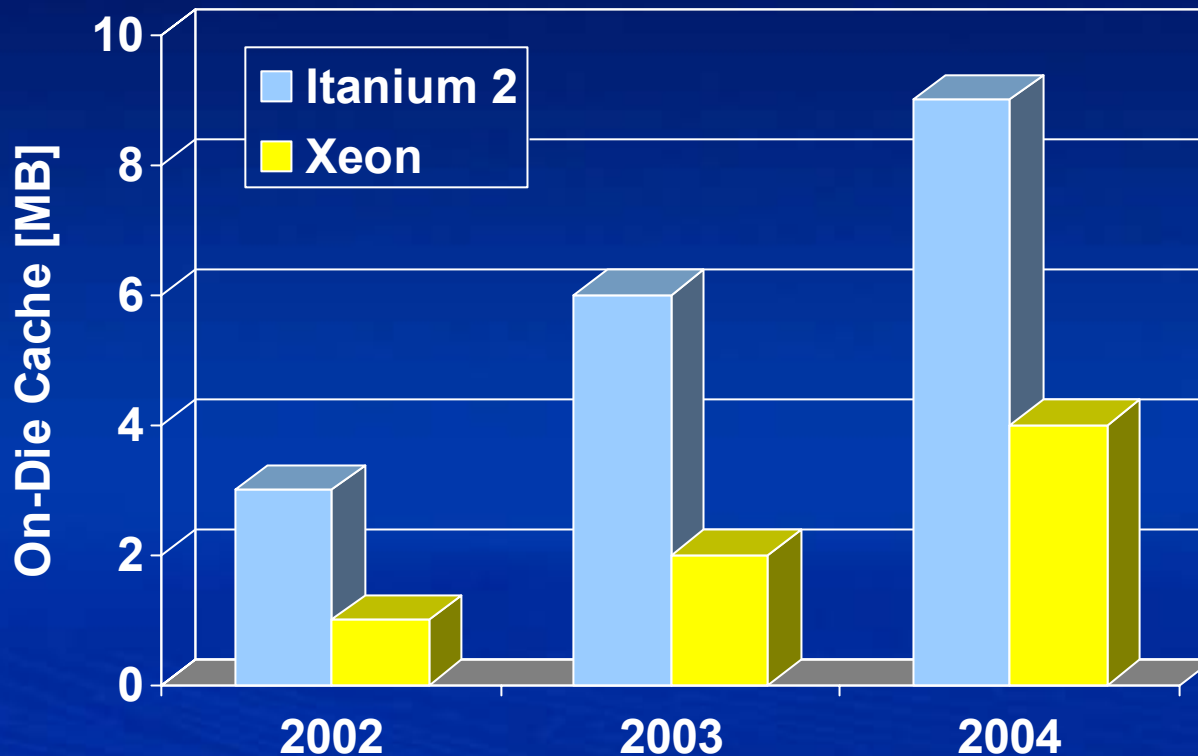
- Static memory has 10X lower active power density
- Lower leakage than logic
- On-die cache provides higher bandwidth and lower latency

SRAM Cell Size Scaling



- SRAM cell size continues to scale ~0.5x per generation
- Larger caches can be incorporated on die

Server Processors On-Die Cache Size Trends

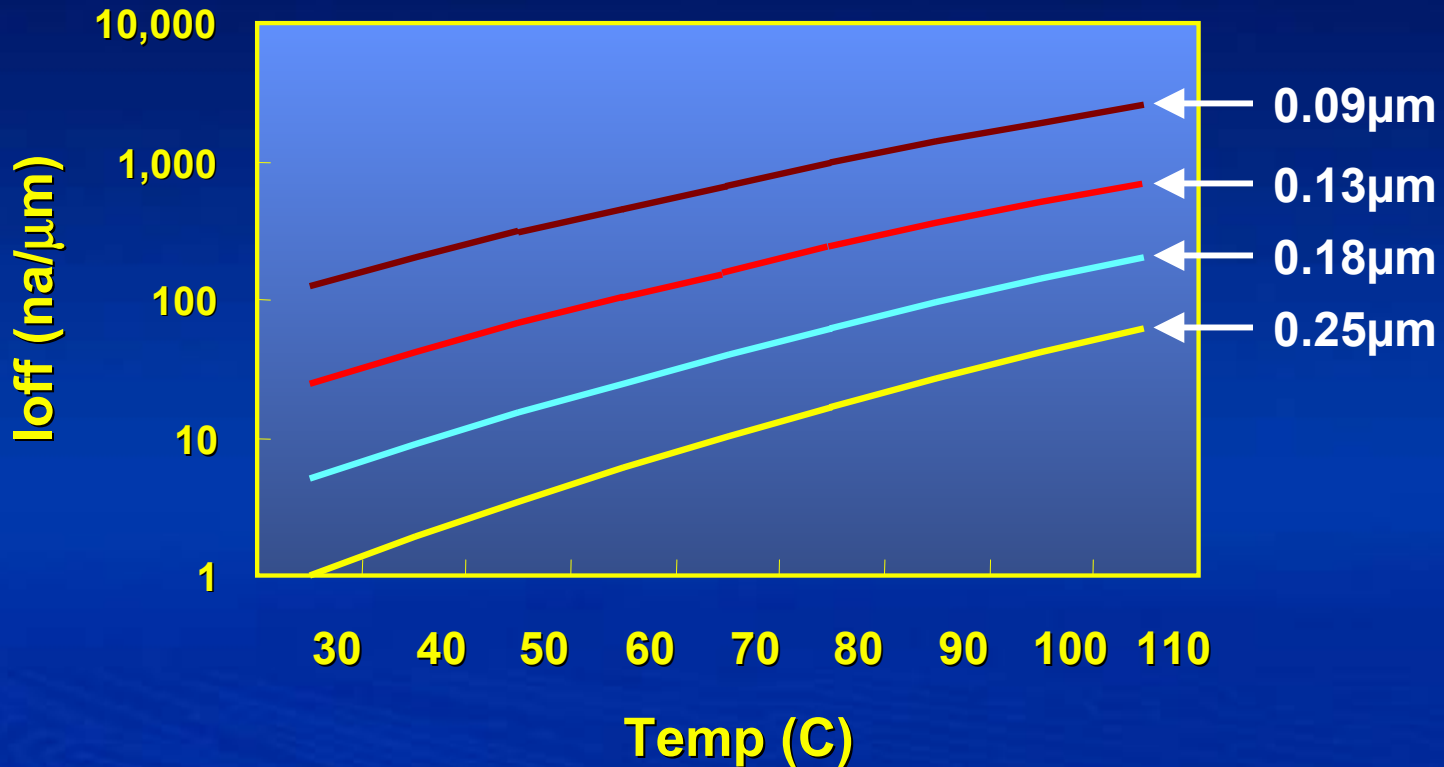


- Increasing cache size is a power efficient way to improve server performance

Agenda

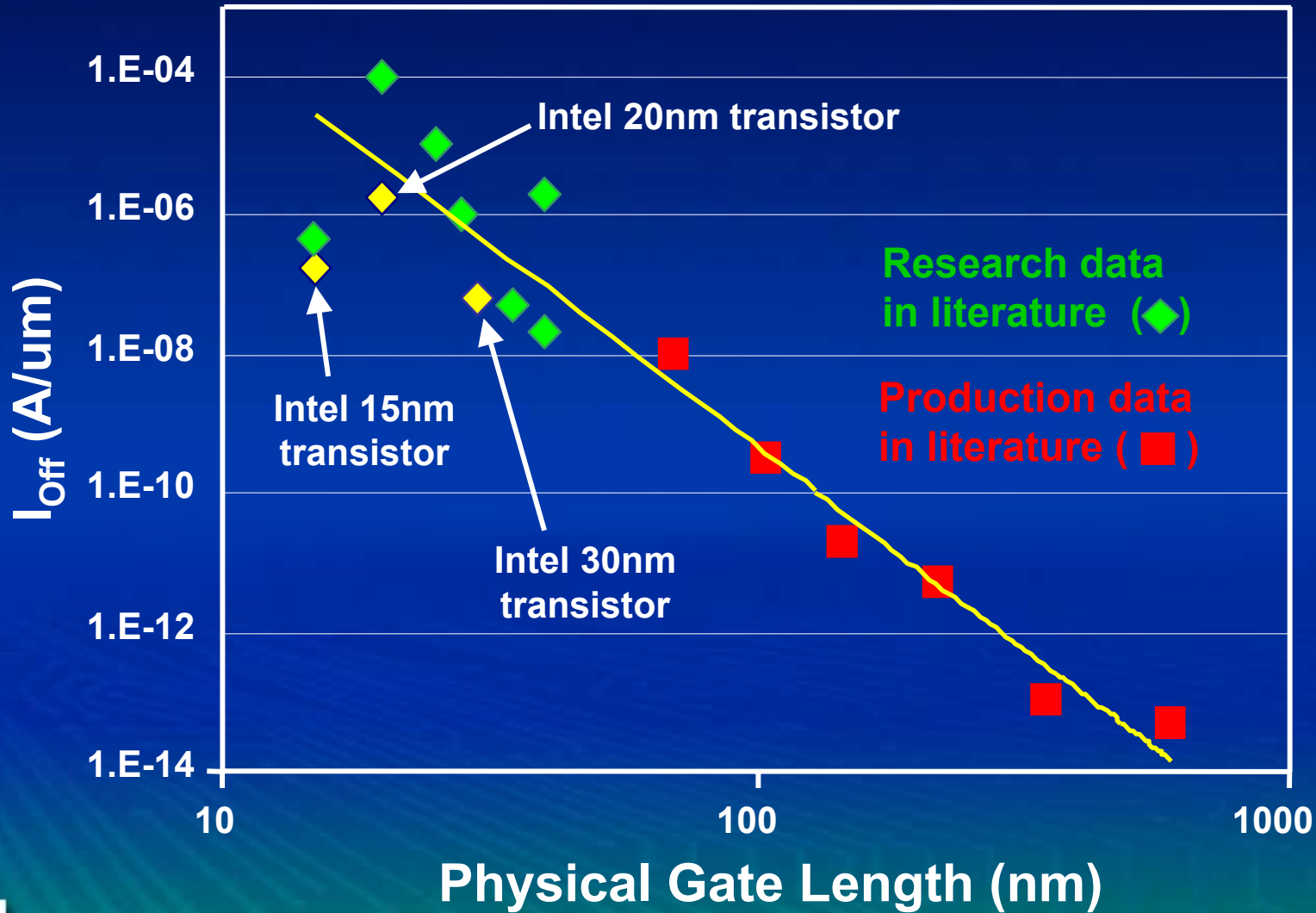
- Microprocessor Design Trends
- Process Technology Directions
- Active Power Management
- Leakage Reduction Techniques
- Thermal Modeling
- Call to Action
- Summary

Leakage Continues to Increase

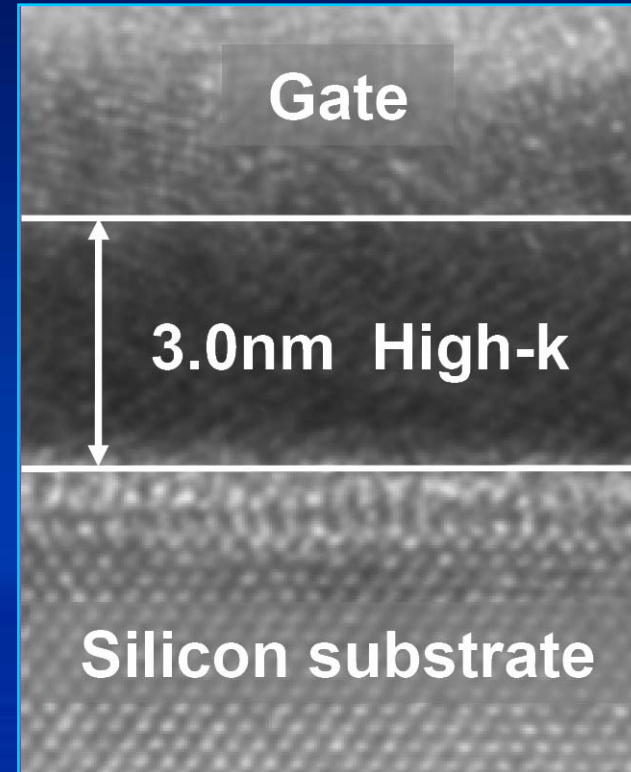
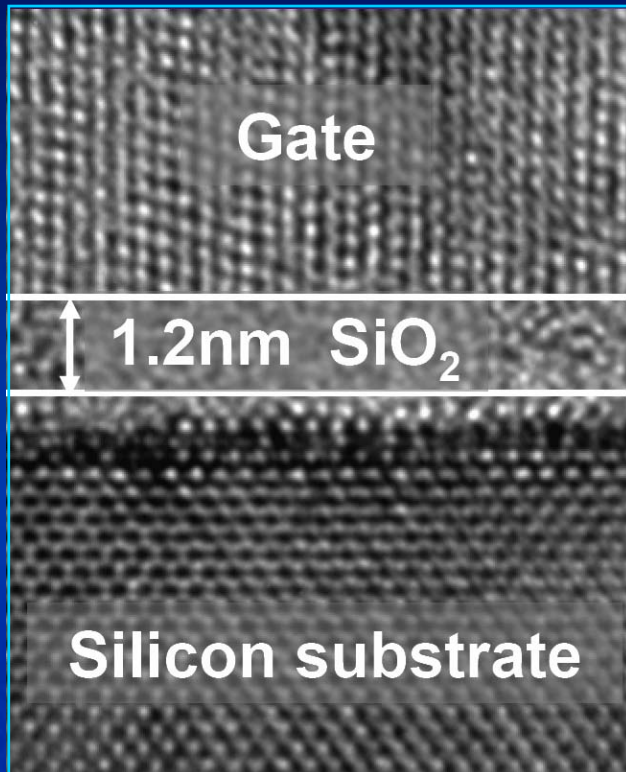


- Design issues:
 - Dynamic circuits may fail
 - Need to guarantee burn-in functionality

Subthreshold Leakage Trend



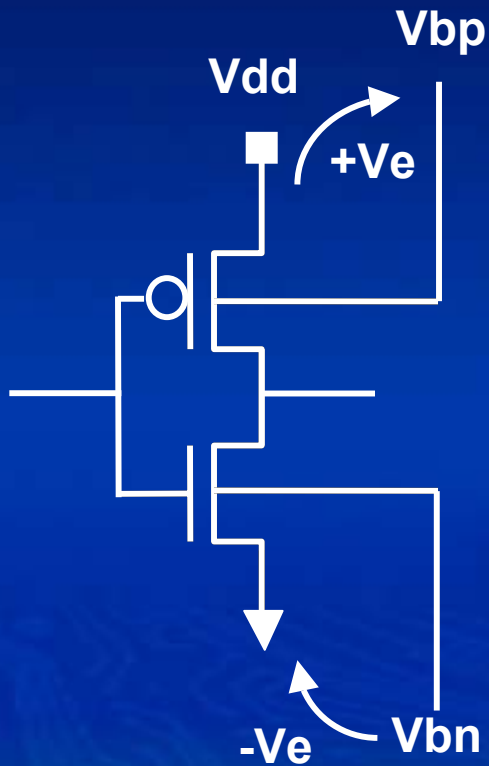
High-K Gate Dielectric



High-K gate dielectric will reduce gate leakage by up to 100x in the 45nm technology node

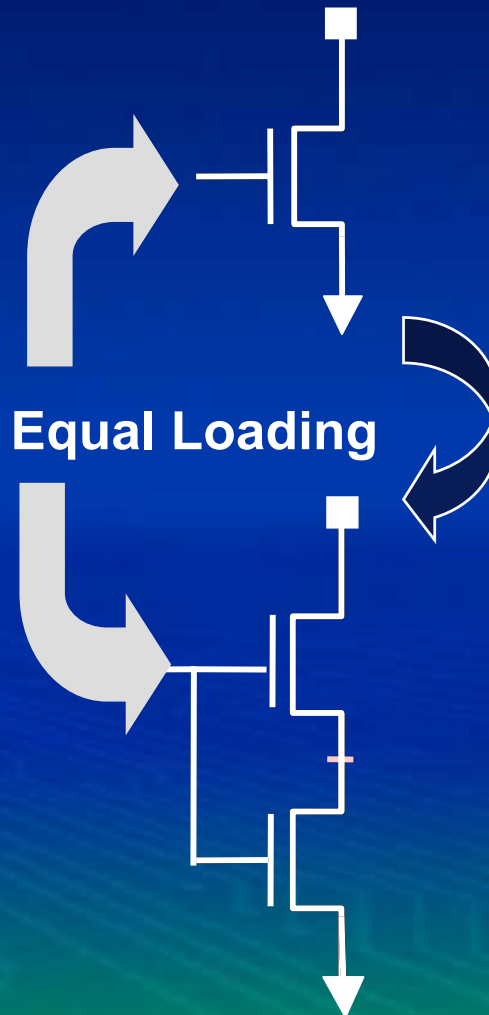
Leakage Reduction Techniques

Body Bias

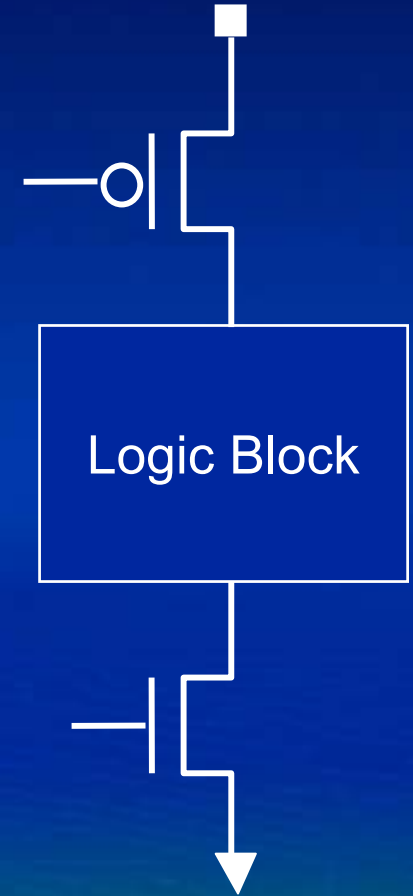


2-10X reduction

Stack Effect

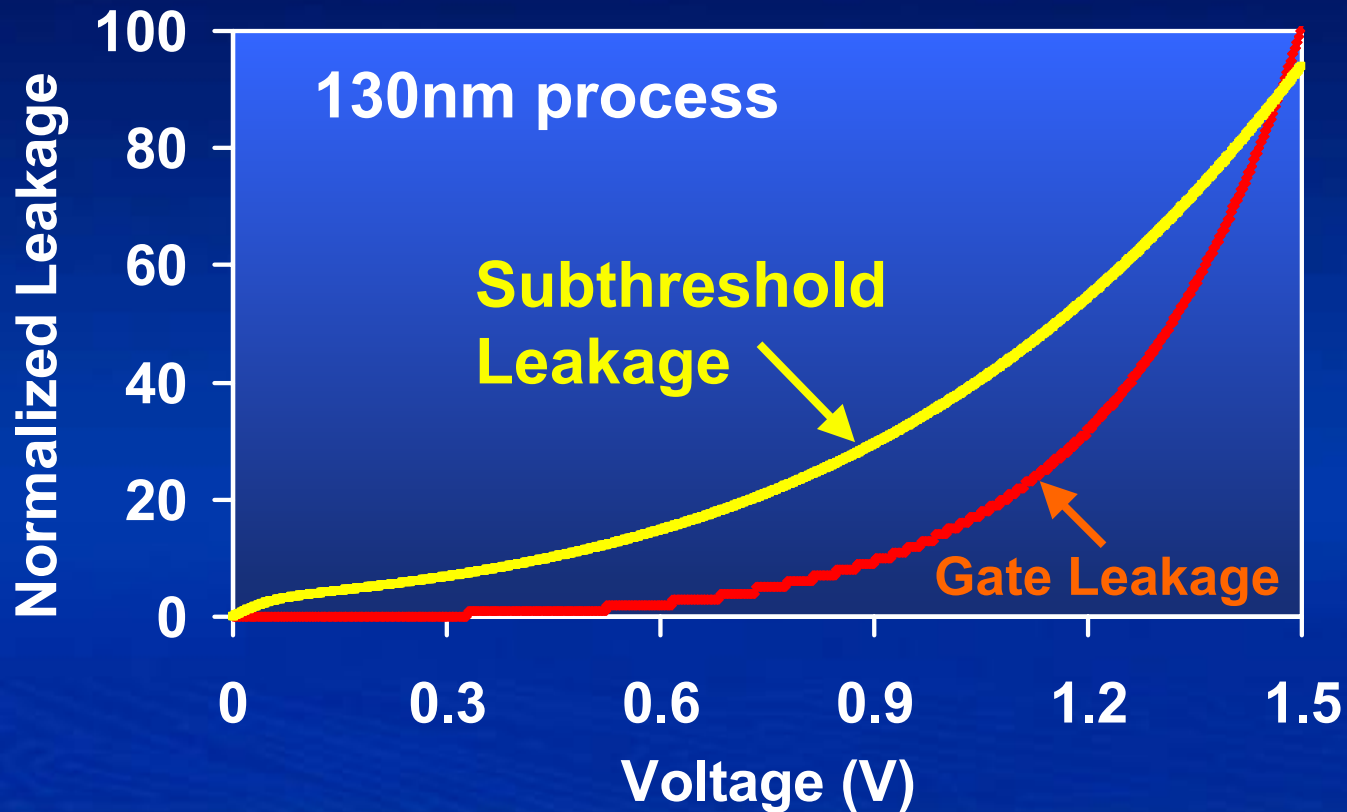


Sleep Transistor



2-1000X reduction

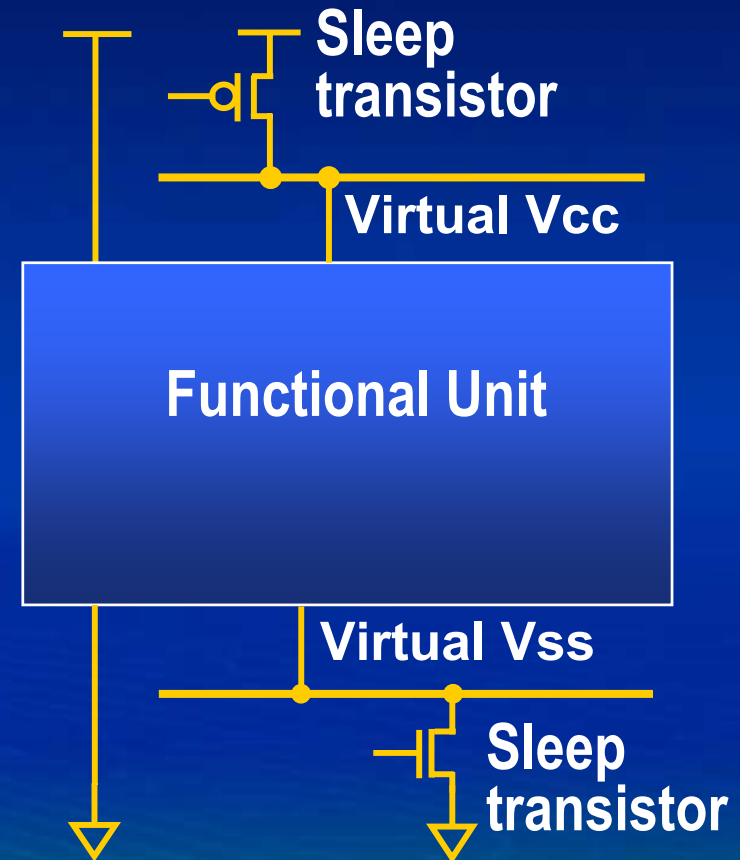
Leakage is a Strong Function of Voltage



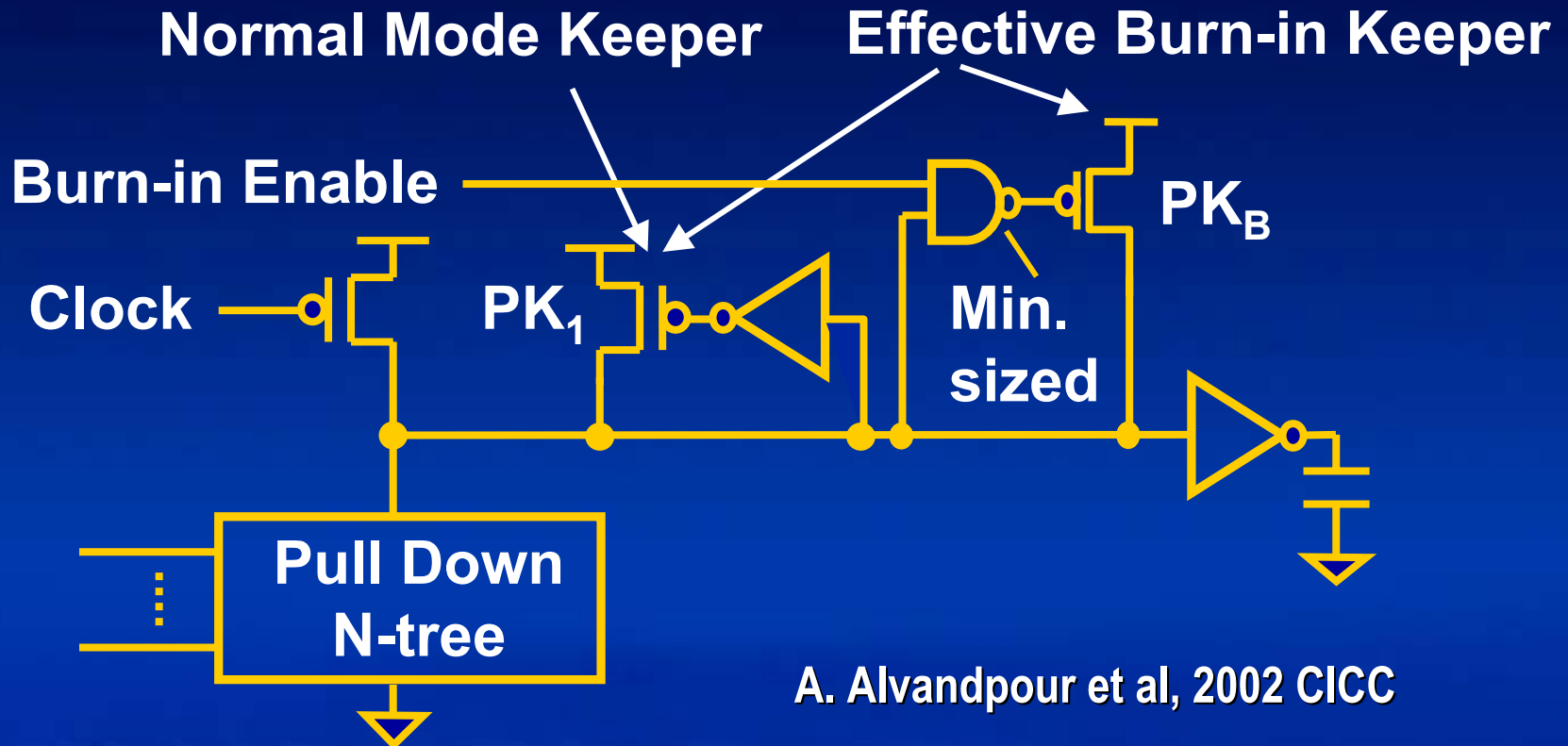
- Subthreshold and gate leakage reduce with lower supply voltage

Standby Leakage Reduction: Sleep Transistor Design

- Motivation: Cut off power supply in sleep-mode
- Insert sleep transistor between main supply and functional unit's supply rails
- Latches tied to main supply rails to retain state
- EDA tools needed to:
 - Size sleep transistor and distribute in layout
 - Model the timing impact



Burn-in Tolerant Dynamic Circuits

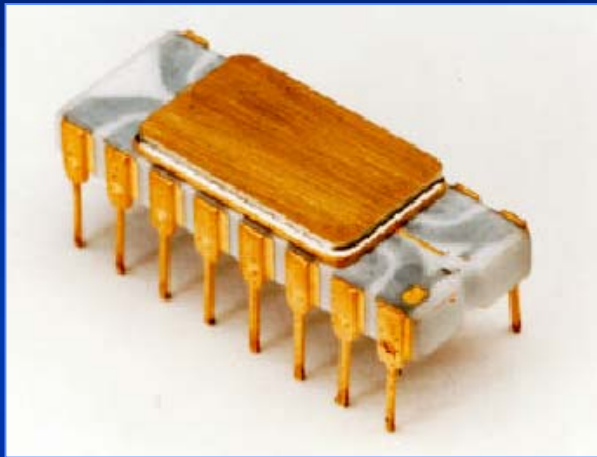


- Leakage sensitive circuits not functional at burn-in
- Larger keepers increase delay at nominal condition
- Conditional keeper enables functional burn-in

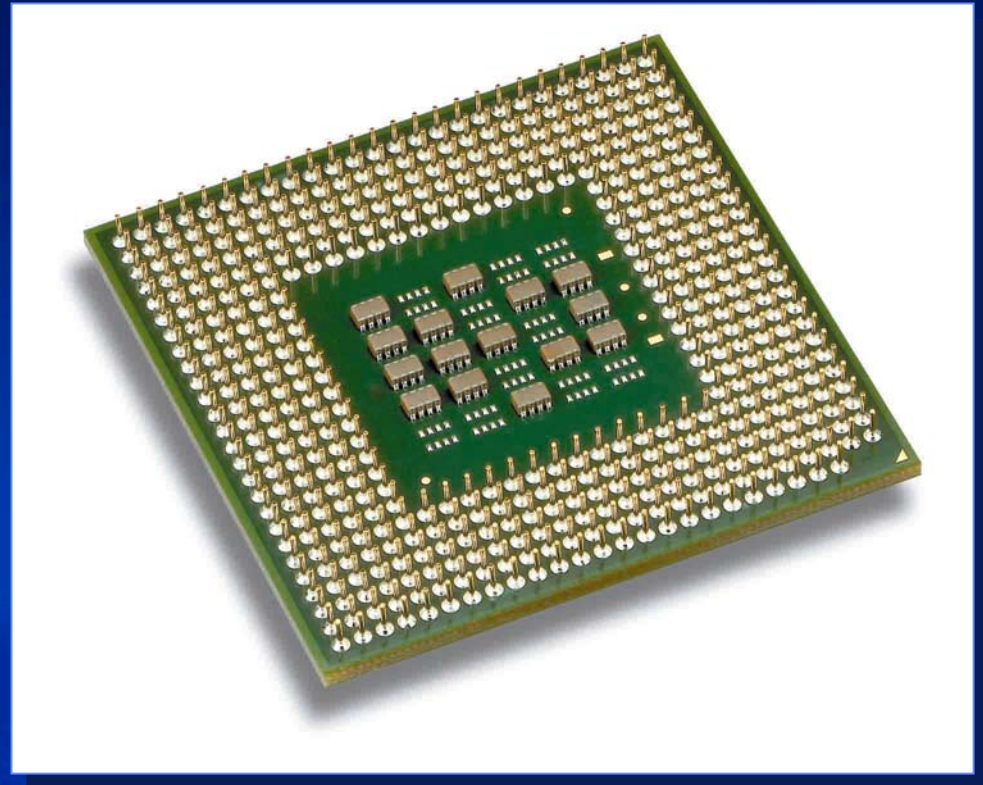
Agenda

- **Microprocessor Design Trends**
- **Process Technology Directions**
- **Active Power Management**
- **Leakage Reduction Techniques**
- **Thermal Modeling**
- **Call to Action**
- **Summary**

Microprocessor Package Evolution

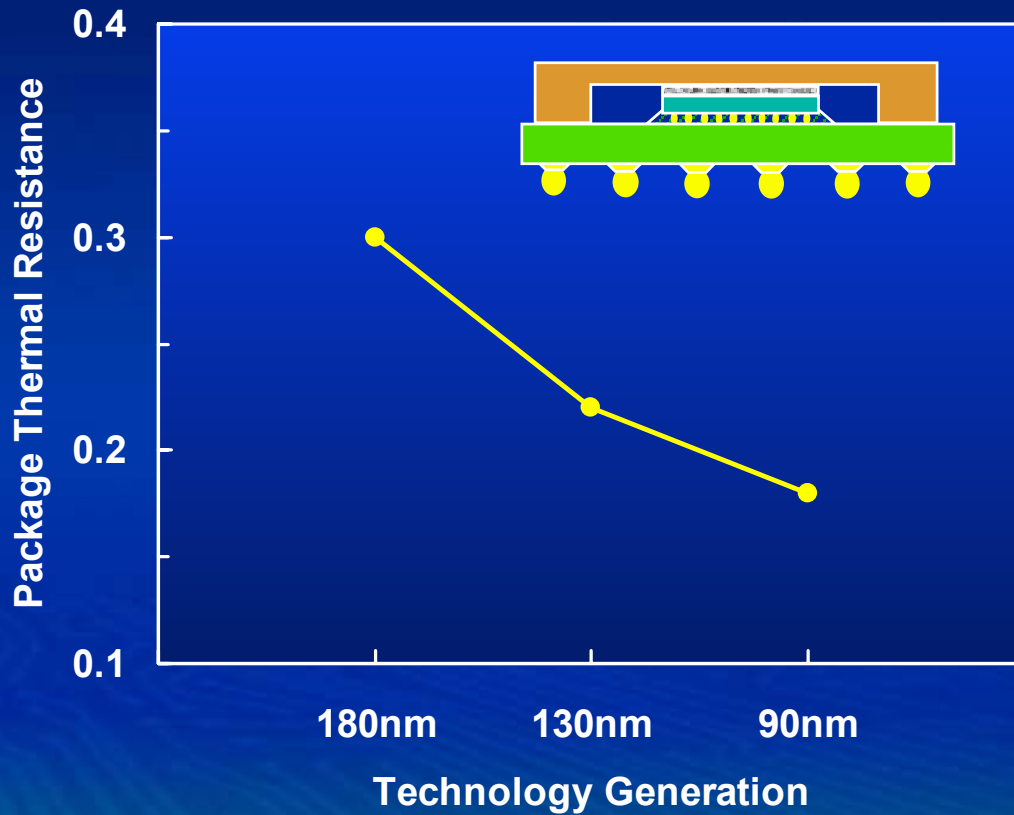


- **1971 – 4004 Processor**
 - 16-pin ceramic package
 - Wire bond attach
 - 750kHz I/O

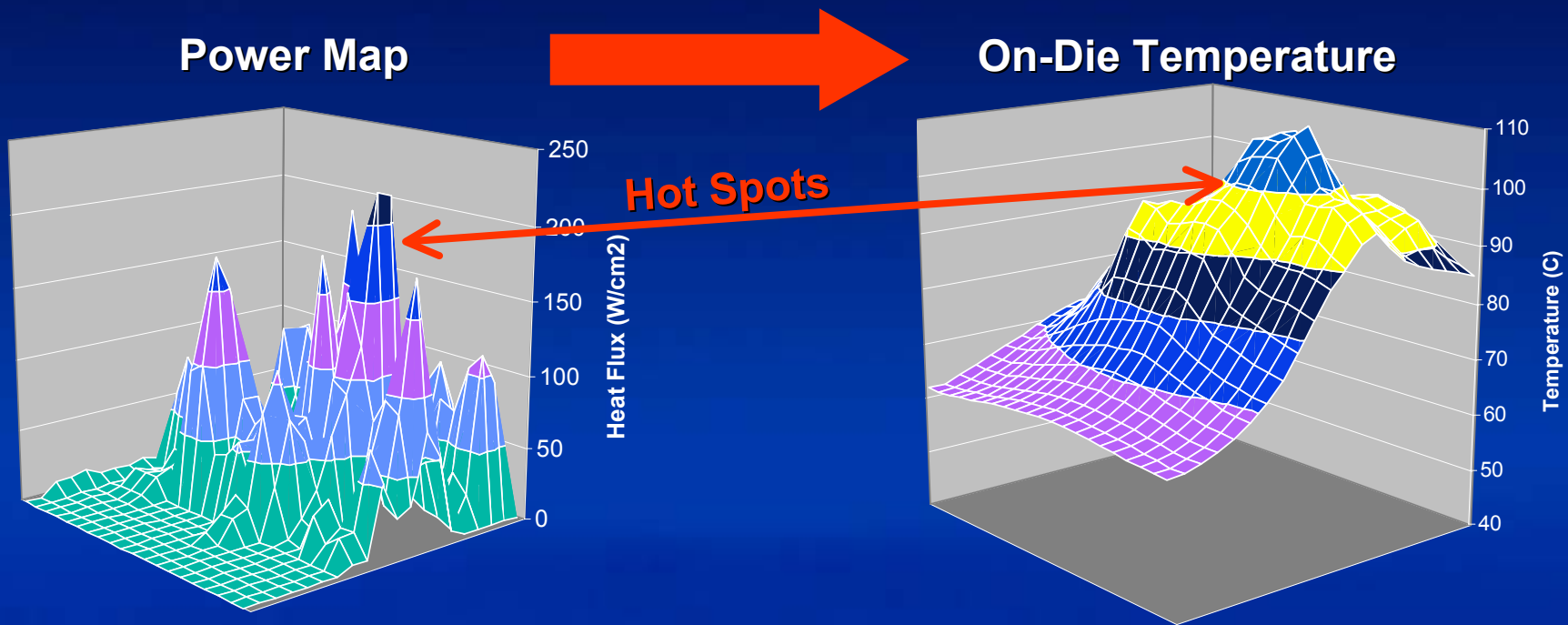


- **2003 - Pentium® 4 Processor**
 - 478-pin organic package
 - Flip-chip attach
 - 200MHz, quad-pumped I/O

Thermal Resistance

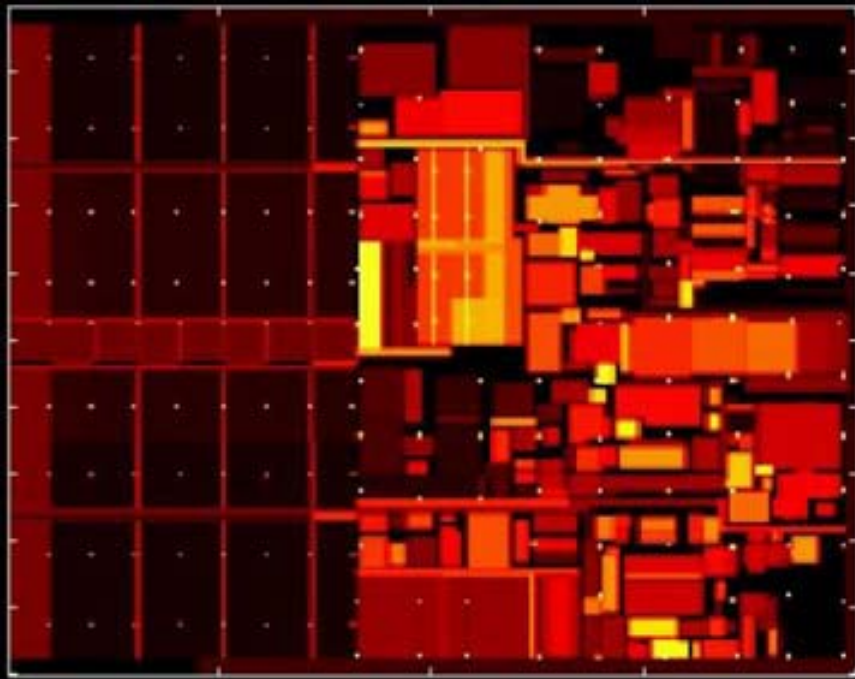


Power Density Models

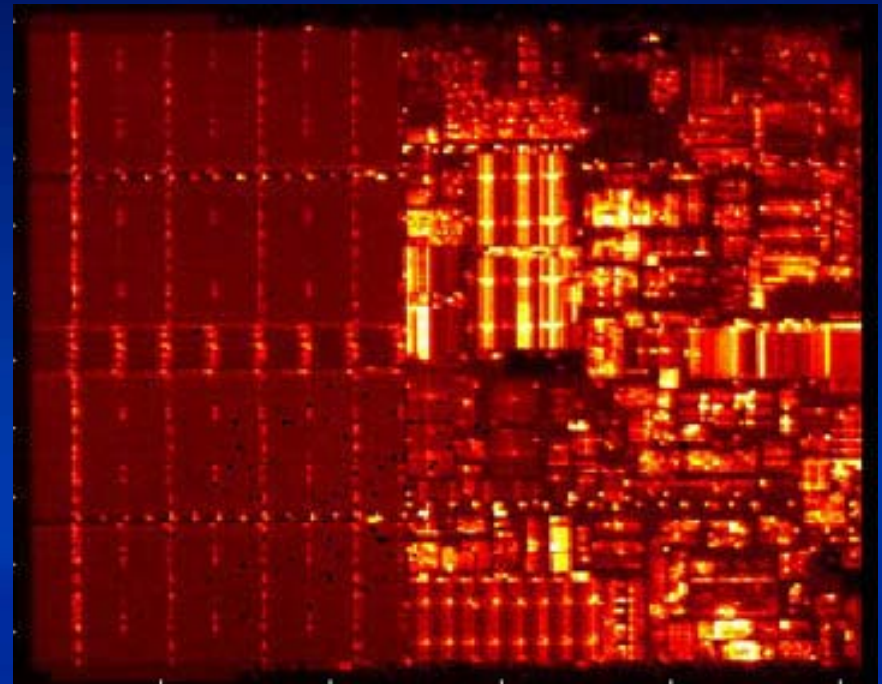


- With increasing power density and large on-die caches, detailed, non-uniform power models are required

Thermal Modeling



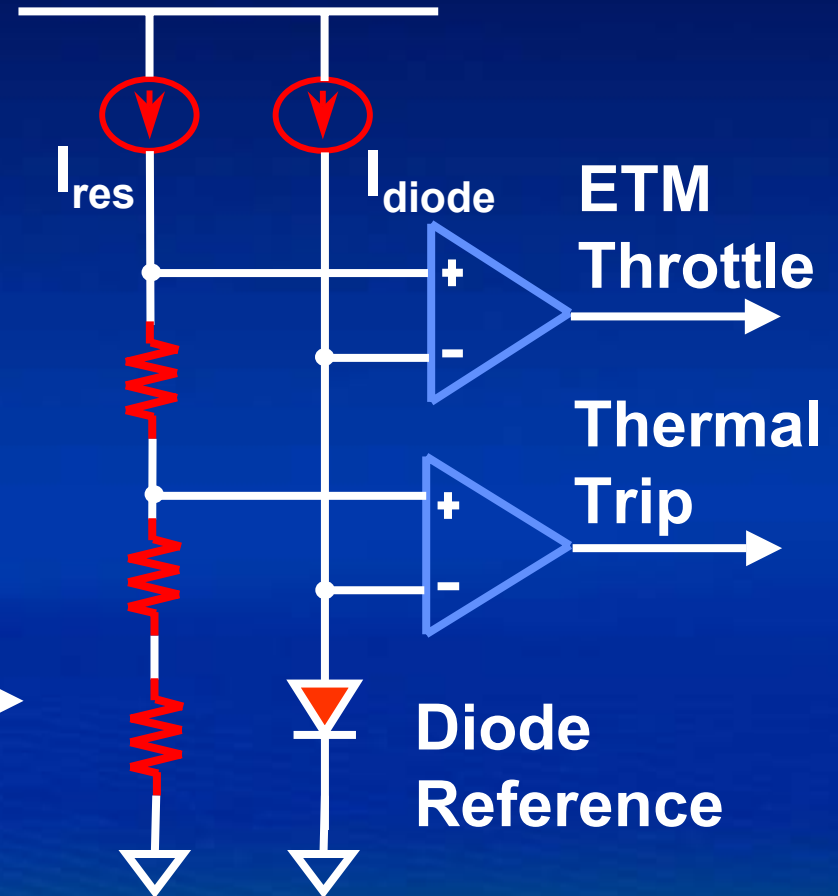
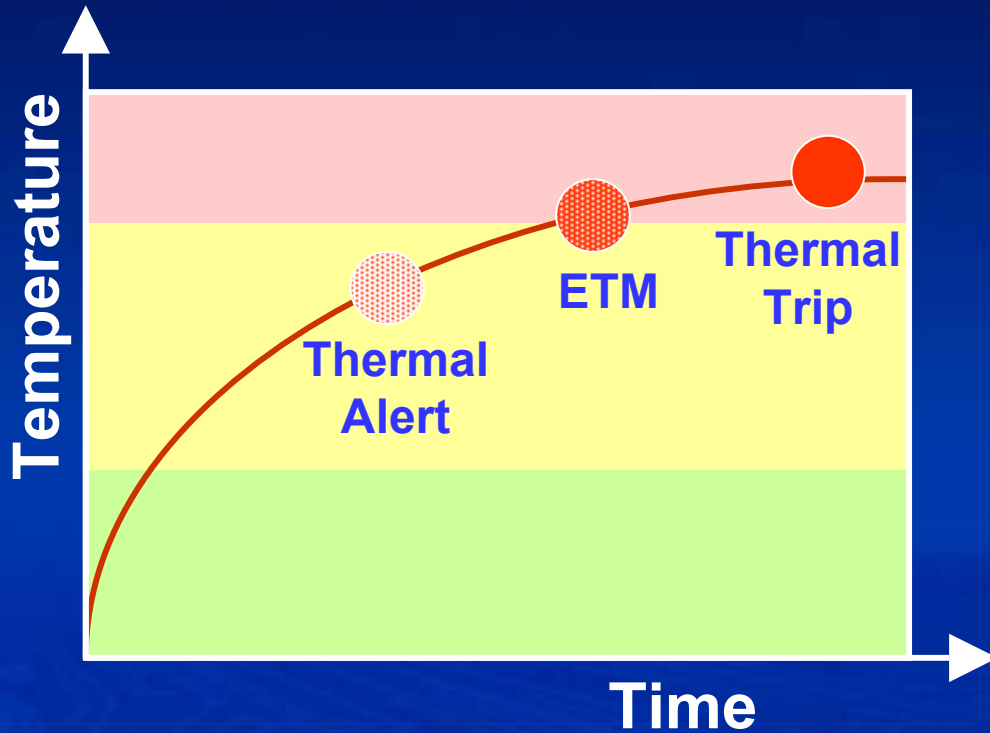
Simulated power density



**Infrared Emission
Microscope measurement**

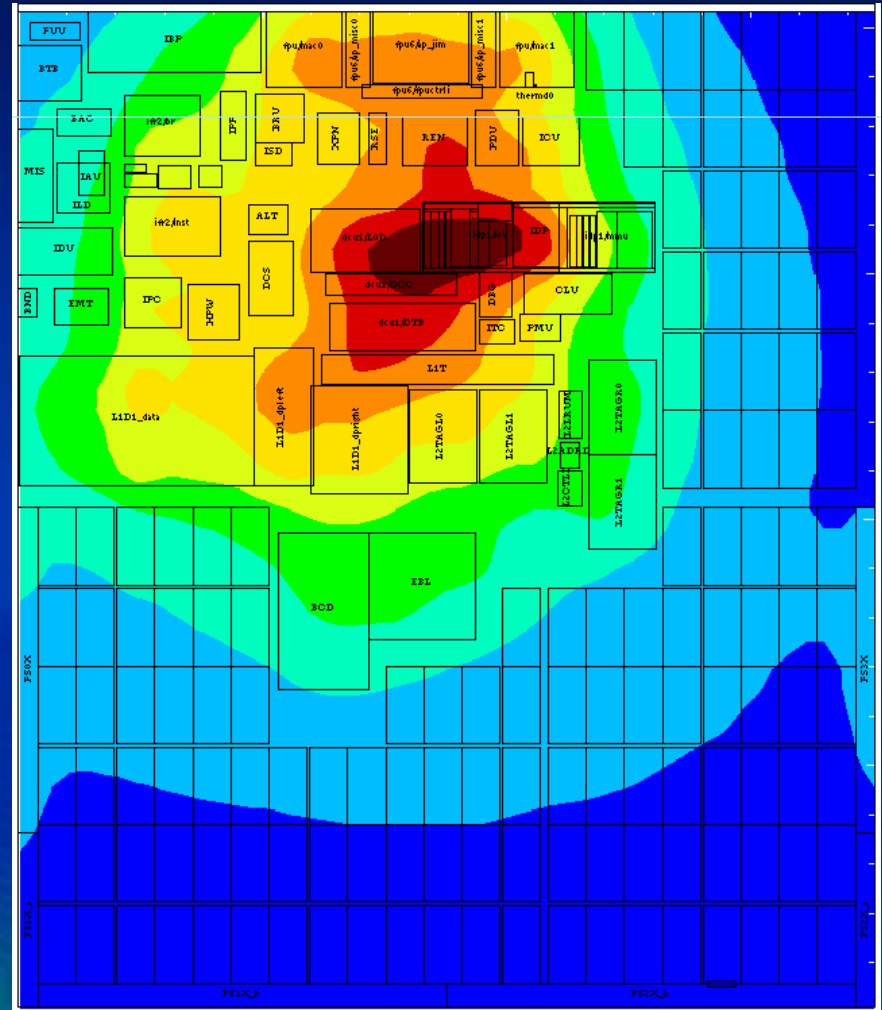
D. Genossar and N. Shamir "Intel® Pentium® M Processor Power Estimation, Budgeting, Optimization and Validation", Intel Technology Journal 5/2003

Thermal Protection Features



Metal Reliability Verification

- Metal routing validated for self-heating (SH) and electromigration (EM)
 - Macro-blocks verified through every geometry for SH/EM
 - Fullchip EM correct-by-construction
 - Fullchip SH verified through all geometries
- Thermal map used at the macro-block level to tighten constraints
 - Hottest areas of die need to meet higher standards



Agenda

- **Microprocessor Design Trends**
- **Process Technology Directions**
- **Active Power Management**
- **Leakage Reduction Techniques**
- **Thermal Modeling**
- **Call to Action**
- **Summary**

Call to Action

- **CAD tools must enable power and leakage reduction techniques with high productivity**
 - All design flows must be power and leakage aware
 - Need support for multiple transistors flavors (Le and Vt) and sleep devices for leakage reduction
- **CAD tools must comprehend process, temperature and voltage variations - worst casing is not practical**
 - Major shift from deterministic to probabilistic design
 - Design optimization must consider parameter variations
- **Simultaneous optimization of power, timing and noise**
 - Need accurate R,L,C extraction tools
 - Explore multiple solutions for noise problems

Summary

- **Moore's Law will continue for at least another decade**
 - 2X transistors growth per technology generation
 - 30nm and smaller transistors realized
- **Power and leakage are a significant challenge**
 - Exploit memory power efficiency → larger caches
 - Dynamic voltage and frequency adjustment
 - Circuit techniques (clock gating, sleep transistors)
- **EDA industry's job is to enable designers to keep pace with Moore's law**
 - Deliver tools and methodologies for increasingly complex designs
 - Focus on leakage and active power reduction