Future Trend of Microprocessor Design: Challenges and Realities

Robert Yung, Ph.D.
Chief technology officer, Enterprise Processors
(yung@intel.com)

Stefan Rusu
Senior Principal Engineer
(stefan.rusu@intel.com)

Kenneth Shoemaker
Senior Principal Engineer
(kenneth.d.shoemaker@intel.com)

Intel Corporation
ESSCIRC / ESSDERC 2002
Sept 25, 2002

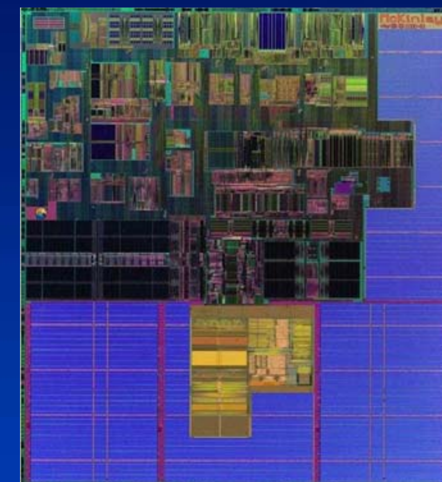
Agenda

- Process Driven Trends
  - Moore's Law
  - Transistors: Frequency, Power, Gate Length
  - Interconnection: Wires
  - Power Dissipation
  - Packaging
- Architecture Driven Trends
  - Increased Parallelism
  - Cache And Memory
  - Input/Output
- Conclusion

Microprocessor Evolution

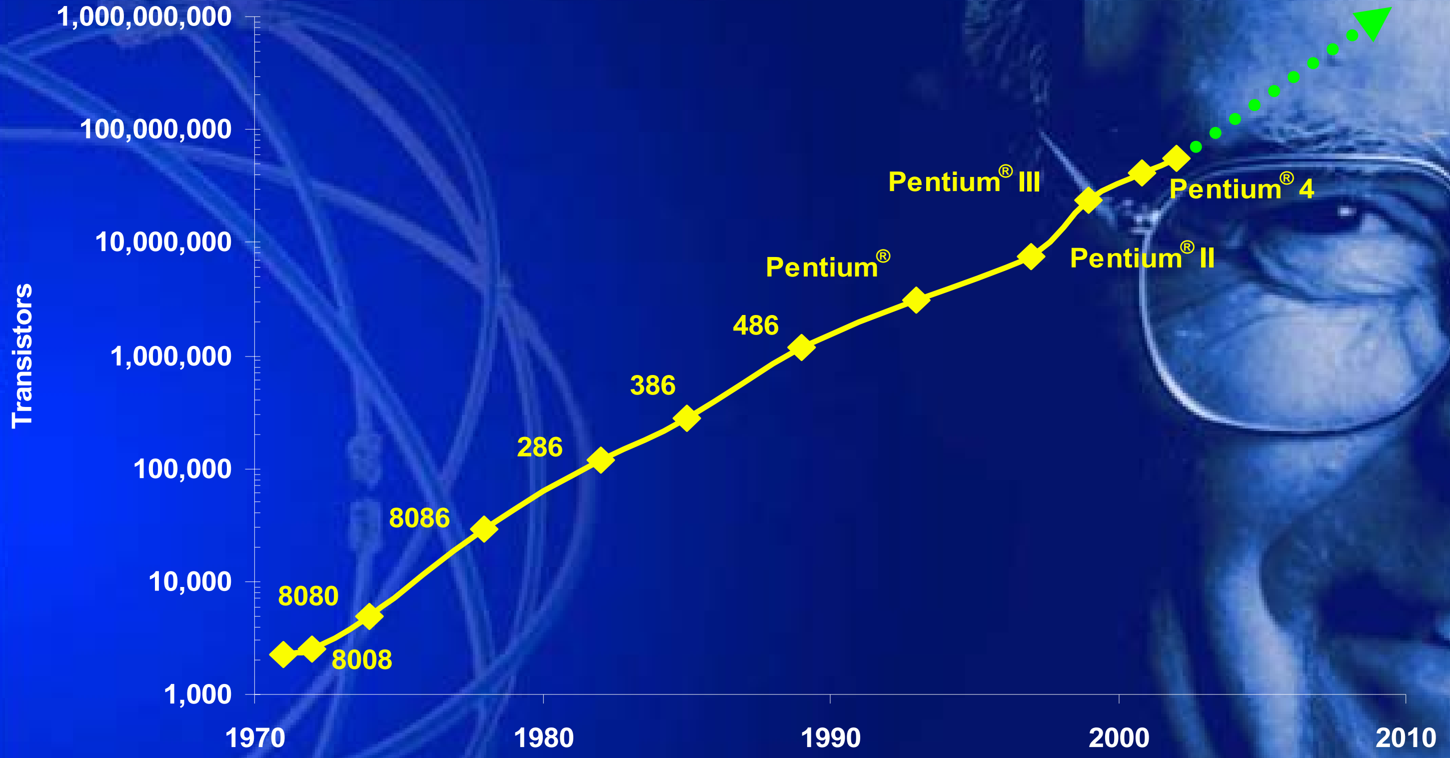
- 4004
  - 1971
  - 2300 transistors
  - 10um process
  - 2”, 50mm wafer
  - 12mm^2
  - 108 kHz

- Pentium® 4 processor
  - 2002 (31 yrs)
  - 55M (24K X)
  - 0.13um (1/77 X)
  - 12”, 300mm (6X)
  - 142mm^2 (12 X)
  - 2.8 GHz (26K X)



- Itanium® 2 processor
  - 2002 (31 yrs)
  - 220M (96K X)
  - 0.18um (1/55 X)
  - 12”, 300mm (6X)
  - 421mm^2 (35 X)
  - 1 GHz (9K X)

Moore's Law Continues



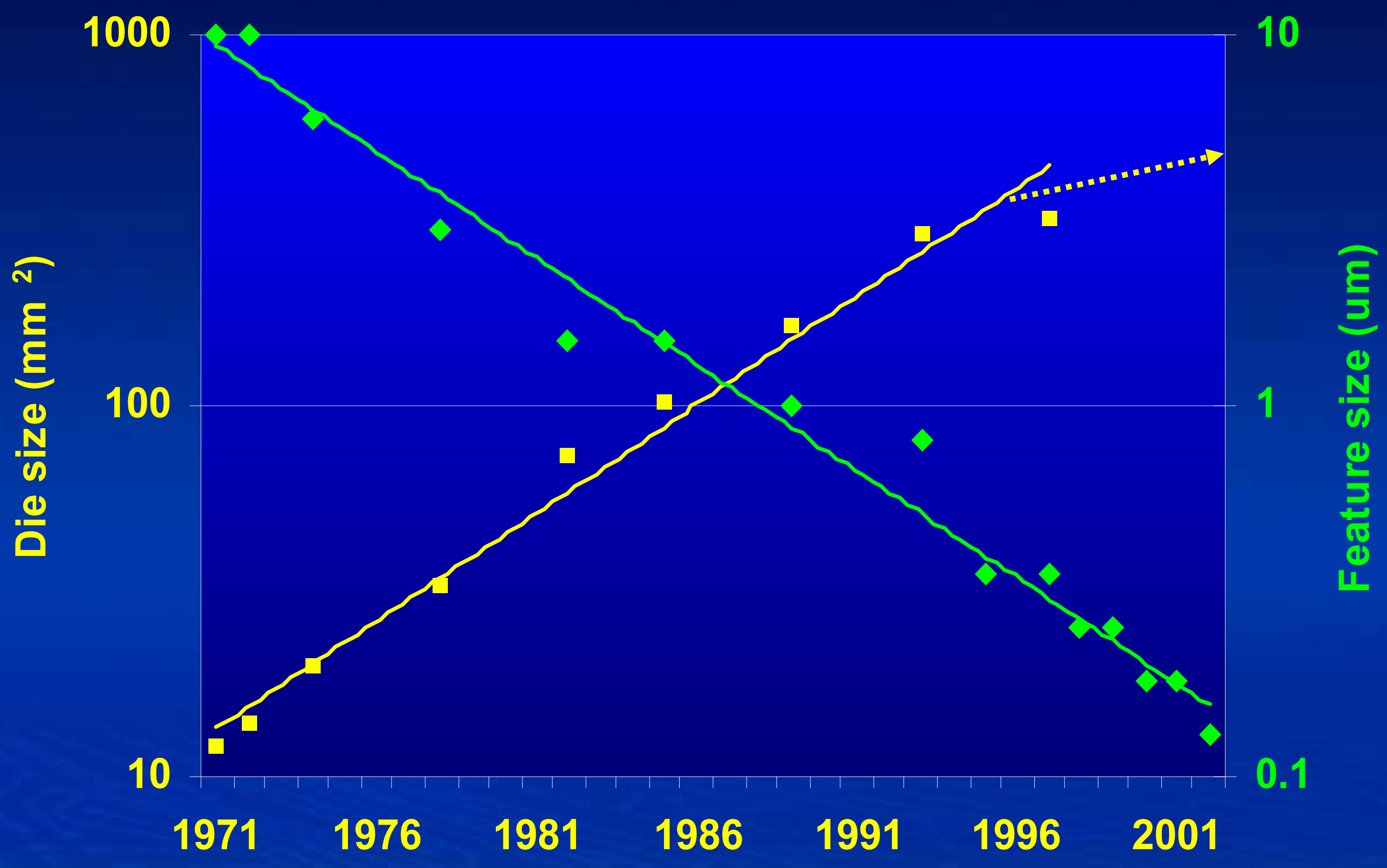
- Transistors per IC doubles every two years
- In less than 30 years
  - 1,000X decrease in size
  - 10,000X increase in performance
  - 10,000,000X reduction in cost
- Heading toward 1 billion transistors before end of this decade

In the Last 25 Years
Life was Easy

- Die sizes increase, allowed by
  - Increasing wafer size
  - Process technology moving from “black art” to “manufacturing science”
- Doubling of transistors every 18 months
- And, only constrained by cost & mfg. limits

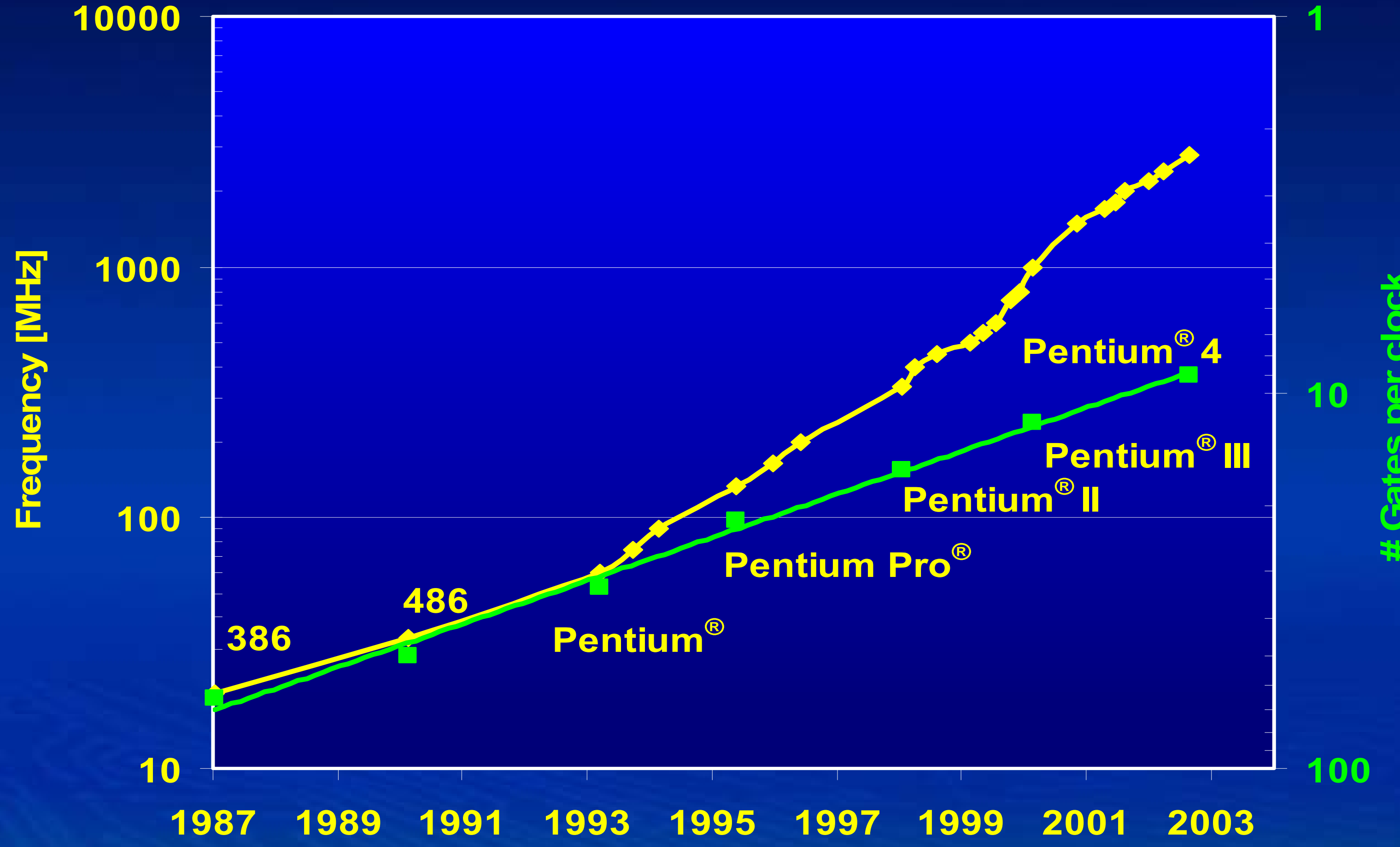
What Are The Future Challenges?

Feature, Die Size Trend



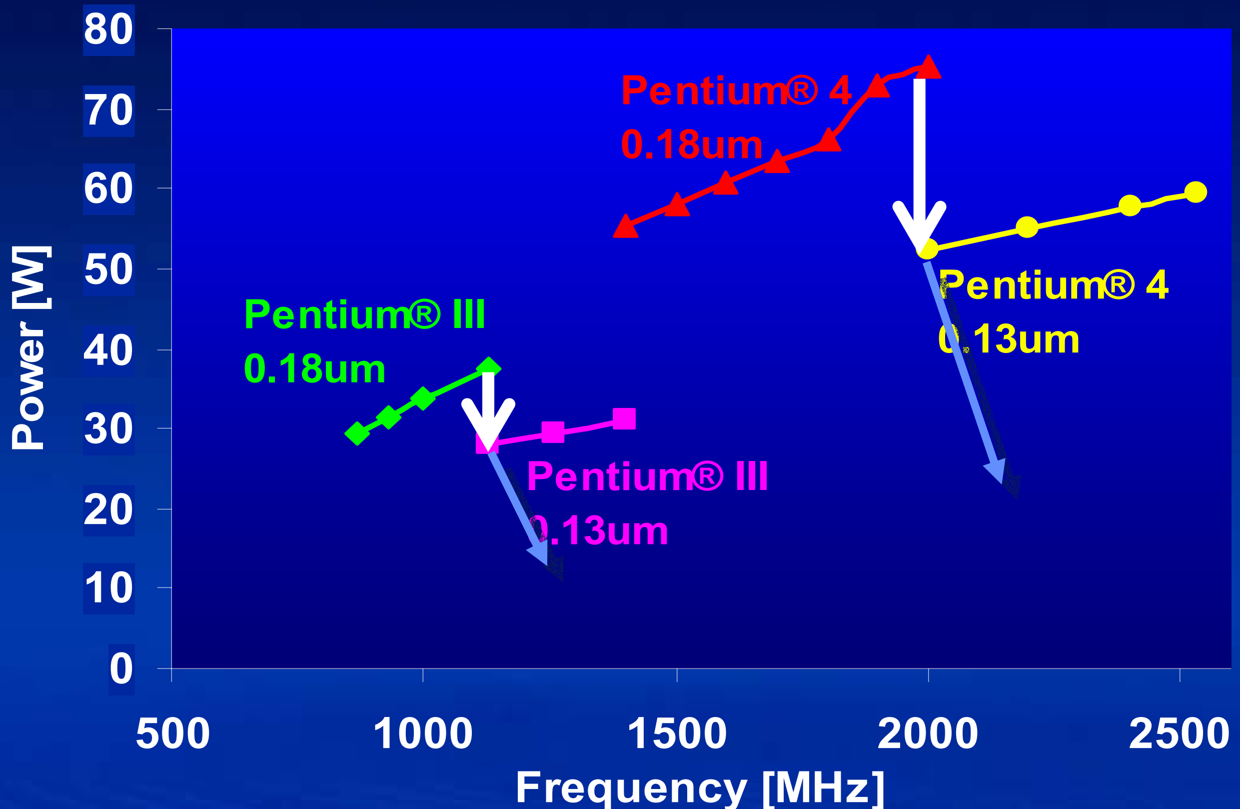
- 30% feature size reduction every 3 and now 2 yrs
- Before mid 1990's, 7% die size increase/yr; lithography limited
- After that, die size growth will be limited by power dissipation

Processor Frequency Trend



- Gates per clock reduces by 25% each generation; leveling out
- Frequency doubles each generation enabled by advanced circuit and architectural techniques

Processor Power Trend



- Lead processor power increases every generation — power constrained
  - Vcc will scale by only 0.8 (not 0.7)
  - Active power will scale by ~0.9 (not 0.5)
  - Active power density will increase by ~30-80% (not constant)
  - Leakage power will make it worse as process shrinks
- Process scaling provides higher performance at lower power

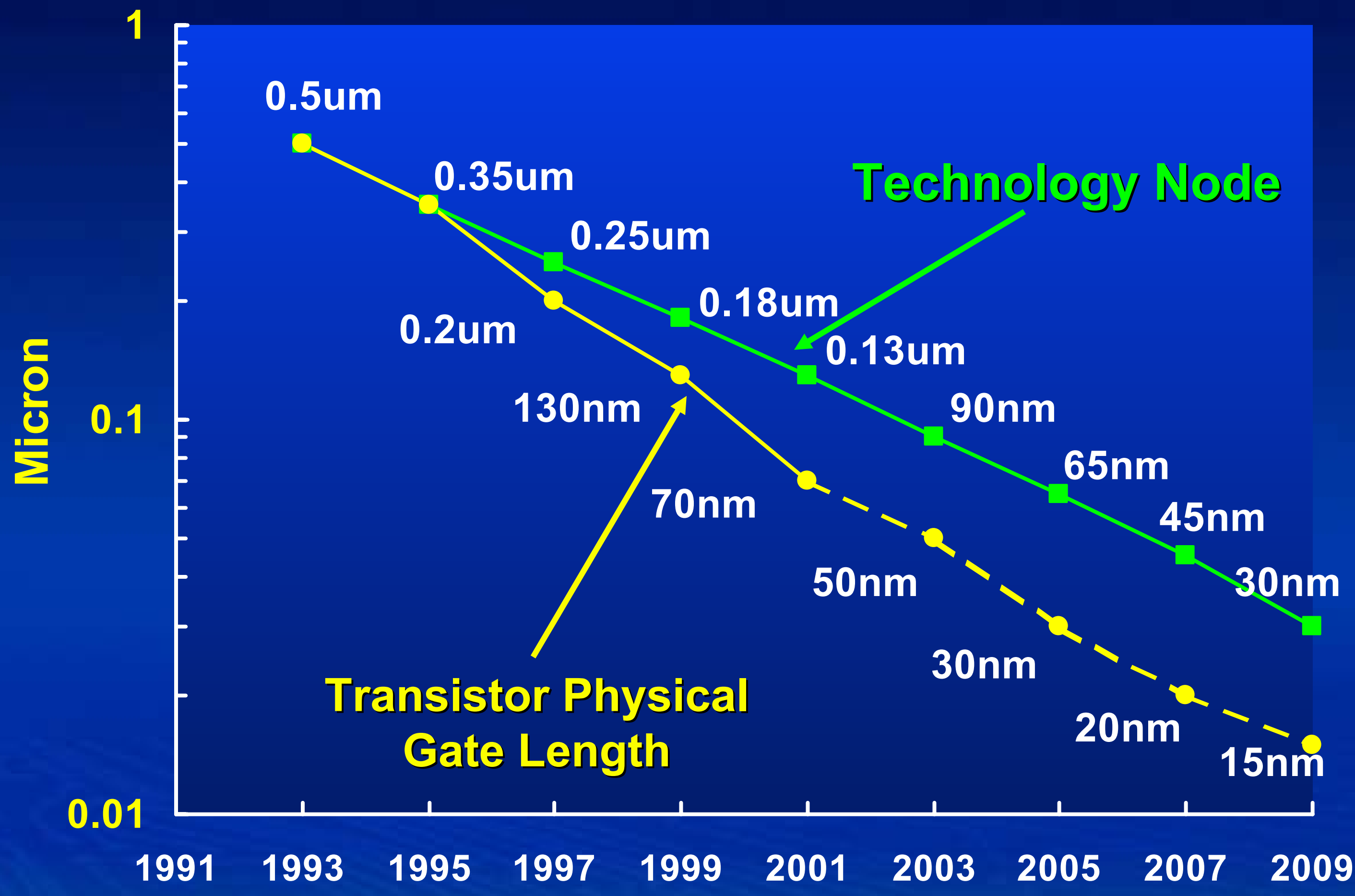
Some Implications

- Moore's Law will continue beyond this decade
  – 2X transistors growth per technology generation
- Die size increase will level out
  – Constraint is power – not manufacturability
- Frequency will continue to increase
  – Faster process, advanced micro-architecture
  – Reduction of gates per clock will slow down
- What is the future look like?
  – Process technology trend
  – Microprocessor and platform architectural trend

Agenda

- Process Driven Trends
  - Moore's Law
  - Transistors: Frequency, Power, Gate Length
  - Interconnection: Wires
  - Power Dissipation
  - Packaging
- Architecture Driven Trends
  - Increased Parallelism
  - Cache And Memory
  - Input/Output
- Conclusion

Transistor Physical Gate Length



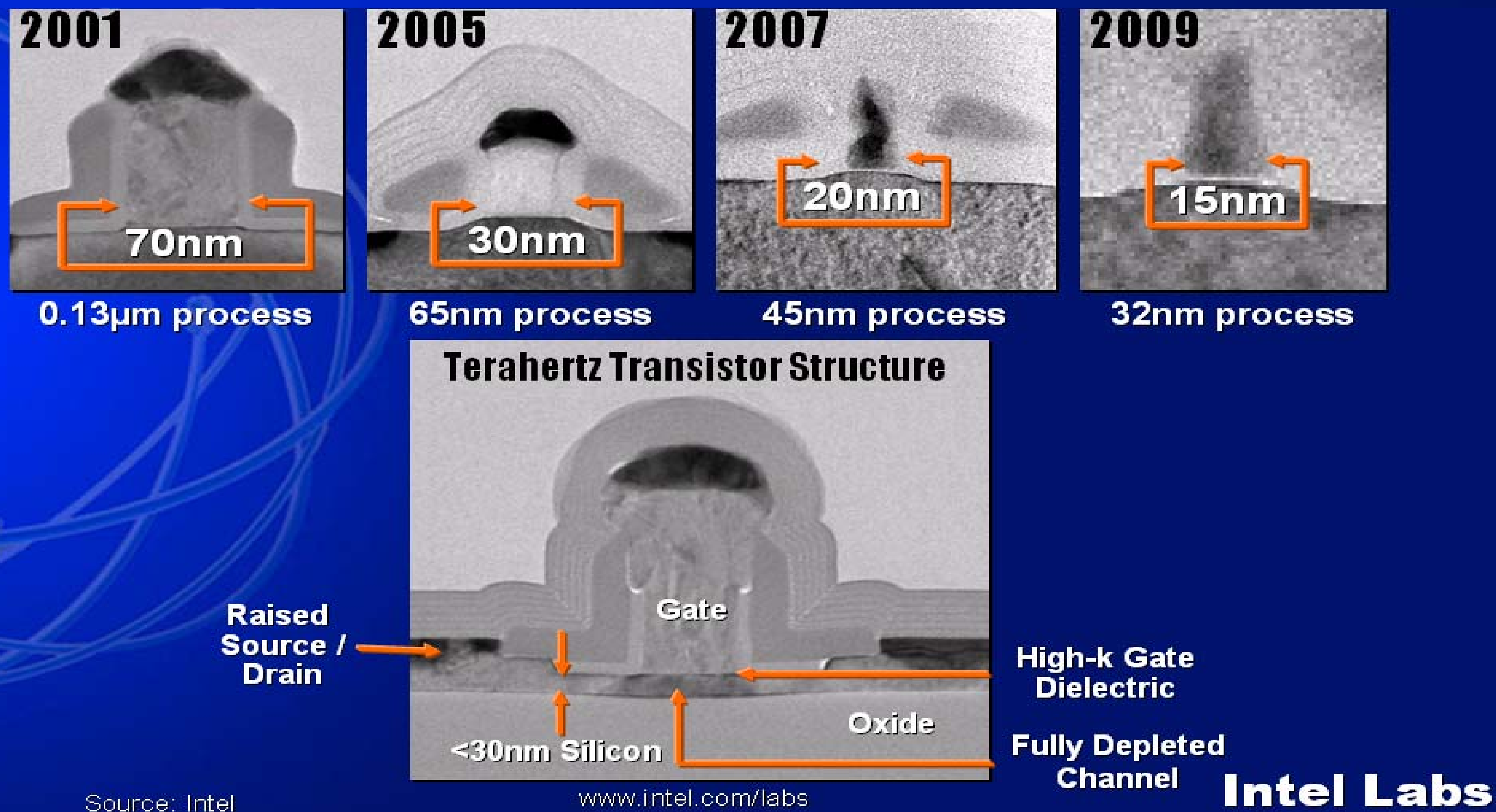
New Process Generation Every 2 Years

Source: Robert Chau, 12/2001

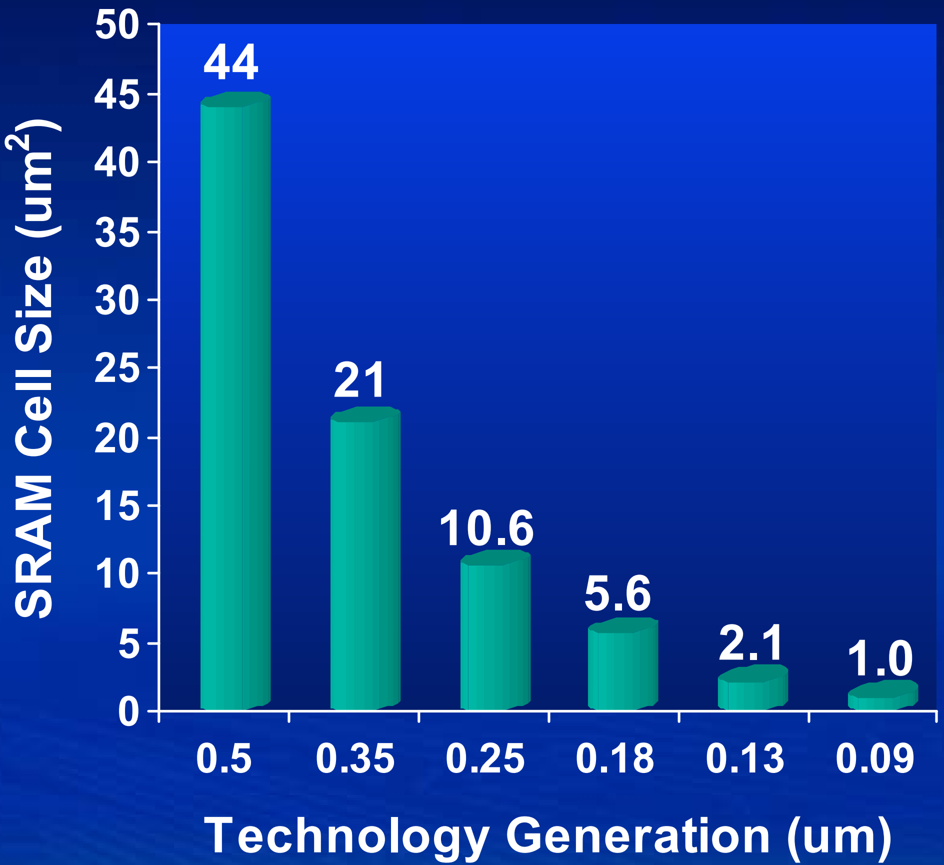
Process Technology Trends

Intel: To the Terahertz Transistor

Transistor Leadership Continues



SRAM Cell Size Scaling

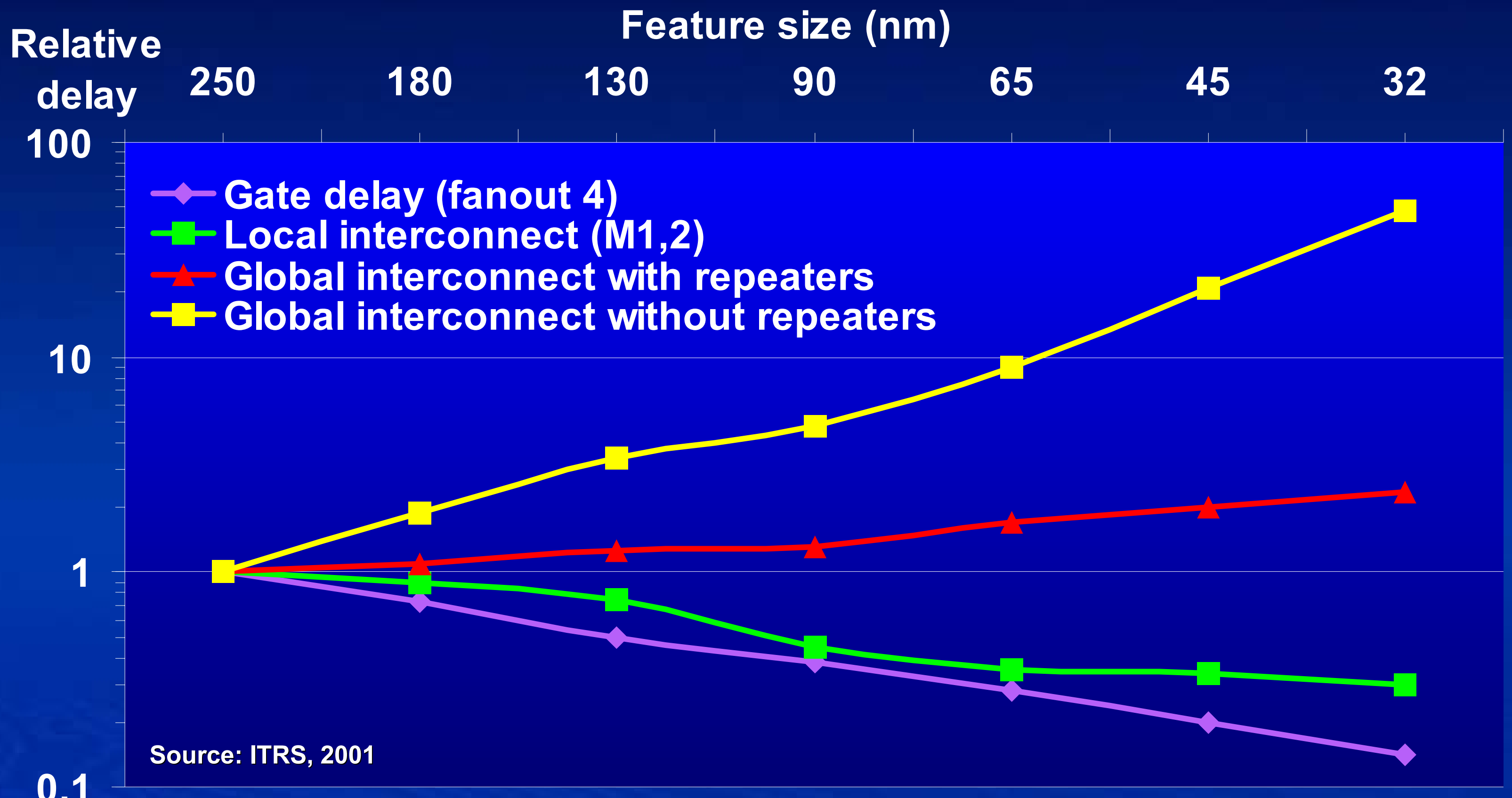


- SRAM cell size will continue to scale ~0.5x per generation
- Larger caches can be incorporated on die

Agenda

- Process Driven Trends
  - Moore's Law
  - Transistors: Frequency, Power, Gate Length
  - Interconnection: Wires
  - Power Dissipation
  - Packaging
- Architecture Driven Trends
  - Increased Parallelism
  - Cache And Memory
  - Input/Output
- Conclusion

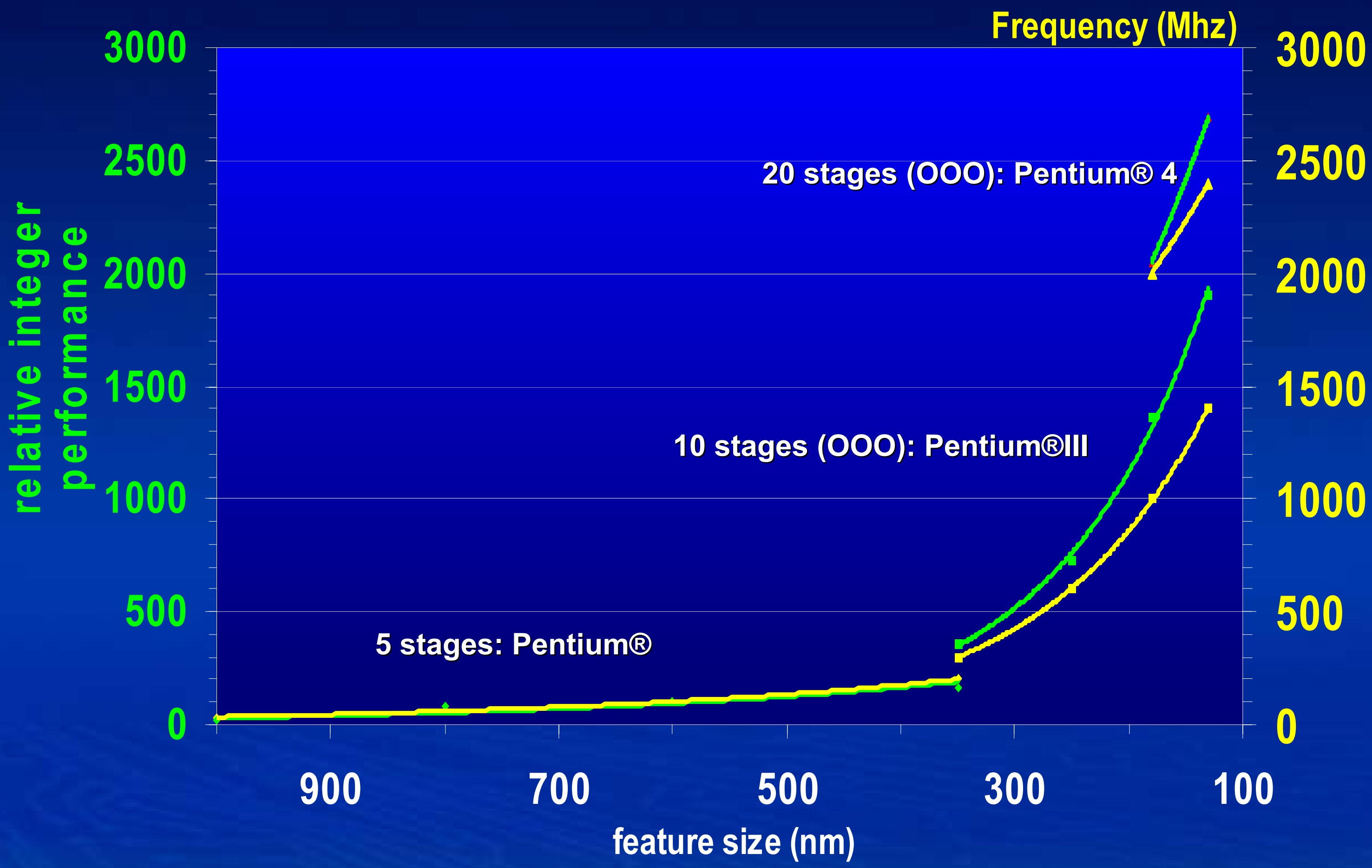
On-chip Interconnect Trend



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

More metal layers may not help

Pipe Length vs. Frequency Trend



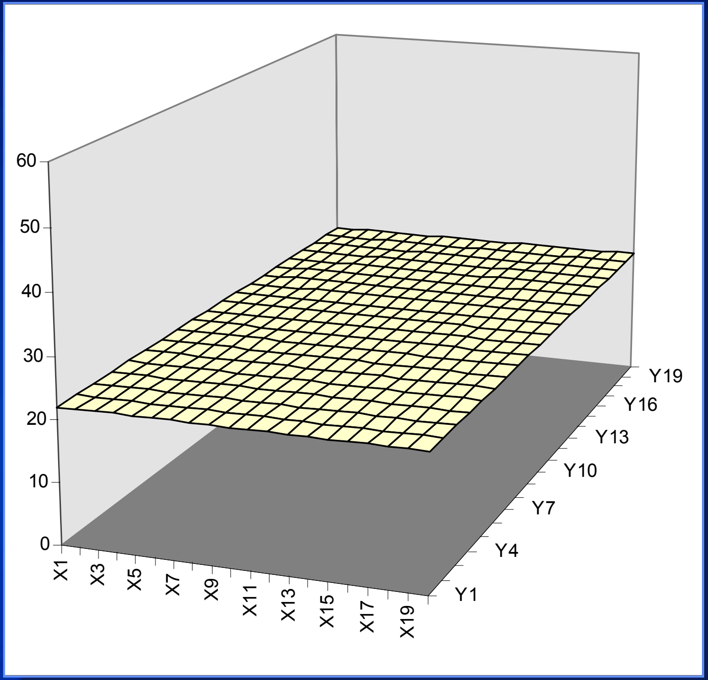
- As feature size reduces, longer pipeline enables higher frequency
- Performance benefits from higher frequency, advanced micro-architectural techniques, larger caches

Agenda

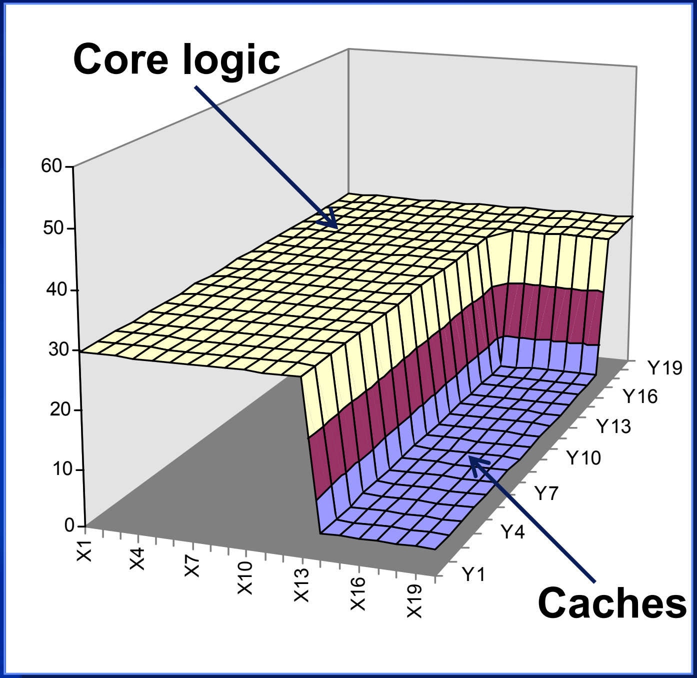
- Process Driven Trends
  - Moore's Law
  - Transistors: Frequency, Power, Gate Length
  - Interconnection: Wires
  - Power Dissipation
  - Packaging
- Architecture Driven Trends
  - Increased Parallelism
  - Cache And Memory
  - Input/Output
- Conclusion

Power Density: Cache vs. Logic

Past: Thermal Uniformity

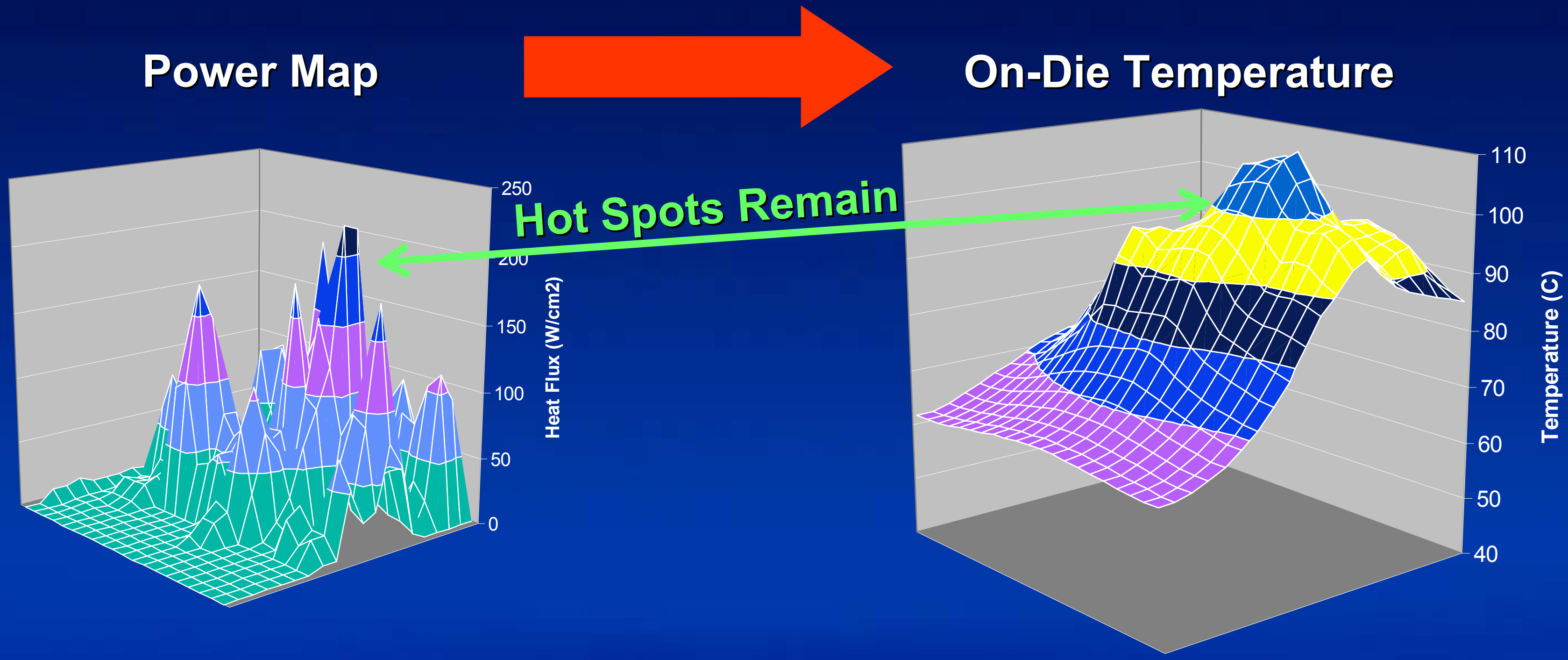


Present: Logic vs. Cache



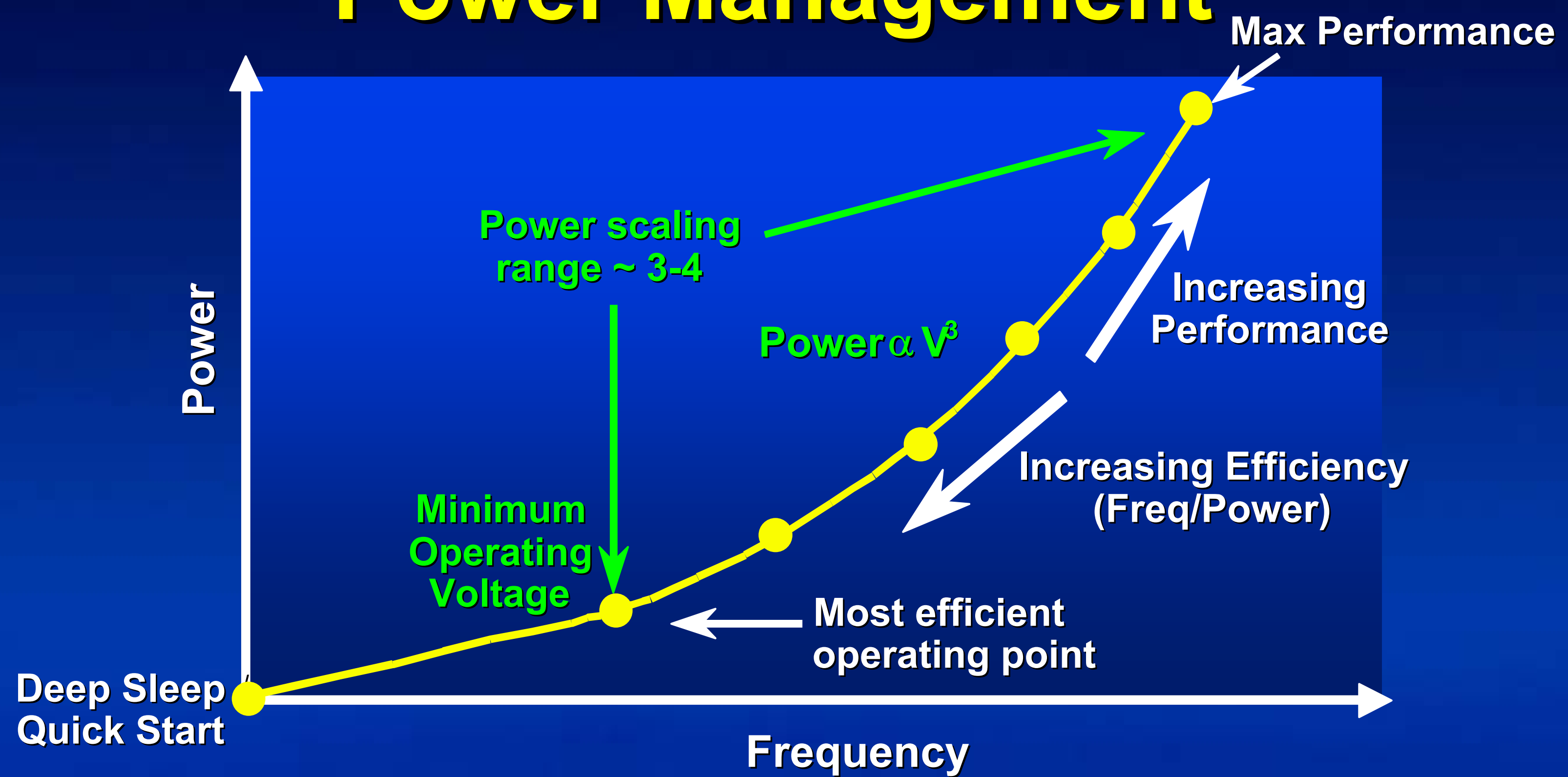
- As die temperature increases, CMOS logic slows down
- With low power density (past), can assume uniformity
- With increasing power density and on-die caches, need to consider simplistic non-uniformity

Power Density: The Future



- With high power density, cannot assume uniformity
  - As die temperature increases, CMOS logic slows down
  - At high die temp., long-term reliability can be compromised

Power Management

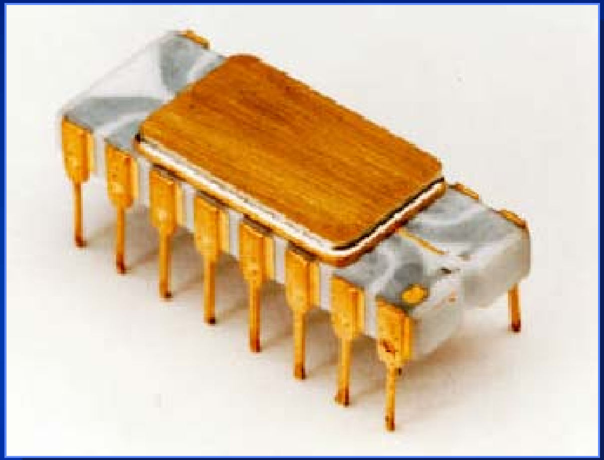


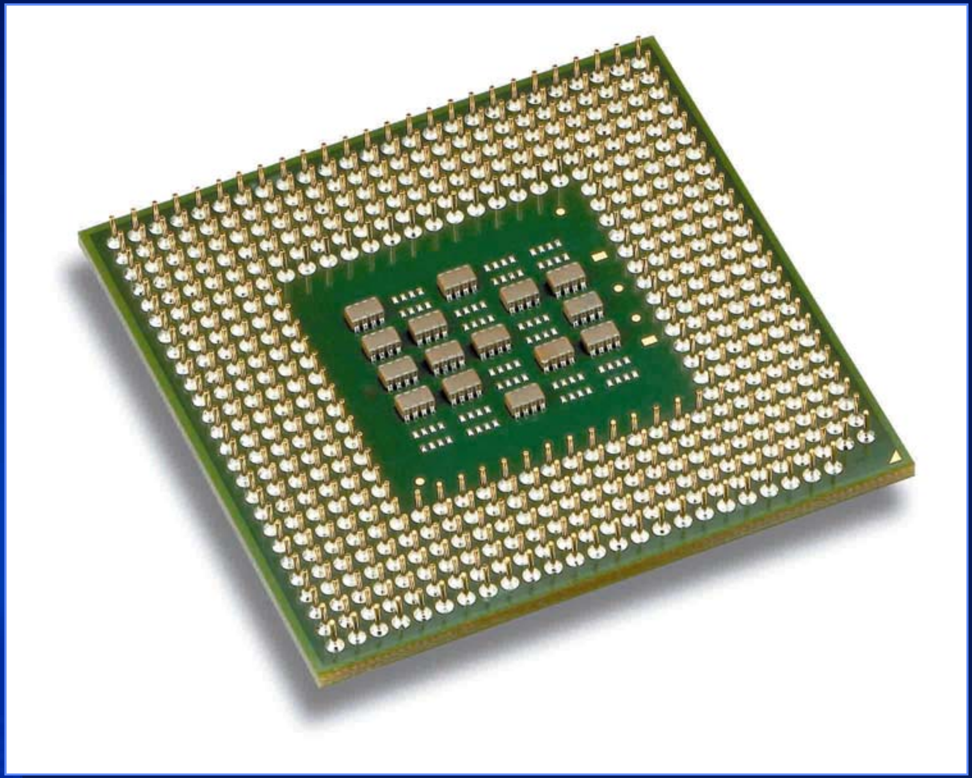
- Intel Speedstep® Technology (Geyserville)
  - Voltage-freq scaling with active thermal feedback
  - Multi-operating states from high perf. to deep sleep
- Throttling to reduce instruction rate
- Power management reduces average and peak power dissipation
- Trend: Static logic, clock gating, split power planes, active power mgmt.

Agenda

- Process Driven Trends
  - Moore's Law
  - Transistors: Frequency, Power, Gate Length
  - Interconnection: Wires
  - Power Dissipation
  - Packaging
- Architecture Driven Trends
  - Increased Parallelism
  - Cache And Memory
  - Input/Output
- Conclusion

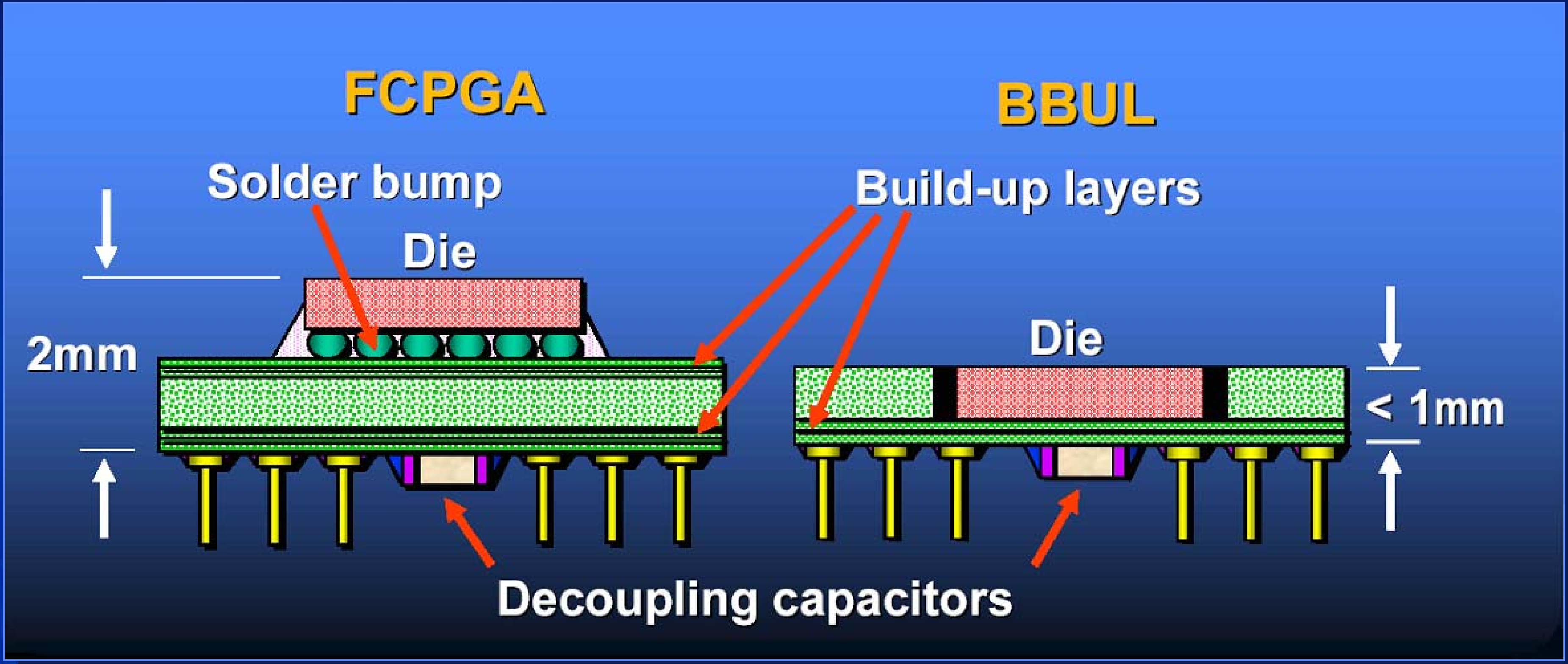
Microprocessor Packaging





- 1971 – 4004 Processor
  - 16-pin ceramic package
  - wire bond attach
  - 750Khz I/O
- 2002 - Pentium® 4 Processor
  - 478-pin organic package
  - flip-chip attach
  - 133Mhz, quad-pumped I/O

FCPGA vs. BBUL

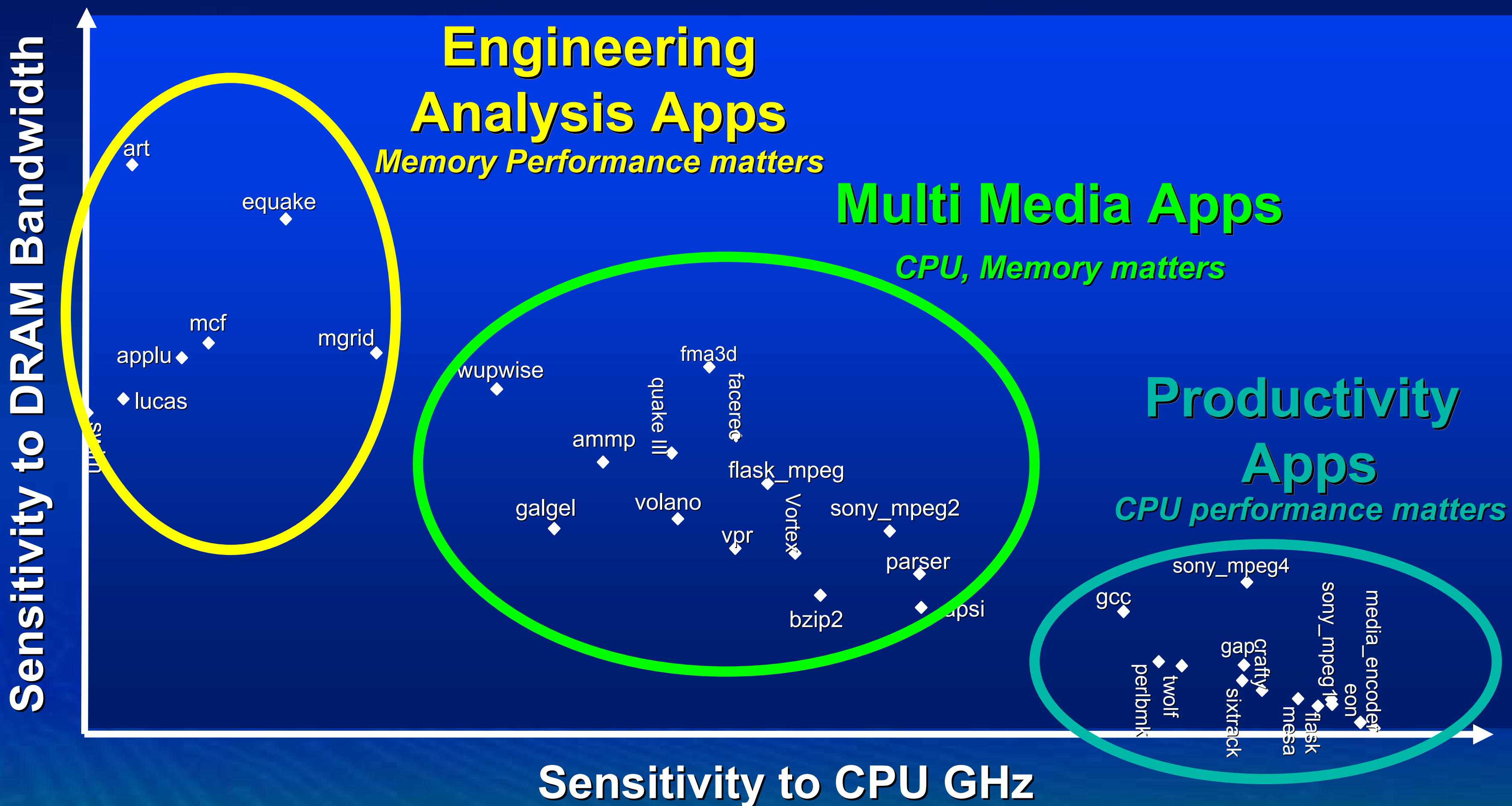


- Package built around die→ shorter profile→ smaller form factor
- Results in lower inductance, higher frequency

Agenda

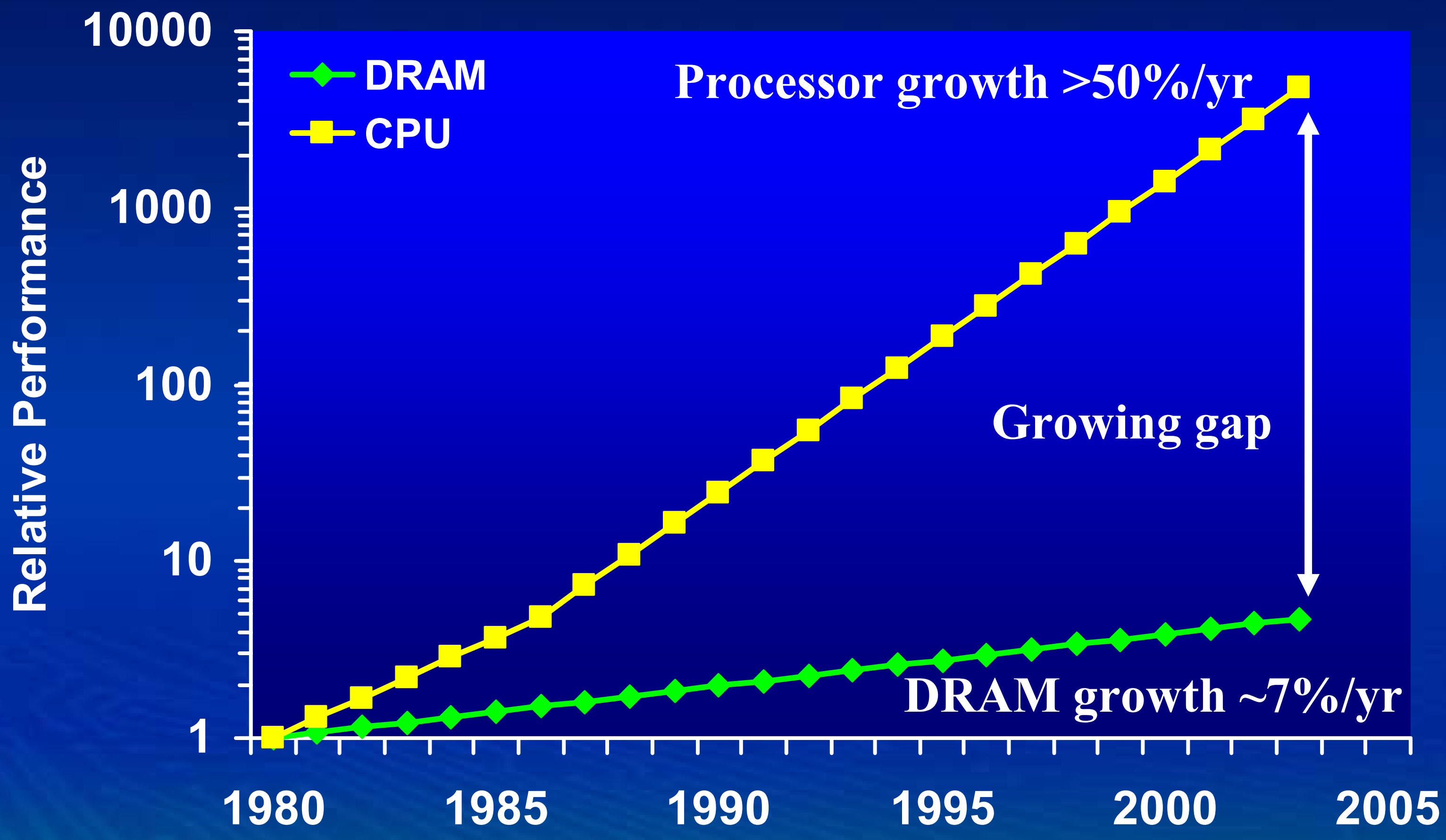
- Process Driven Trends
  - Moore's Law
  - Transistors: Frequency, Power, Gate Length
  - Interconnection: Wires
  - Power Dissipation
  - Packaging
- Architecture Driven Trends
  - Increased Parallelism
  - Cache And Memory
  - Input/Output
- Conclusion

CPU, Memory Sensitivity of Apps



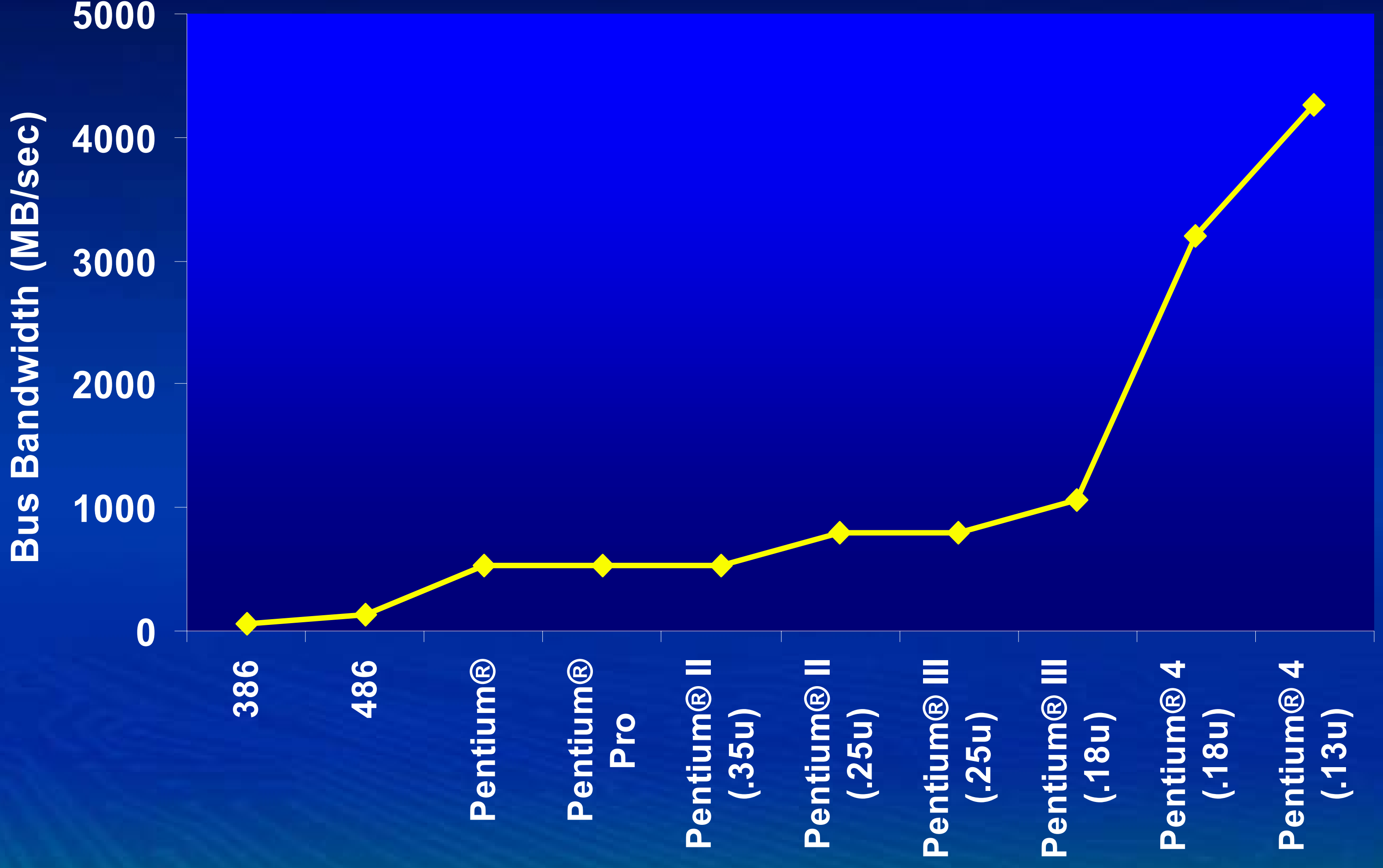
Apps Show Different Sensitivity To Bandwidth And CPU Frequency

Processor-DRAM Gap (latency)



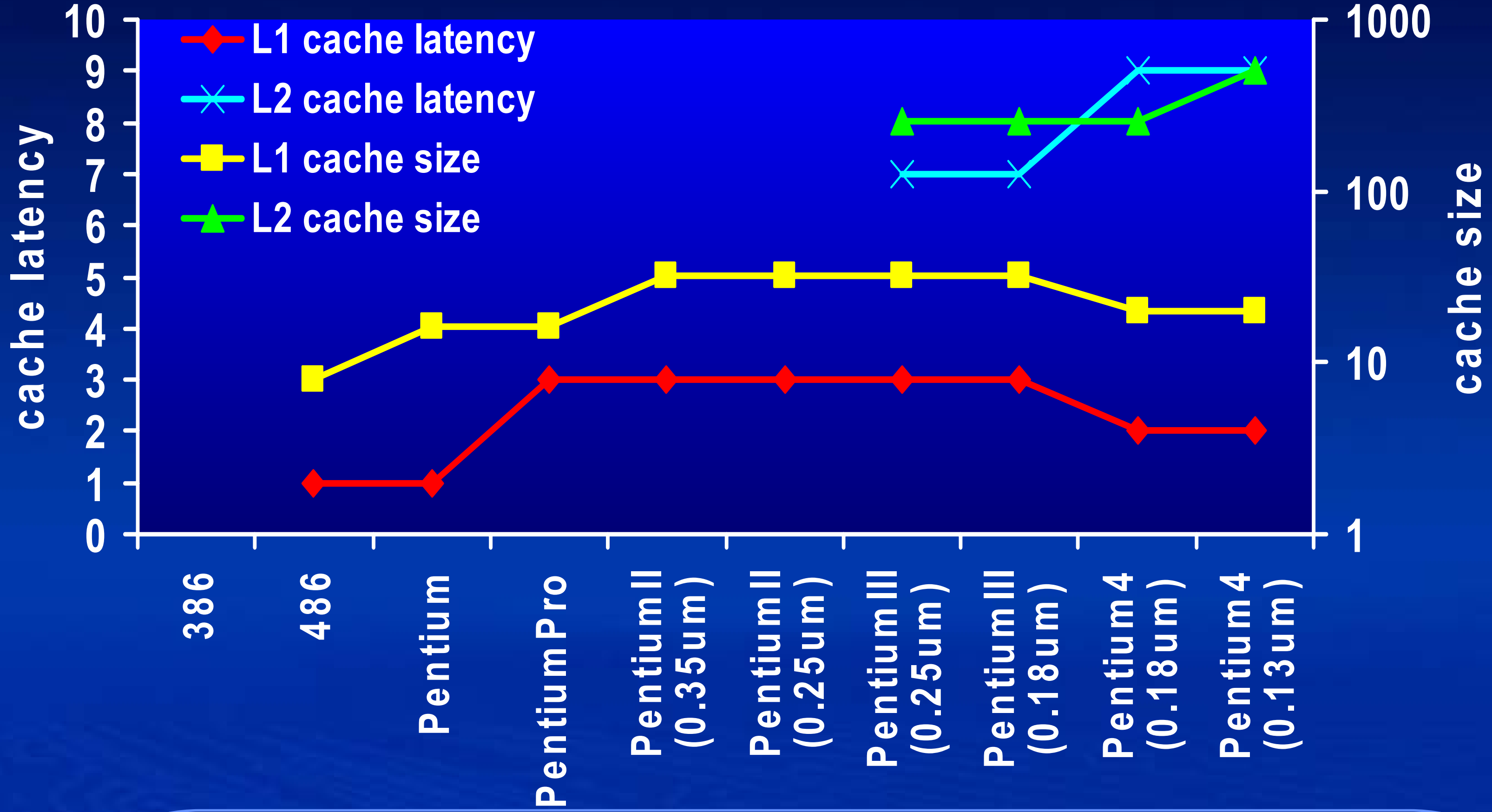
Processor-DRAM Gap Grows >40% Year

Bus Bandwidth Trend



Memory And I/O Bandwidth Are Crucial For High Performance

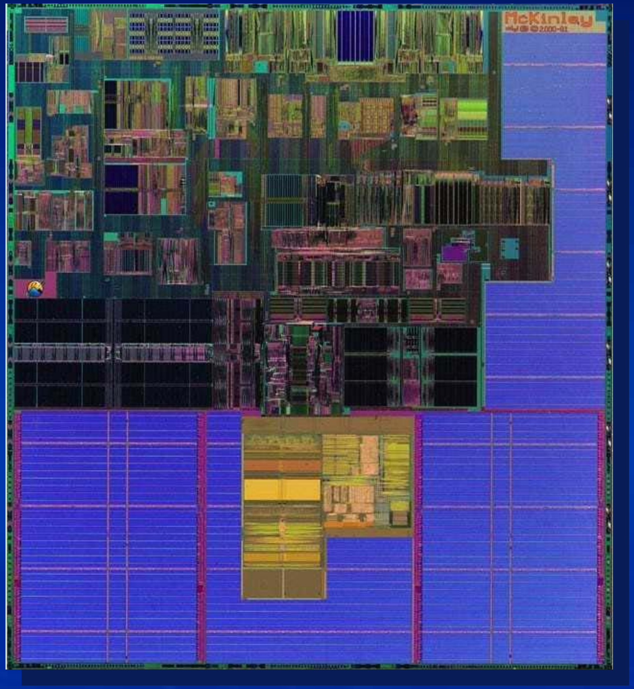
Cache Memory Trend



- Hierarchy of caches reduce widening CPU-memory gap
  - Reduce average miss rates
  - Reduce average memory access latency

Itanium® 2 Processor

- Transistors: 221M
  - Caches, I/O: 3.3MB or ~170M (75%)
  - Core: ~51M (25%)
- Die size: 19.5 x 21.6mm = 421 mm^2
  - Caches, I/O: L3C ~50%; others ~16%
  - Core: 142mm^2 (34%)



Caches becoming an increasing portion of the die because of its performance impact and low power density

Conclusion

- Moore's Law will continue beyond this decade
  - 2X transistors growth per technology generation
  - 30nm and smaller transistors realized
- Die size increase will level out
  - Constraint is power – not manufacturability
  - Increasing cache sizes and multi-cores on die enable performance increase within power constraint
- Towards 10Ghz microprocessor in this decade
  - Faster process
  - Advanced architectural and circuit techniques
- Processor-Memory gap continues to grow
  - Larger caches help reduce impact
  - Innovative processor-cache memory design crucial to continual performance scaling