A Journey Through History From Mainframes to Smartphones

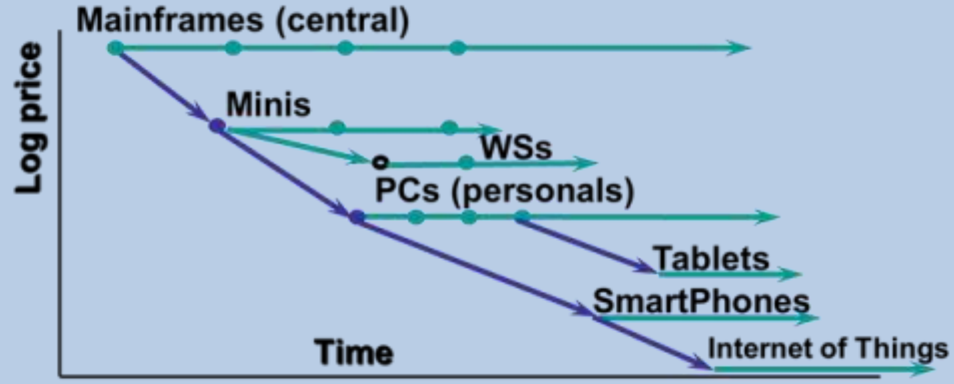
Dileep Bhandarkar, Ph. D.
IEEE Life Fellow

Electronic Design Process Symposium
April 2015

Disclaimer

This presentation is based on personal
Experiences over the last 40+ years in industry
and
Is not presented on behalf of
current or past employers.

Disruptions Come from Below!



Bell's Law:
hardware technology, networks, and interfaces allows new, smaller, more specialized computing devices to be introduced to serve a computing need.

Performance

Mainframes

Minicomputers

RISC Systems

Desktop PCs

Notebooks

Smart Phones

Volume

The First 50 Years after Shockley's Transistor Invention

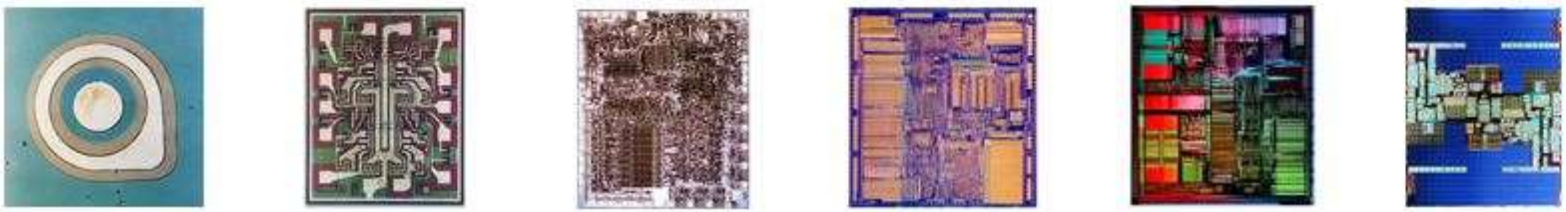
Directions | Contact Us

COMPUTER HISTORY MUSEUM

Welcome | Timeline | People | Companies | Resources | Glossary | Search | Search Exhibit | GO

MOORE'S LAW "Transistor density on integrated circuits doubles about every two years." *

1950s	1960s	1970s	1980s	1990s	2000s
Silicon Transistor	TTL Quad Gate	8-bit Microprocessor	32-bit Microprocessor	32-bit Microprocessor	64-bit Microprocessor



1 Transistor	**16 Transistors**	**4500 Transistors**	**275,000 Transistors**	**3,100,000 Transistors**	**592,000,000 Transistors**

Microelectronic silicon computer "chips" have grown in capability from a single transistor in the 1950s to hundreds of millions of transistors per chip on today's microprocessor and memory devices. From the first documented semiconductor effect in 1833 to the transition from transistors to integrated circuits in the 1960s and 70s, this website explores key milestones in the development of these extraordinary engines that power the computing and communications revolution of the information age.

*Source: "Moore's Law: Raising the Bar" (Intel Corporation 2005)

Photo credits: Fairchild Camera and Instrument Corporation, Intel Corporation (Note that images are not to scale)

1958: Jack Kilby's Integrated Circuit

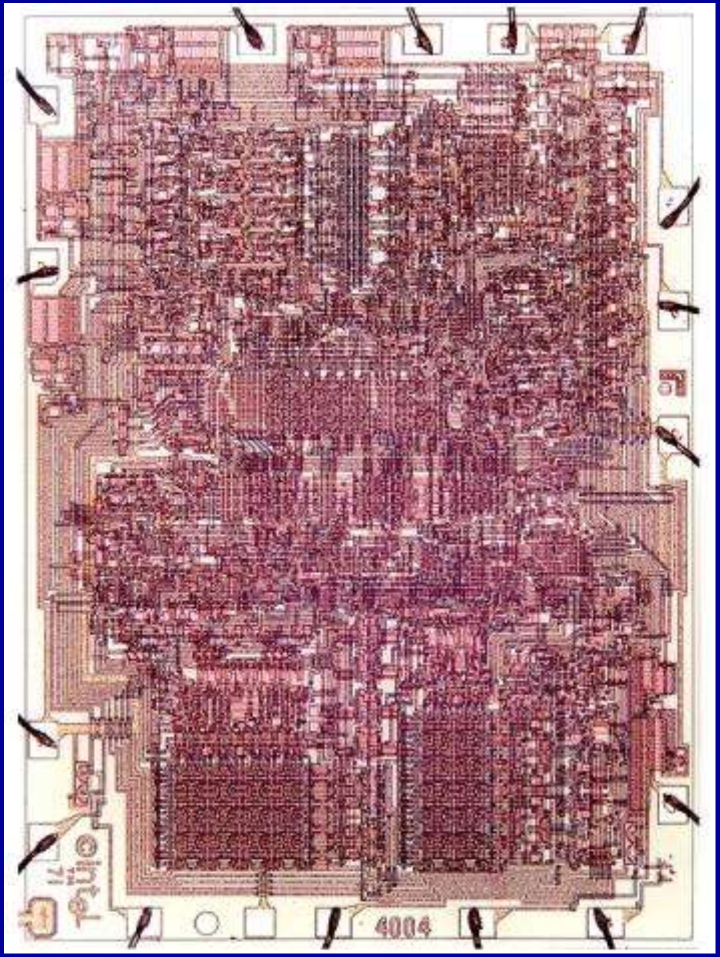
SSI -> MSI -> LSI -> VLSI -> OMGWLSI

Dennard Scaling

Device or Circuit Parameter	Scaling Factor
Device dimension tox, L, W	1/K
Doping concentration Na	K
Voltage V	1/K
Current I	1/K
Capacitance eA/t	1/K
Delay time per circuit VC/I	1/K
Power dissipation per circuit VI	$1/K^2$
Power density VI/A	1

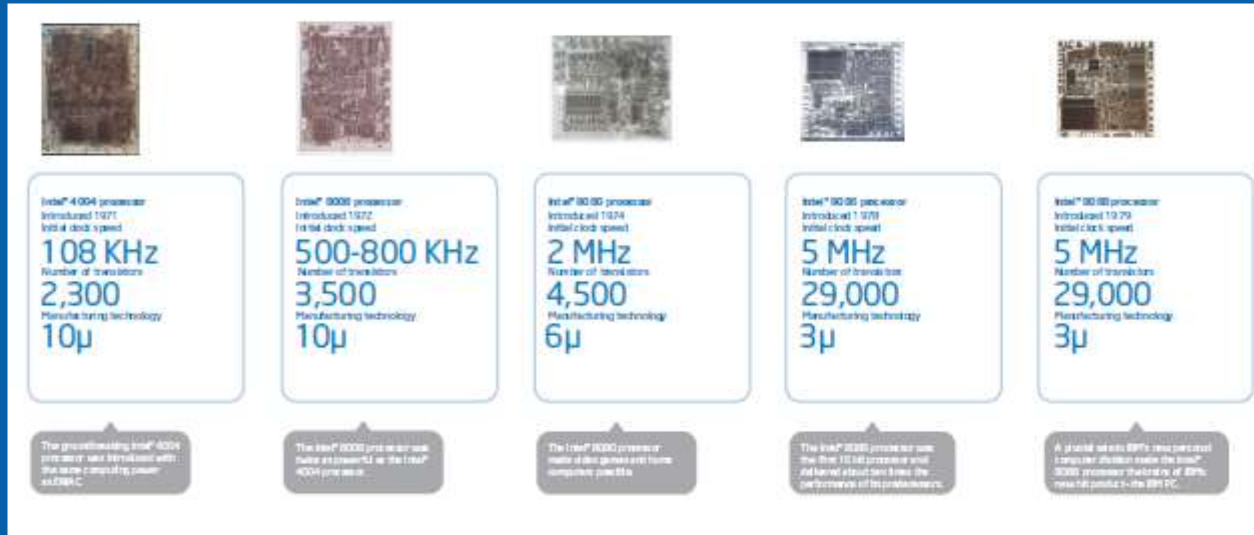
Dennard's 1974 paper summarizes transistor or circuit parameter changes under ideal MOSFET device scaling conditions, where K is the unitless scaling constant.
The benefits of scaling : as transistors get smaller, they can switch faster and use less power.
Each new generation of process technology was expected to reduce minimum feature size by approximately 0.7x (K ~1.4). A 0.7x reduction in linear features size provided roughly a 2x increase in transistor density.
Dennard scaling broke down around 2004 with unscaled interconnect delays and our inability to scale the voltage and the current due to reliability concerns.
But our the ability to etch smaller transistors has continued spawning multicore designs.

1971: 4004 Microprocessor



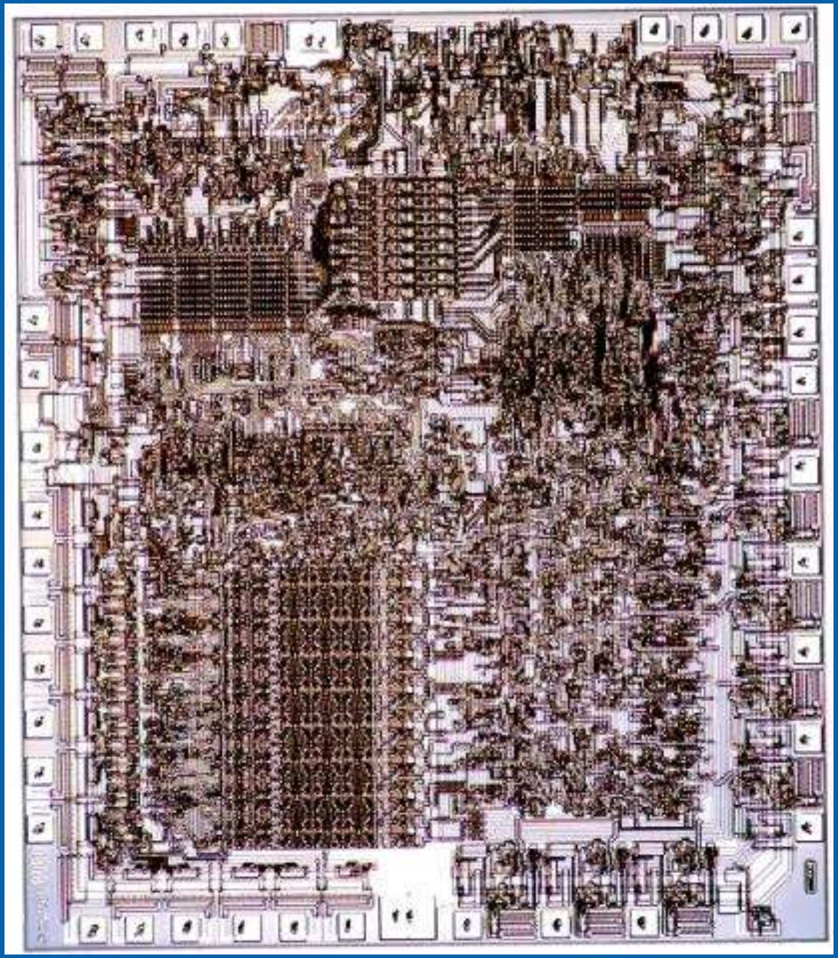
- The 4004 was Intel's first microprocessor. This breakthrough invention powered the Busicom calculator and paved the way for embedding intelligence in inanimate objects as well as the personal computer.

Introduced November 15, 1971
108 KHz, 50 KIPs , 2300 10μ transistors



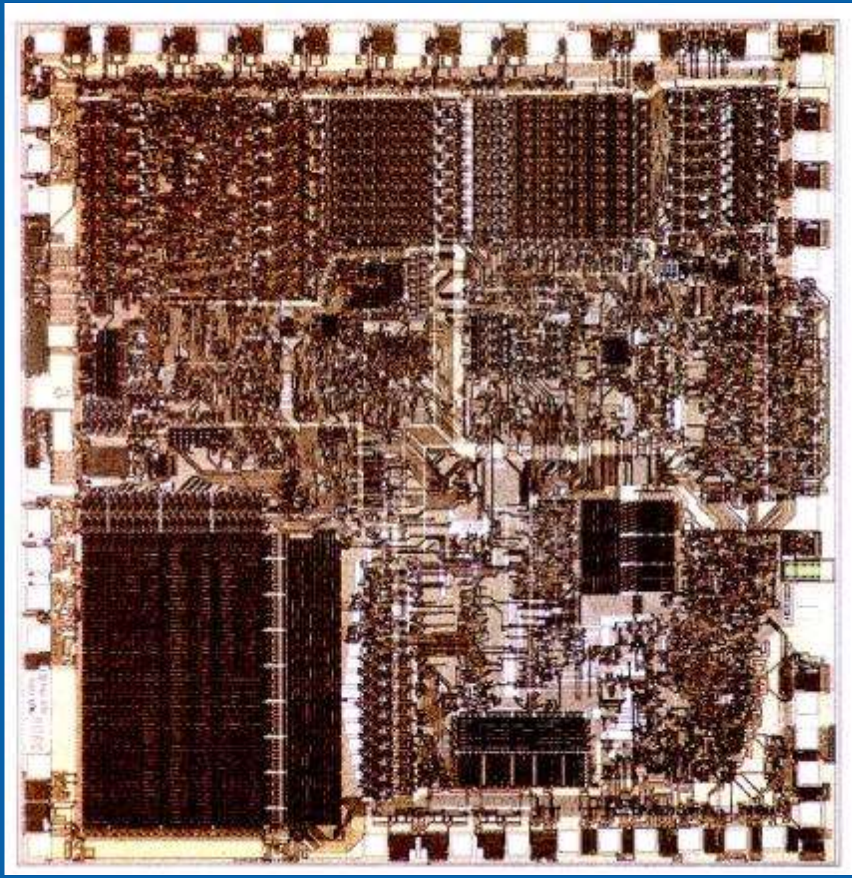
The First 25 Years of Microprocessors ~2000x Frequency & > 2000x Transistors

1974: 8080 Microprocessor



- The 8080 became the brain of the first personal computer--the Altair, allegedly named for a destination of the Starship *Enterprise* from the *Star Trek* television show. Computer hobbyists could purchase a kit for the Altair for $395.
- Within months, it sold tens of thousands, creating the first PC back orders in history
- 2 MHz
- 4500 transistors
- 6 µm

1978-79: 8086-8088 Microprocessor



- A pivotal sale to IBM's new personal computer division made the 8088 the brain of IBM's new hit product--the IBM PC.
- The 8088's success propelled Intel into the ranks of the *Fortune 500*, and *Fortune* magazine named the company one of the "Business Triumphs of the Seventies."
- 5 MHz
- 29,000 transistors
- 3 µm

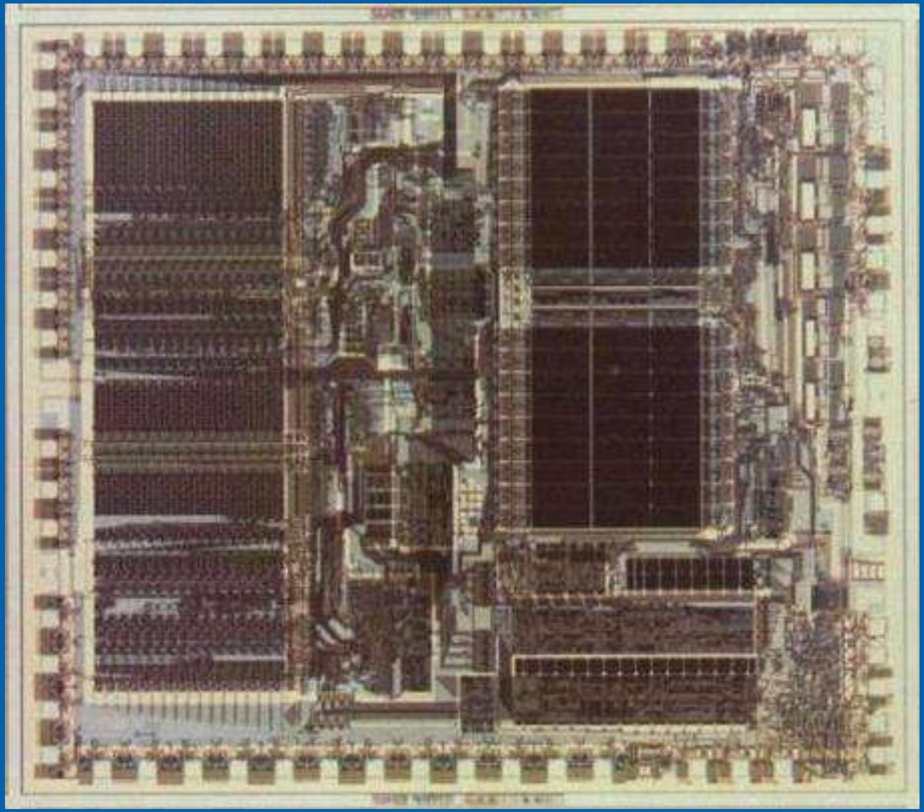
1981: First IBM PC

The IBM Personal Computer ("PC")

- PC-DOS Operating System
- Microsoft BASIC programming language, which was built-in and included with every PC.
- Typical system for home use with a memory of 64K bytes, a single diskette drive and its own display, was priced around $3,000.
- An expanded system for business with color graphics, two diskette drives, and a printer cost about $4,500.

“There is no reason anyone would want a computer in their home.” Ken Olsen, president Digital Equipment Corp (1977)

1979: Motorola 68000

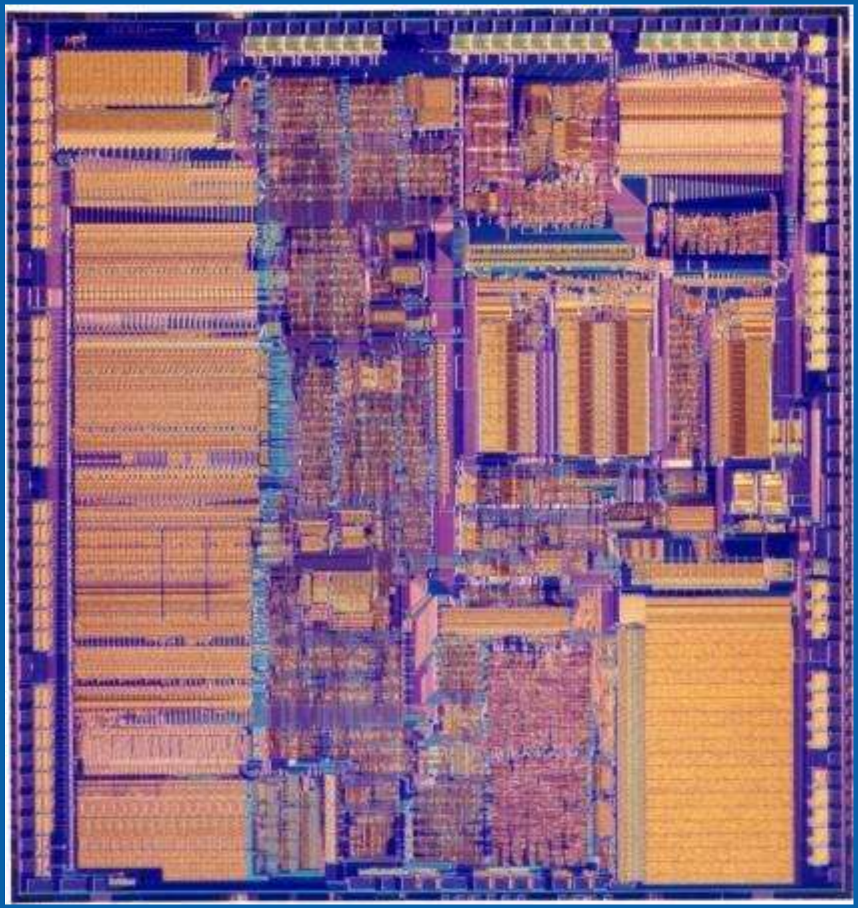


1984: Apple Macintosh

The 68000 became the dominant CPU for Unix-based workstations from Sun and Apollo

It was also used for personal computers such as the Apple Lisa, Macintosh, Amiga, and Atari ST

1985: Intel386™ Microprocessor



- The Intel386™ microprocessor featured 275,000 transistors--more than 100 times as many as the original 4004. It was a Intel’s first 32-bit chip.
- The 80386 included a paging translation unit, which made it much easier to implement operating systems that used virtual memory.
- 16 MHz
- 1.5µm

RISC vs CISC WARS

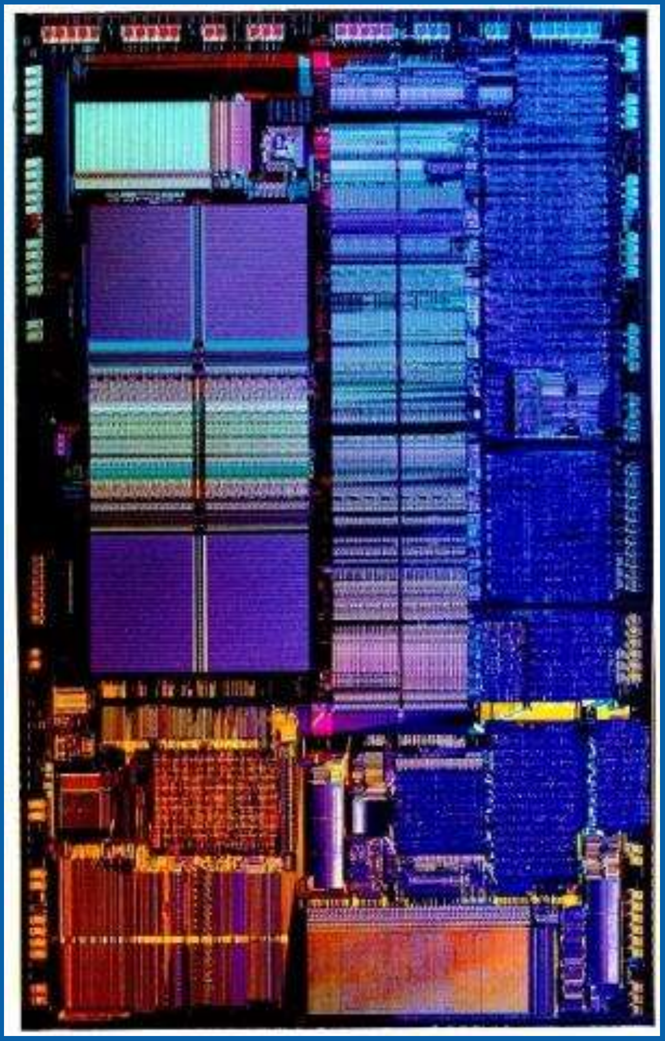
- Sun SPARC
- MIPS R2000, R3000, R4000, R6000, R10000
- HP PA-RISC
- IBM Power and Power PC
- DEC Alpha 21064, 21164, 21264

In 1987, the introduction of RISC processors based on Sun's SPARC architecture spawned the now famous RISC vs CISC debates. RISC processors from MIPS, IBM (Power, Power PC), and HP (PA-RISC) started to gain market share.

- RISC was "better" for in order designs
- Out of order microarchitectures leveled the playing field
- Semiconductor Technology and Volume Economics matter!
- PC Volumes and Pentium Pro design changed the industry

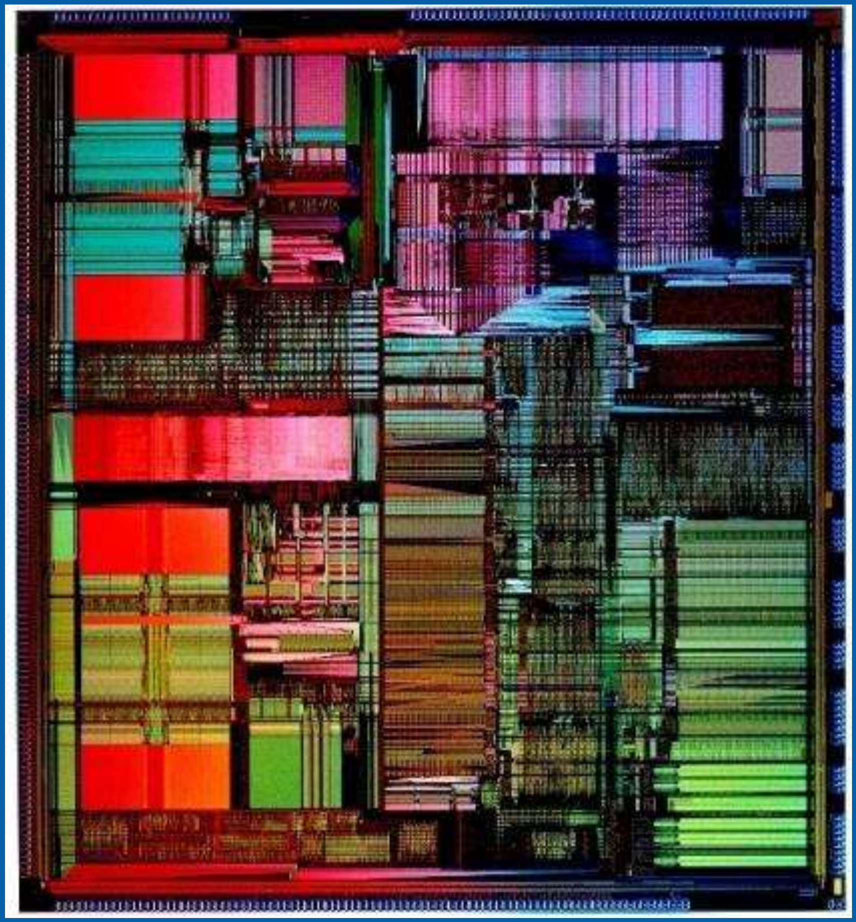
The difference between theory and practice is always greater in practice than it is in theory!

1989: Intel486™ DX CPU Microprocessor



- The Intel486™ processor was the first to offer a "large" 8KB unified instruction and data on-chip cache and an integrated floating-point unit.
- Due to the tight pipelining, sequences of simple instructions (such as ALU reg, reg and ALU reg, im) could sustain a single clock cycle throughput (one instruction completed every clock).
- 25 MHz
- 1.2 M transistors
- 1 µm

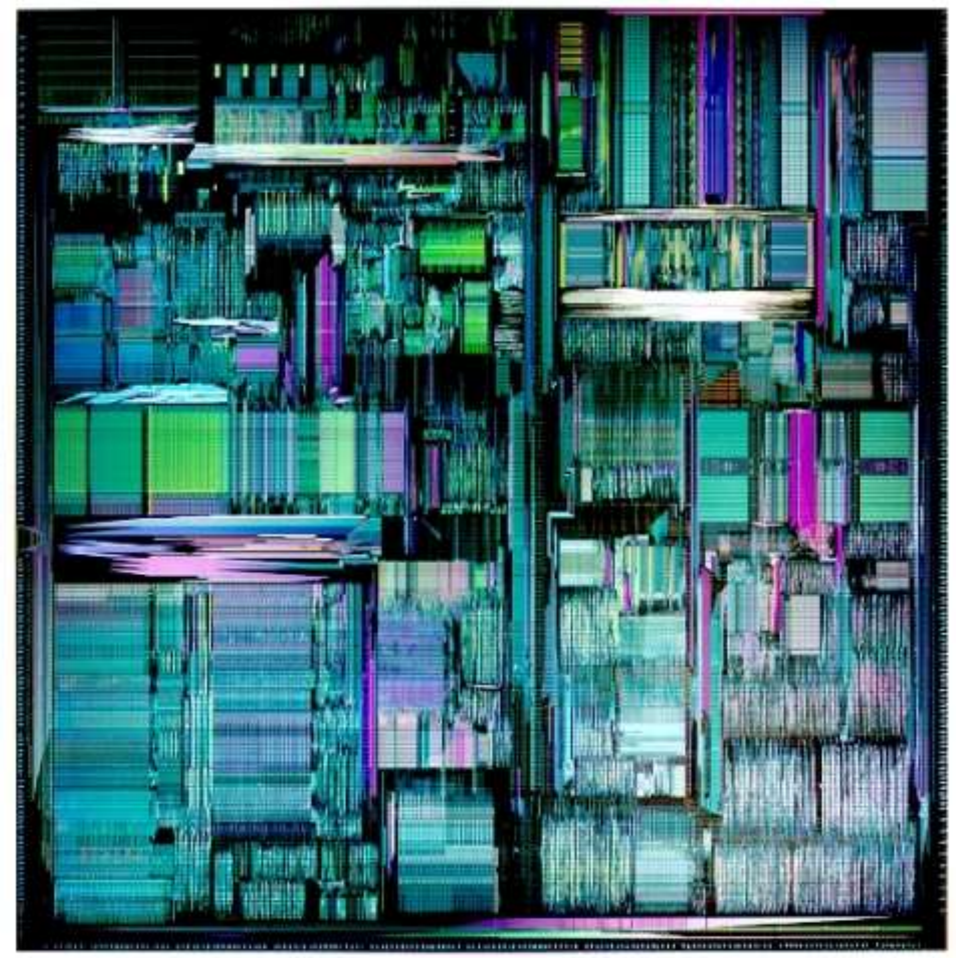
1993: Intel® Pentium® Processor



- The Intel Pentium® processor was the first superscalar x86 microarchitecture. It included dual integer pipelines, a faster floating-point unit, wider data bus, separate instruction and data caches
- Famous for the FDIV bug!
- 22 March 1993
- 66 MHz
- 3.1 M transistors
- 0.8 μm

PC Performance Gets Interesting!

1995: Intel® Pentium® Pro Processor

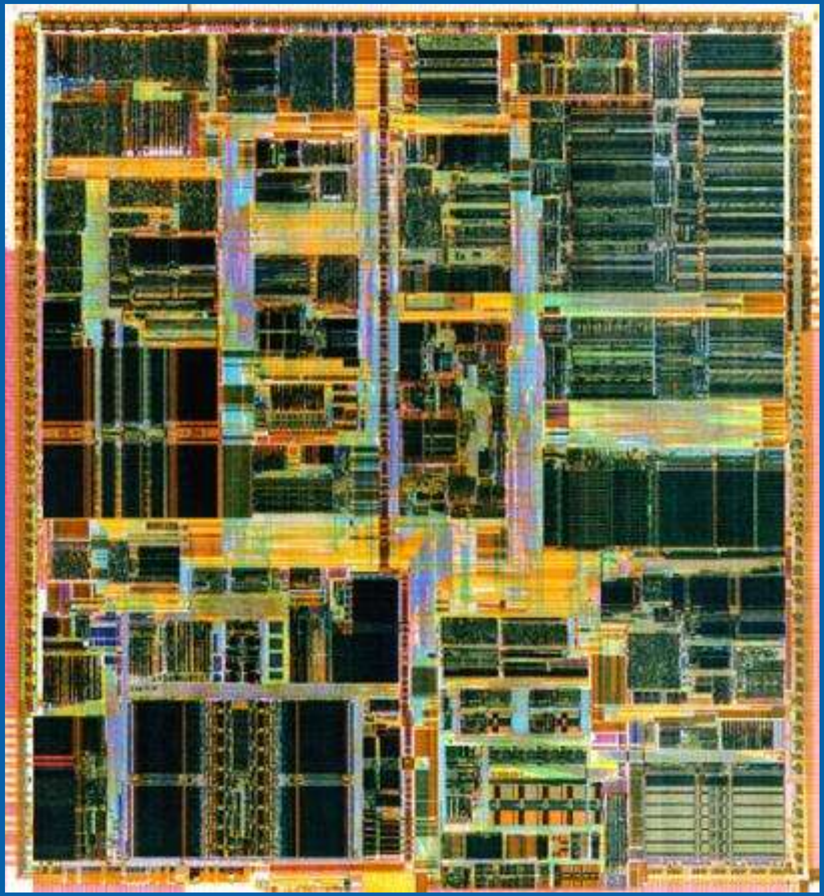


- Intel® Pentium® Pro processor was designed to fuel 32-bit server and workstation applications. Each processor was packaged together with a second L2 cache memory chip on the back-side bus.
- 5.5 million transistors.
- 1 November 1995
- 200 MHz
- 0.35µm
- 1st x86 to implement out of order execution
- Front side bus with split transactions
- The P6 micro-architecture lasted 3 generations from the Pentium Pro to Pentium III
- The Pentium Pro processor slightly outperformed the fastest RISC microprocessors on integer benchmarks, but floating-point performance was significantly lower

X86 Gets Ready for Workstation & Server Markets

1997-98: Intel® Pentium® II Processor

Klamath
Deschutes

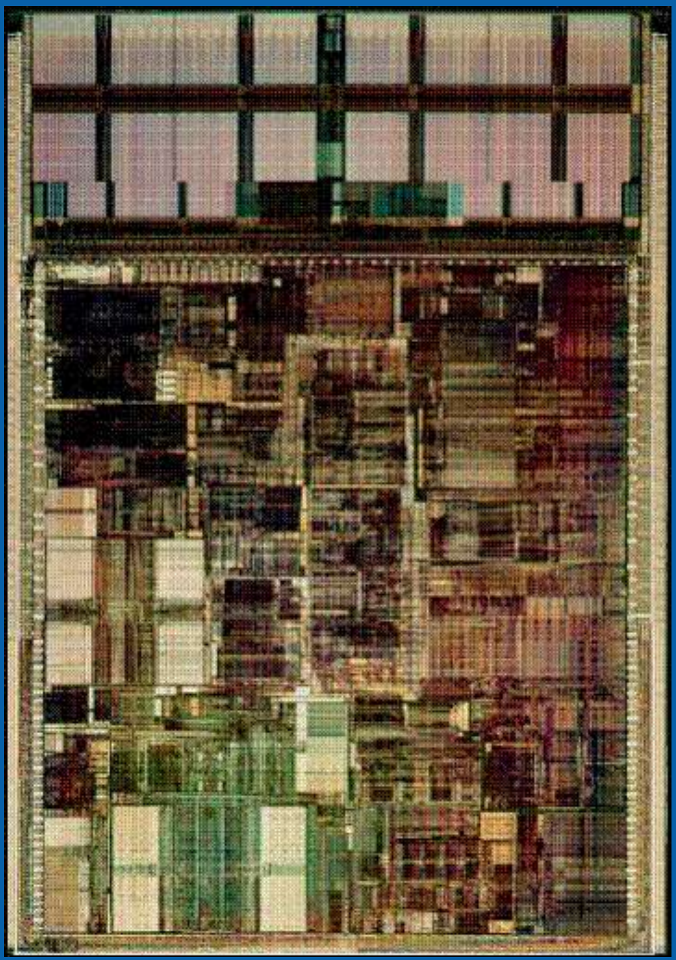


- The 7.5 million-transistor 0.35 µm Pentium II processor was introduced with 512 KB L2 cache in external chips on the CPU module clocked at half the CPU's 300 MHz frequency in a "Slot 1" SECC module.
- **1998: Intel Pentium II Xeon** processors (0.25 µm Deschutes) were launched with a full-speed custom 512 KB, 1 MB, or 2 MB L2 cache using a larger Slot 2 to meet the performance requirements of mid-range and higher servers and workstations

Driving PC Technology Higher

Mendocino

1998: Intel® Celeron® Processor

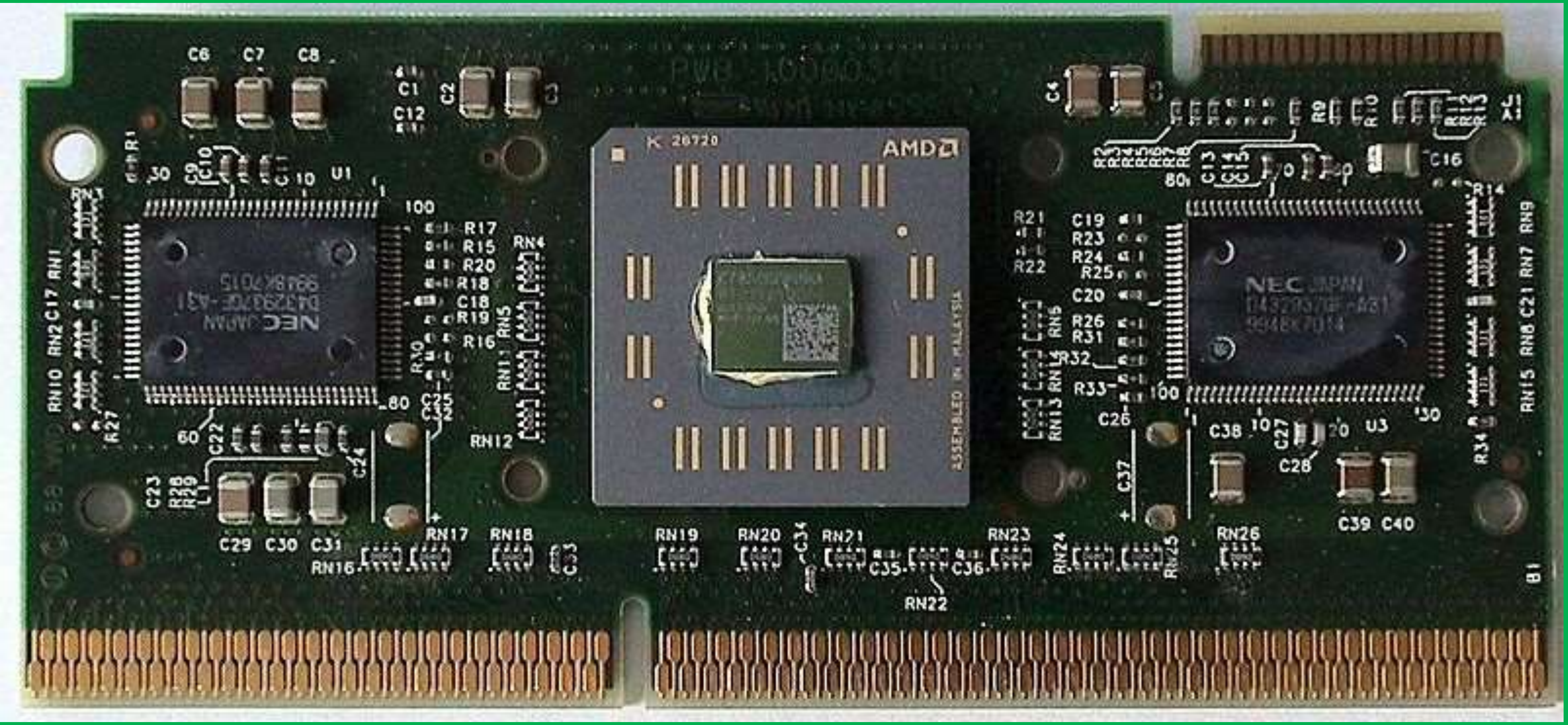


- The Intel® Celeron® processors were designed for the sub $1000 Value PC market segment in response to Cyrix 6x86 (M1)
- The first Celeron processor (Covington) in April 1998 was just a 266 MHz Pentium II without a L2 cache
- Mendocino: First x86 with integrated L2 cache -128 KB
- 19M transistors
- 300 MHz
- 0.25µm
- 24 August 1998

Making PCs More Affordable

Clock Frequency Uber Alis

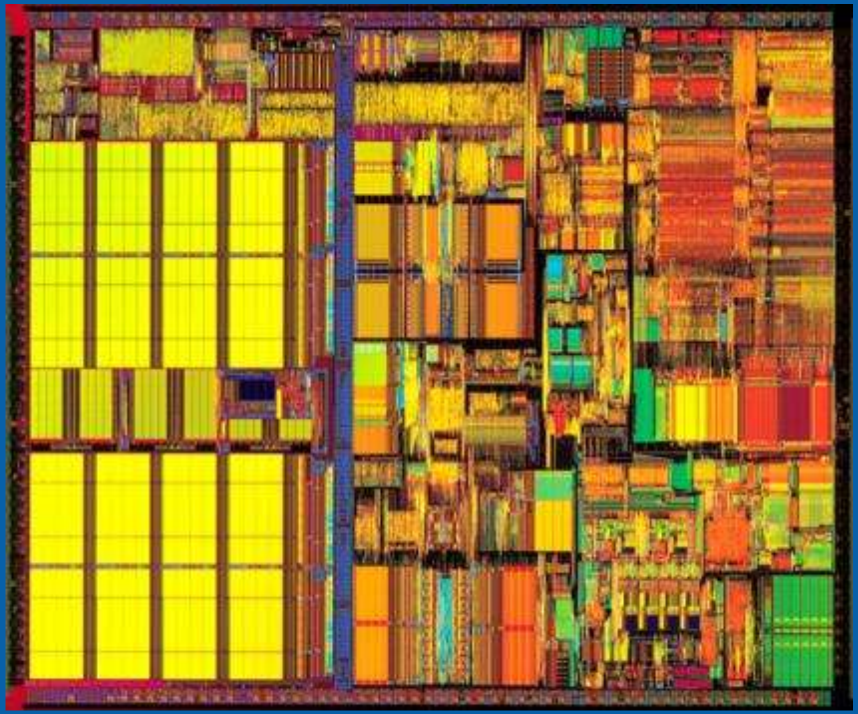
1999: AMD Athlon



Won the Race to 1 GHz

Coppermine

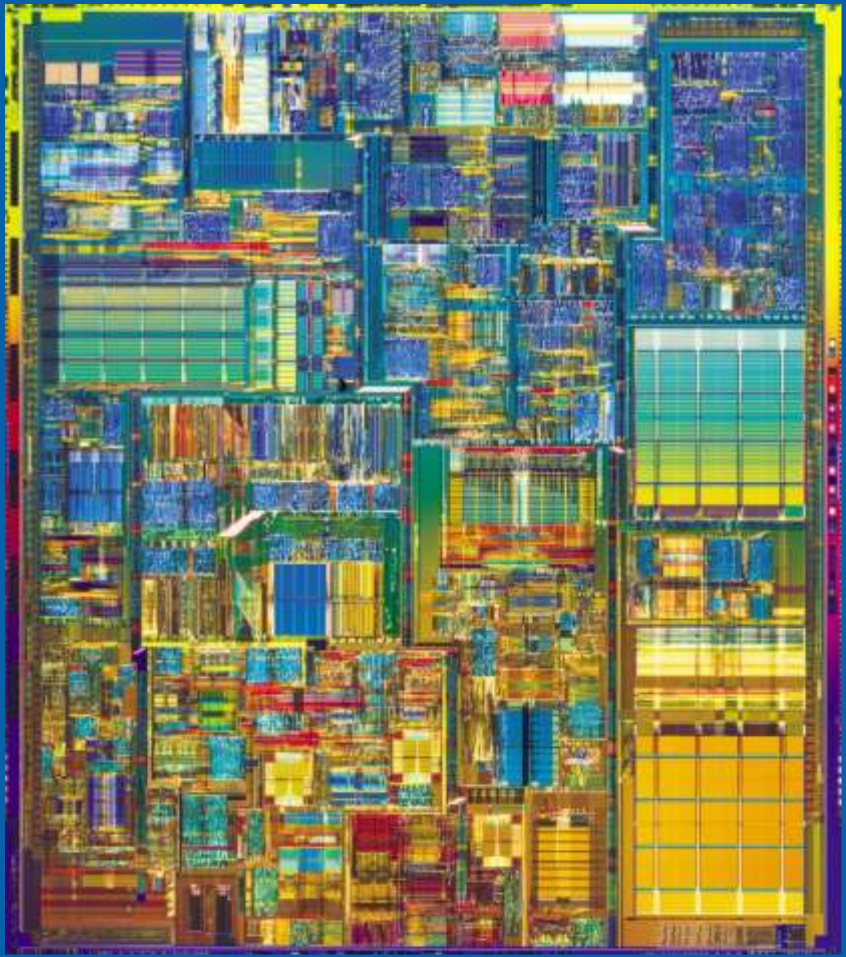
1999: Intel® Pentium® III Processor – 0.18μm



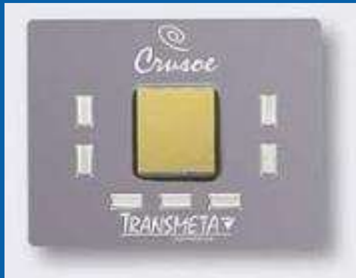
- 25 Oct 1999
- Integrated 256KB L2 cache
- 733 MHz
- 28 M transistors
- 1st Intel microprocessor to hit 1 GHz on 8-Mar-2000, a few days after AMD Athlon!

Willamette

2000: Intel® Pentium® 4 Processor – 0.18µm



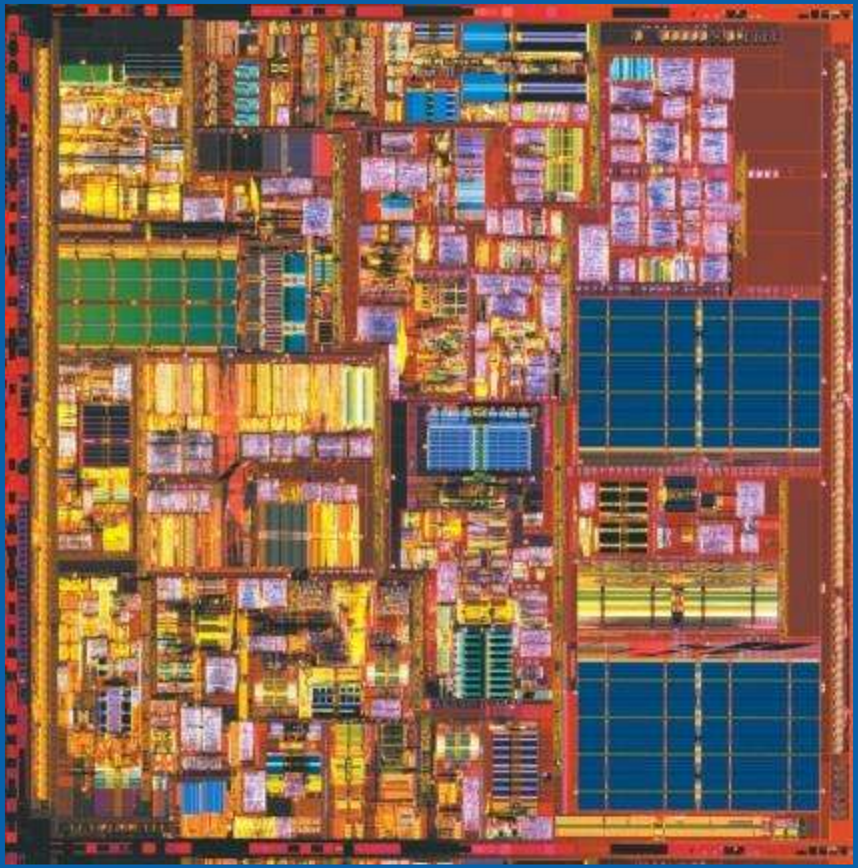
- The Intel® Pentium® 4
- processor's initial speed
- was 1.5 GigaHertz.
- 20 Nov 2000
- 256K integrated L2 cache
- Double clocked "Fireball" inner core
- Deep 20 stage pipeline
- 100 MHz quad pumped
- bus
- 42 M transistors
- Hit 2 GHz on 27 Aug 2001
- ~55 Watts
- No Mobile Pentium 4!



Desktop Processors Not Mobile Friendly!

Northwood

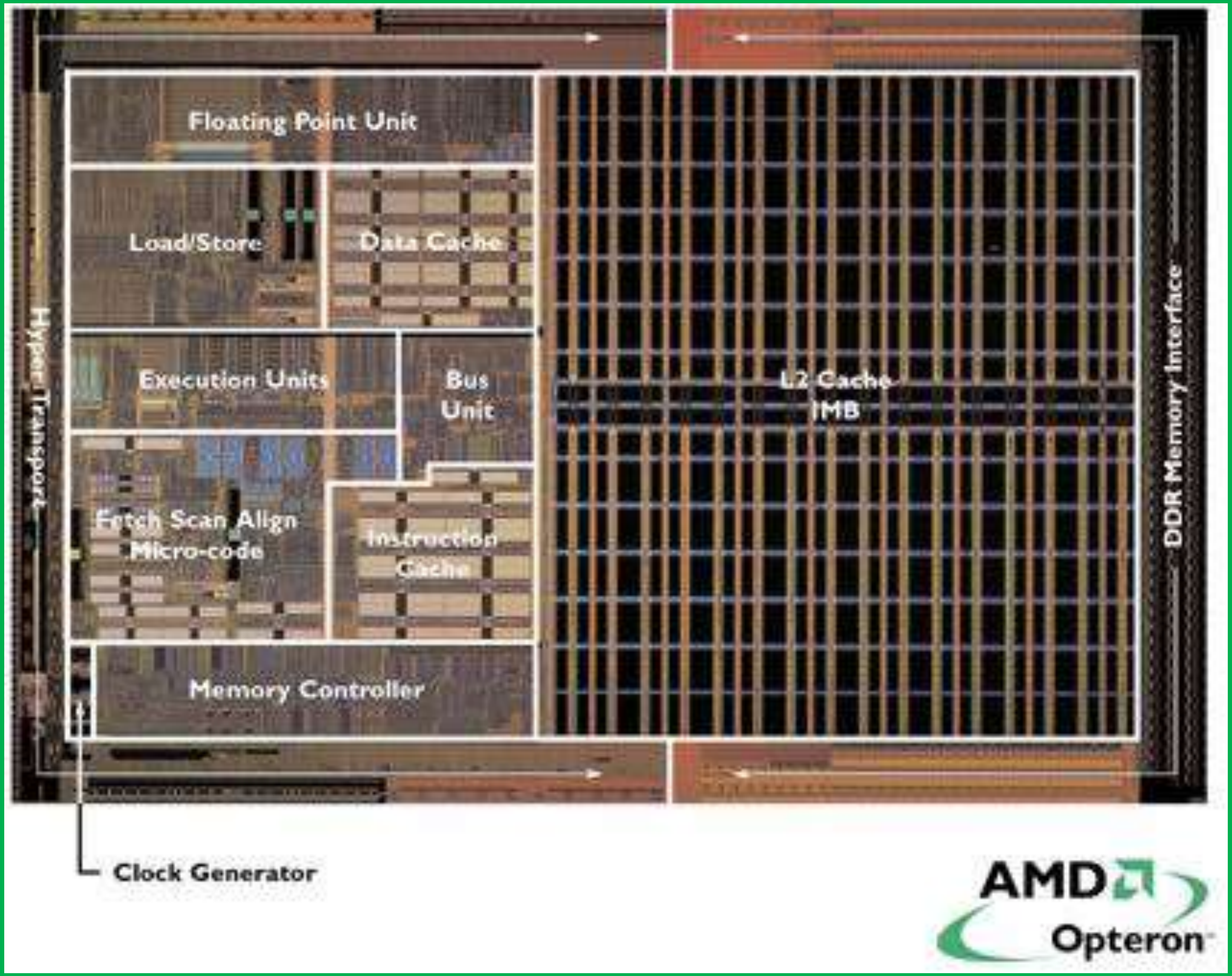
2001: Intel® Pentium® 4 Processor – 0.13µm



- 27 August 2001
- 55 million transistors
- 2 GHz
- 512KB L2 cache
- In 2002 Intel released a Xeon branded CPU, codenamed "Prestonia" with Intel's Hyper-Threading Technology
- 14 Nov 2002: 3.06 GHz
- 23 June 2003: 3.2 GHz

Simultaneous Multi Threading Improves Throughput Performance

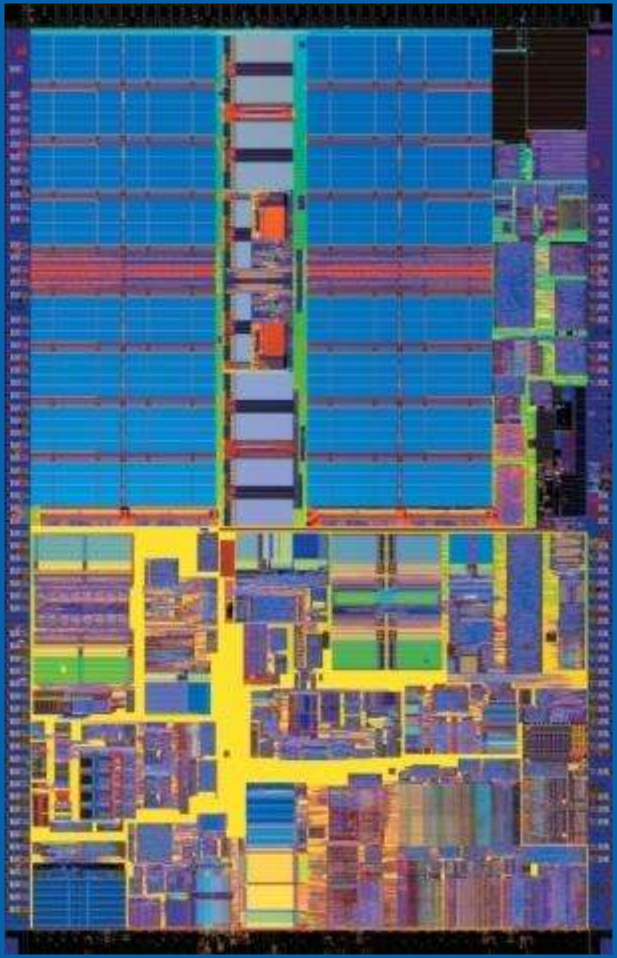
2003: AMD Opteron – First 64 bit x86



64 bits Comes to PC Platforms

Banias

2003: Intel® Pentium® M Processor



- The first Intel® Pentium® M processor, the Intel® 855 chipset family, and the Intel® PRO/Wireless 2100 network connection were the three components of Intel® Centrino™ mobile technology, with built-in wireless LAN capability and breakthrough mobile performance. It enabled extended battery life and thinner, lighter mobile computers.
- Dedicated Processor Optimized for Notebook Segment
- 12 March 2003
- 130 nm
- 1.6 GHz
- 77 million transistors
- 1 MB integrated L2 cache

The move away from core frequency to performance begins!

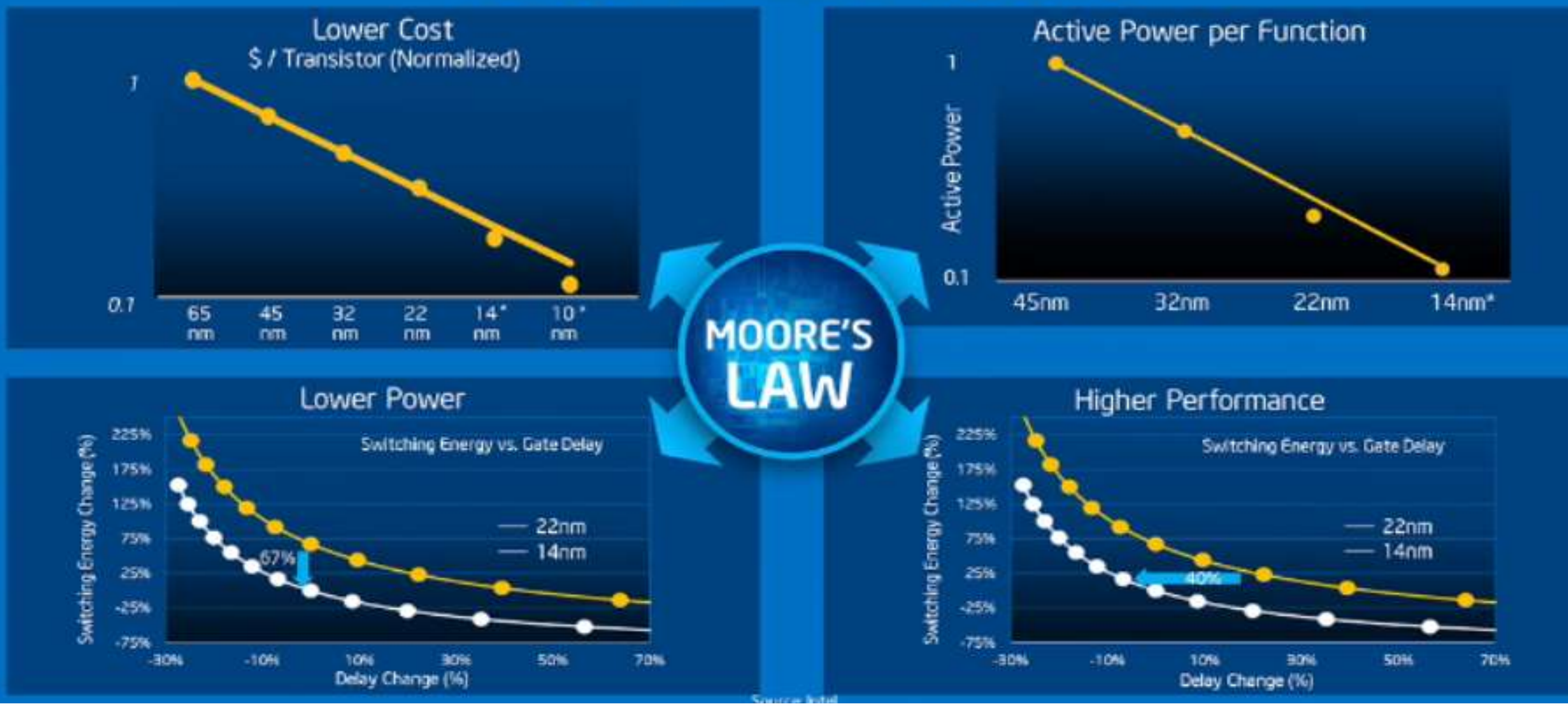
Post Dennard Scaling

THE MULTICORE ERA

NEW DEVICE STRUCTURES

ENERGY EFFICIENCY

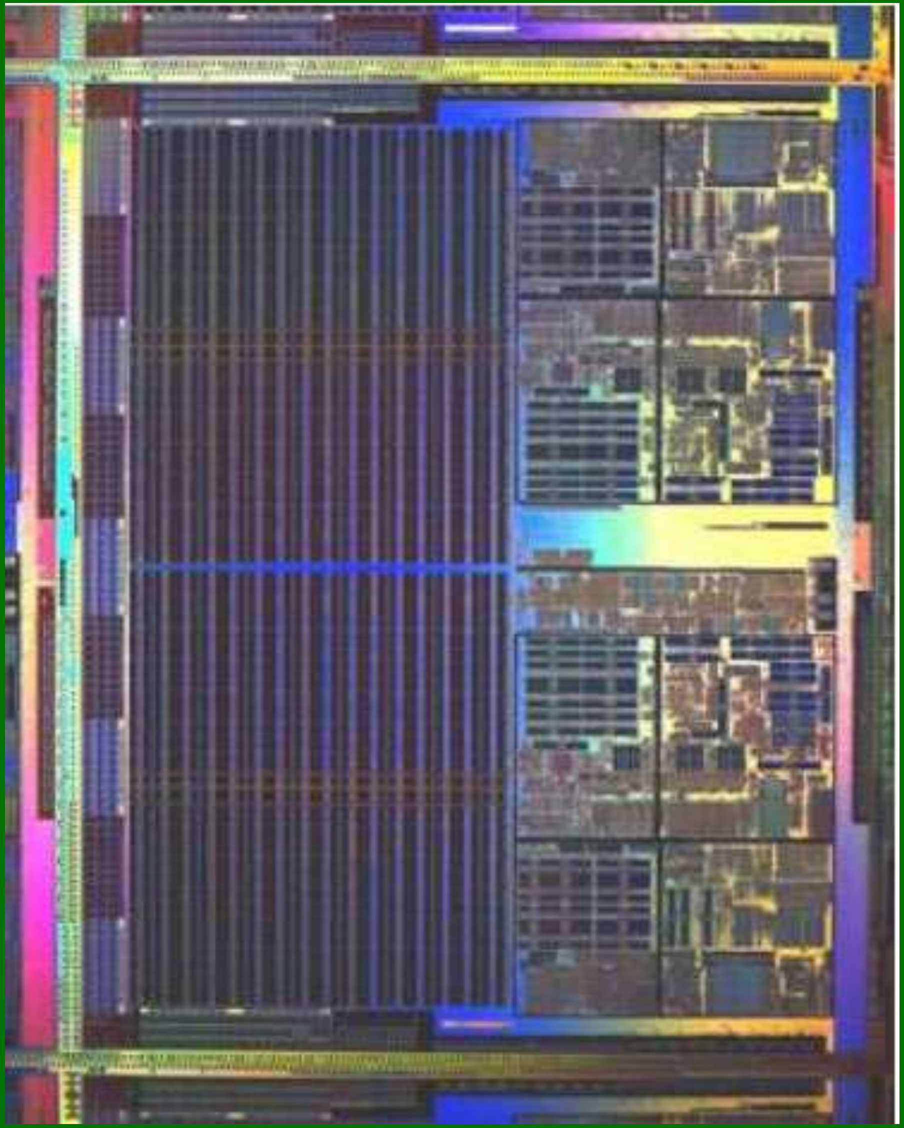
Getting Benefits of Moore's Law Across all Value Vectors



2001	2003	2005	2007	2009	2011

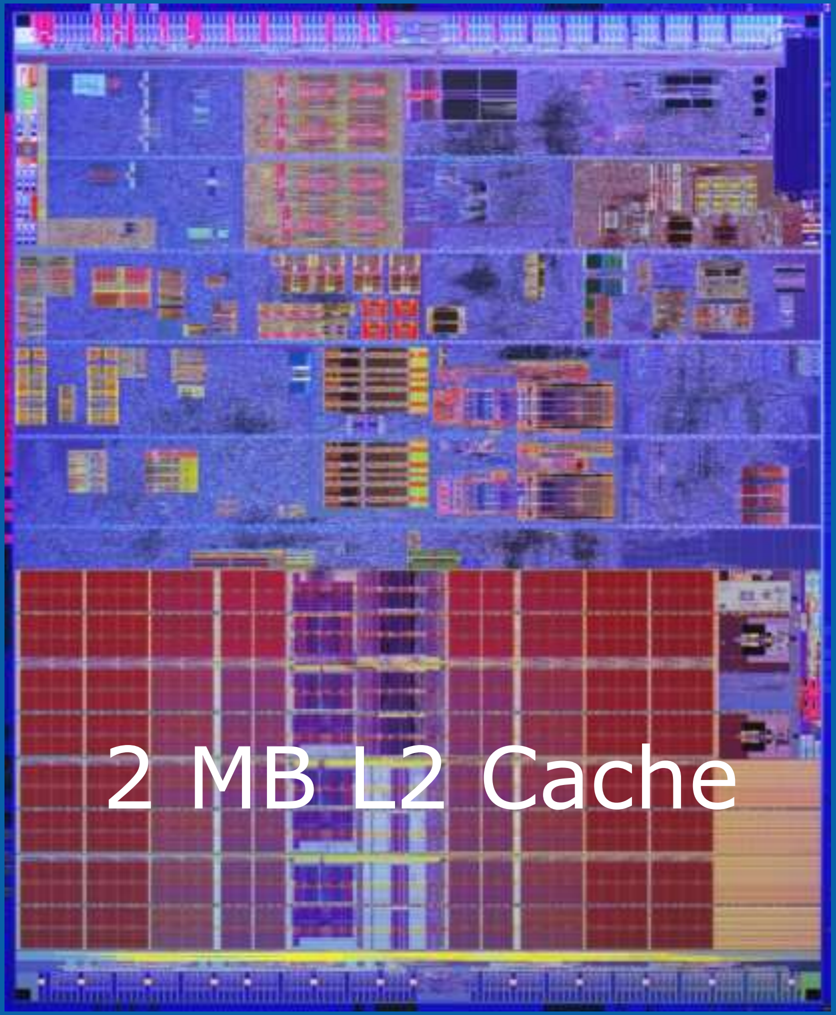
Something New Needed Every Two Process Generations to Keep Moore's Law Going

2005: First Dual Core Opteron



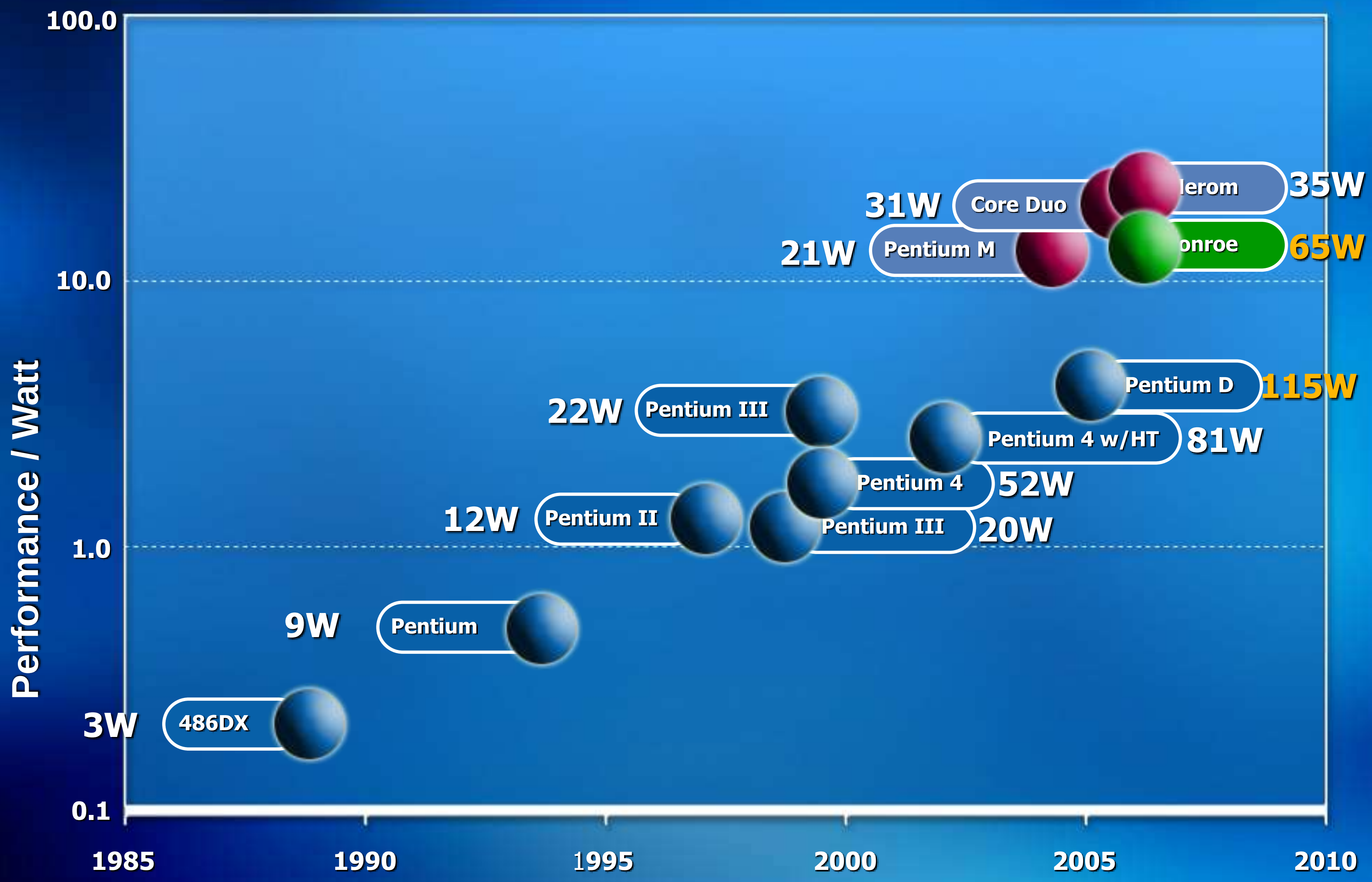
Beginning of the Multi-Core Era!

2005: Last Netburst Microarchitecture Core (65nm)



Last of the Power Hungry Speed Demons!

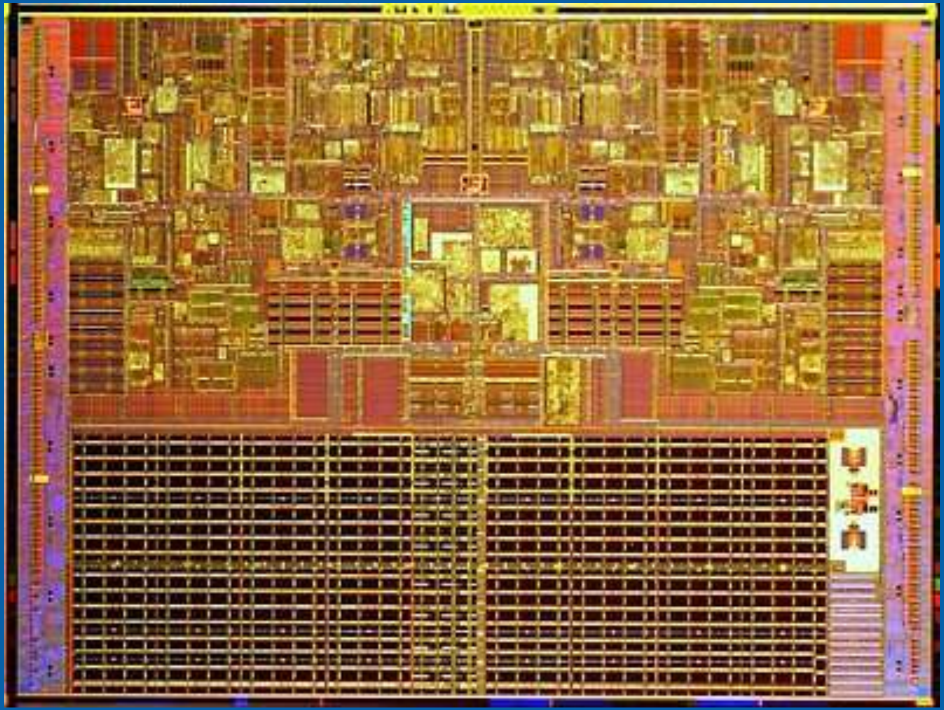
Increasing Energy Efficiency



Specint_rate2000; source: Intel; some data estimated.

Yonah

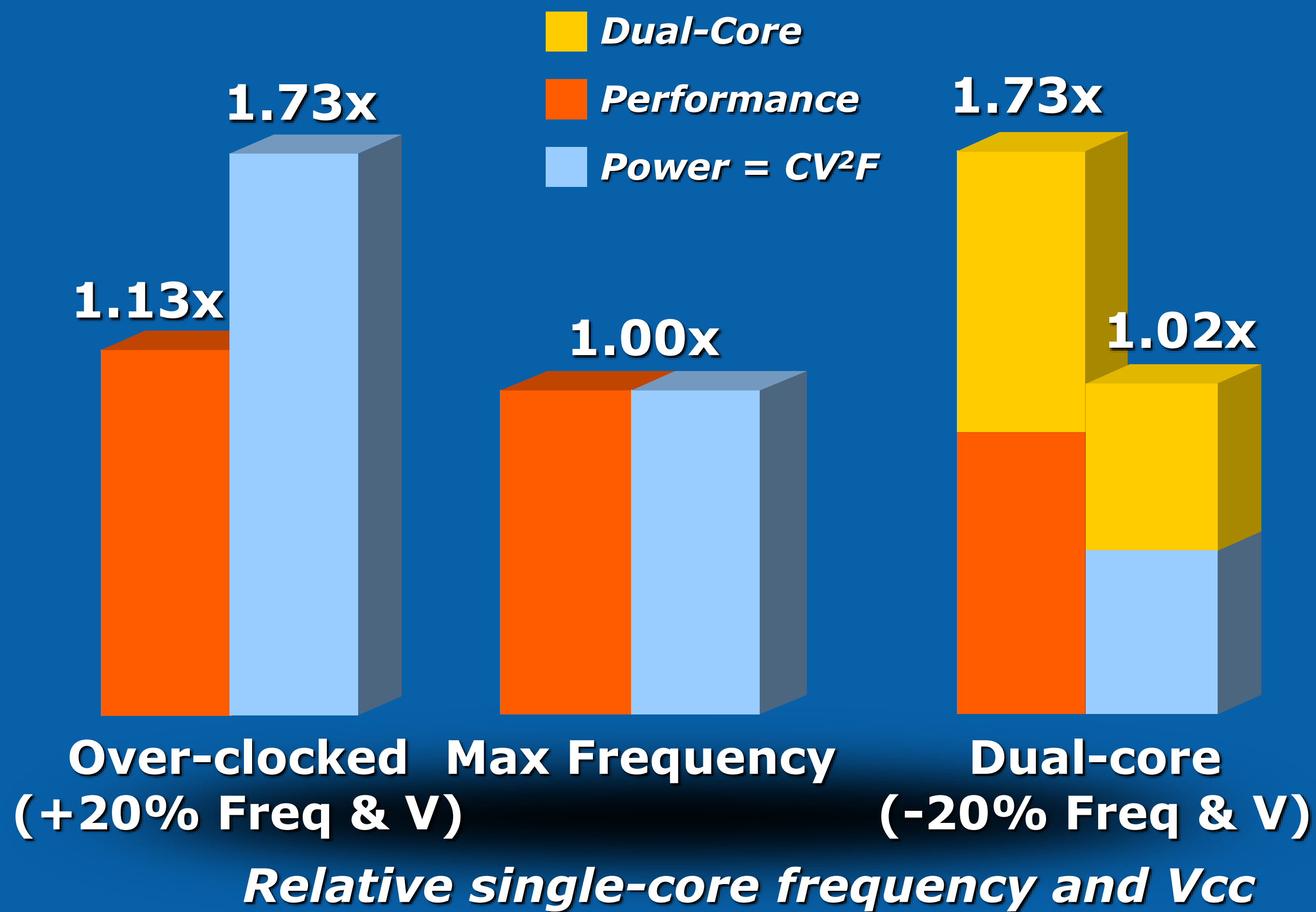
2006: Intel's 1st Monolithic Dual Core



- January 2006
- Intel® Core^{TM} Duo Processor
- 90 mm^2
- 151M transistors
- 65 nm
- First Intel processor to be used in Apple Macintosh Computers

The Convergence to Multiple Mobile Cores Begins Finally!

Why Multi-Core? Energy-Efficient Performance!



- End of Dennard Scaling
- Instruction Level Parallelism harder to find
- Increasing single-stream performance often requires non-linear increase in design complexity, die size, and power

Moore's Law Enables Microprocessor Advances

Chatting with Gordon Moore
http://www.youtube.com/watch?v=xzxpO0N5Amc

1.0µm 0.8µm 0.6µm 0.35µm 0.25µm 0.18µm 0.13µm 90nm 65nm

Intel 486™ Processor

Pentium® Processor

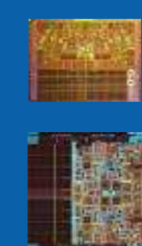
Pentium® II/III Processor

Pentium® 4 Processor

Intel® Core™ Duo Processor

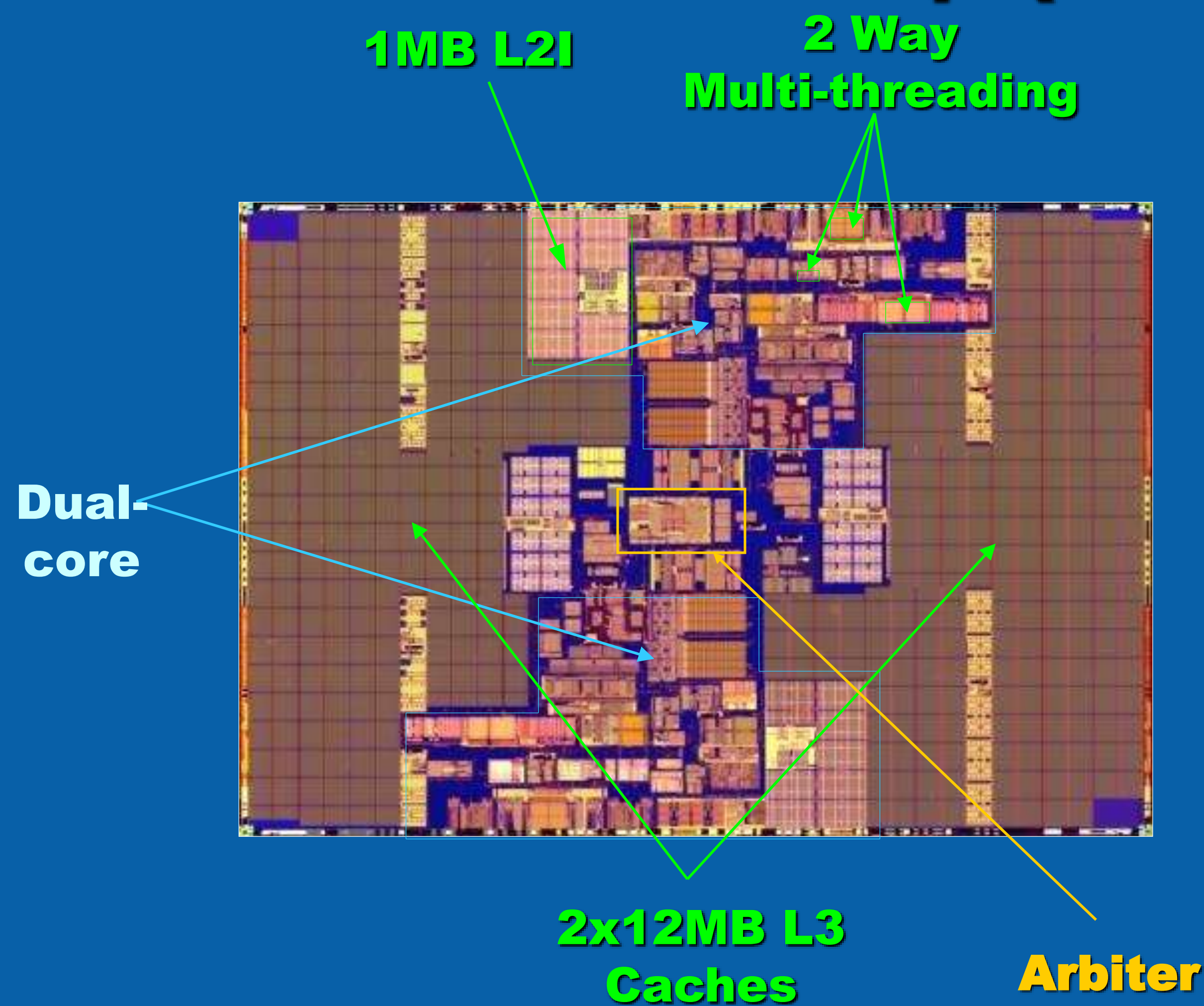
Intel® Core™ 2 Duo Processor

New Designs serve High End first and waterfall to more mainstream segments as die size decreases in subsequent nodes



Source: Intel

2006: Itanium 2: First Billion Transistor Dual Core Chip (90nm)

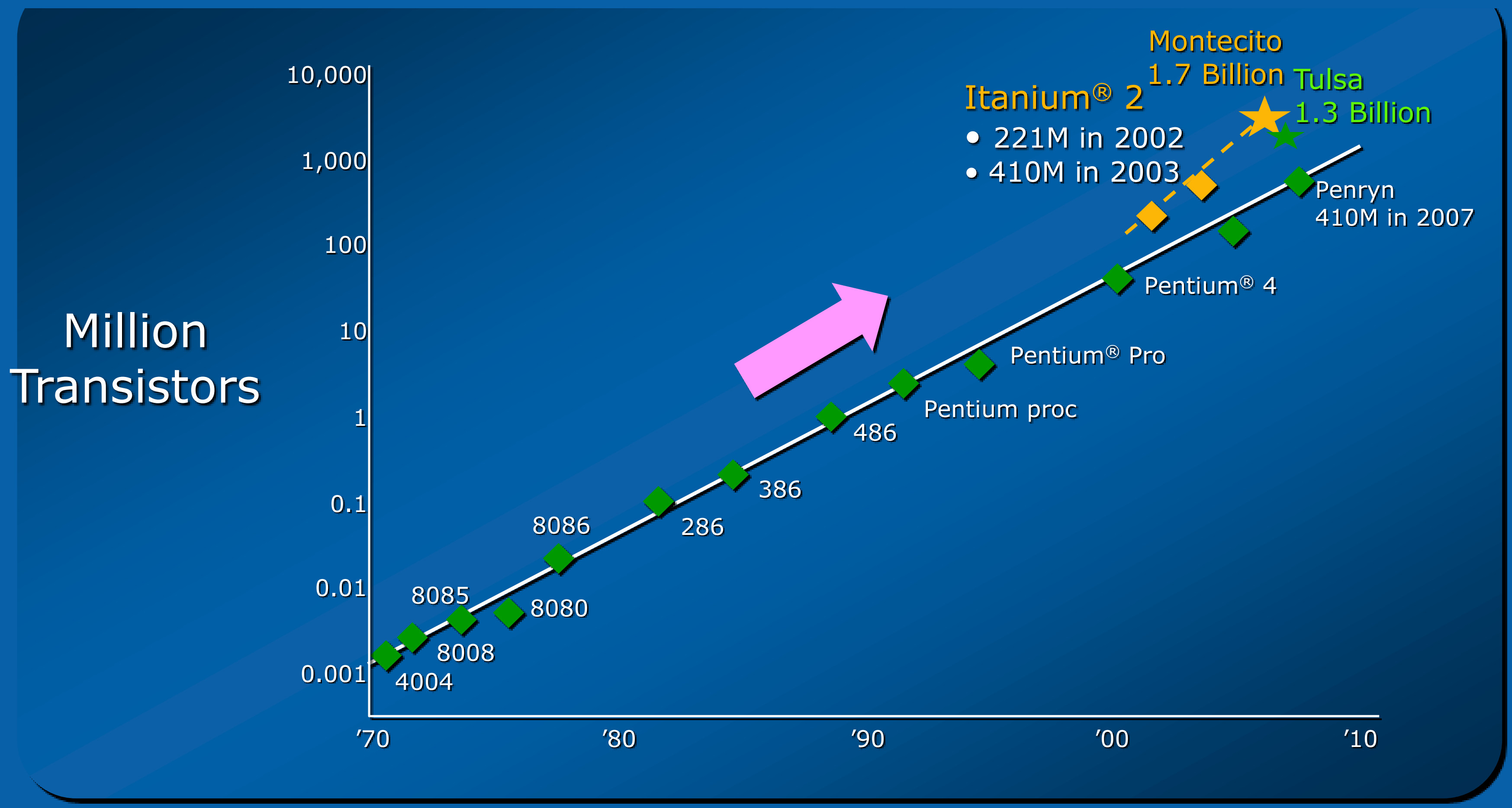


1.72 Billion Transistors
(596 mm²)

From 2300 to >1Billion Transistors

In < 40 Years of Moore's Law

Moore's Law video at http://www.cs.ucr.edu/~gupta/hpca9/HPCA-PDFs/Moores_Law_Video_HPCA9.wmv



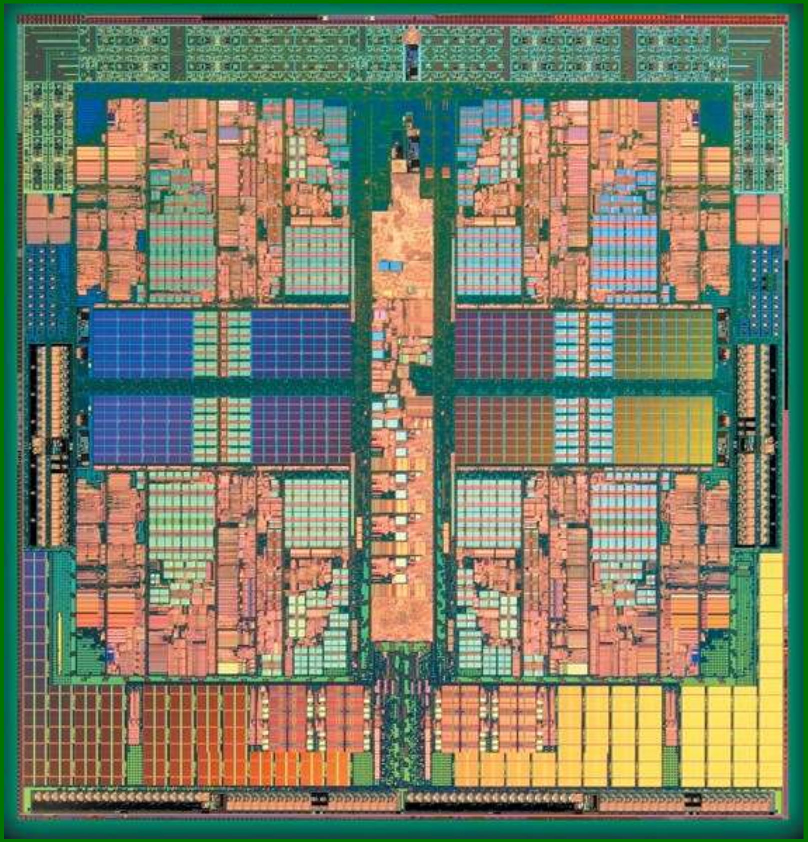
More than 1 Billion Transistors in 2006!

Multi-Core Era

Who Has The Most Cores?

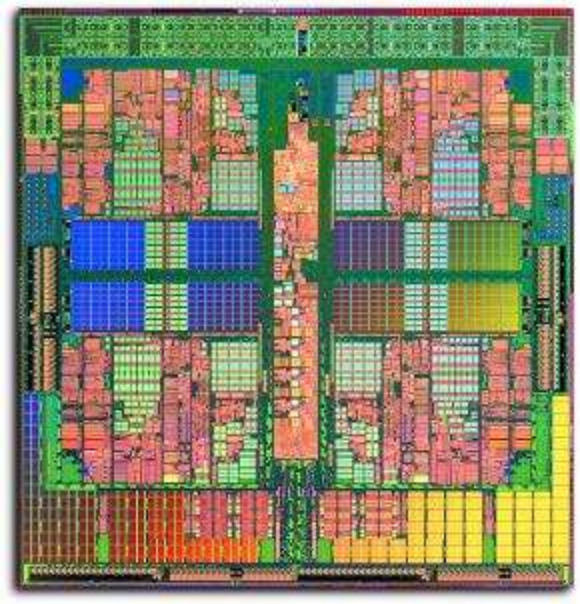
4 is Better Than 2!

2007: AMD Barcelona
First Monolithic x86 Quad Core

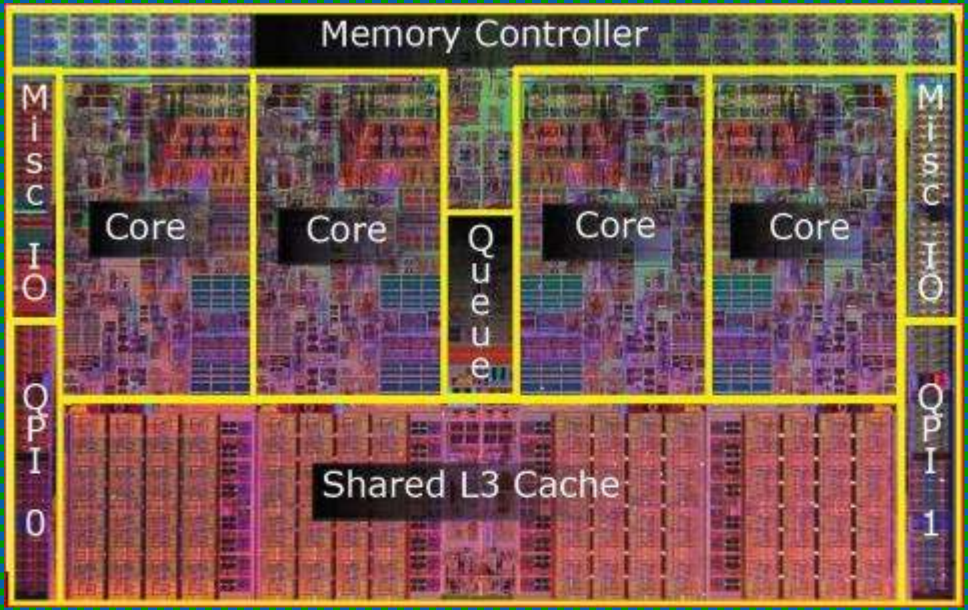


283mm² design with 463M transistors to implement four cores and a shared 2MB L3 cache in AMD’s 65nm process

2008-9: Performance Race Gets Serious With Quad Core



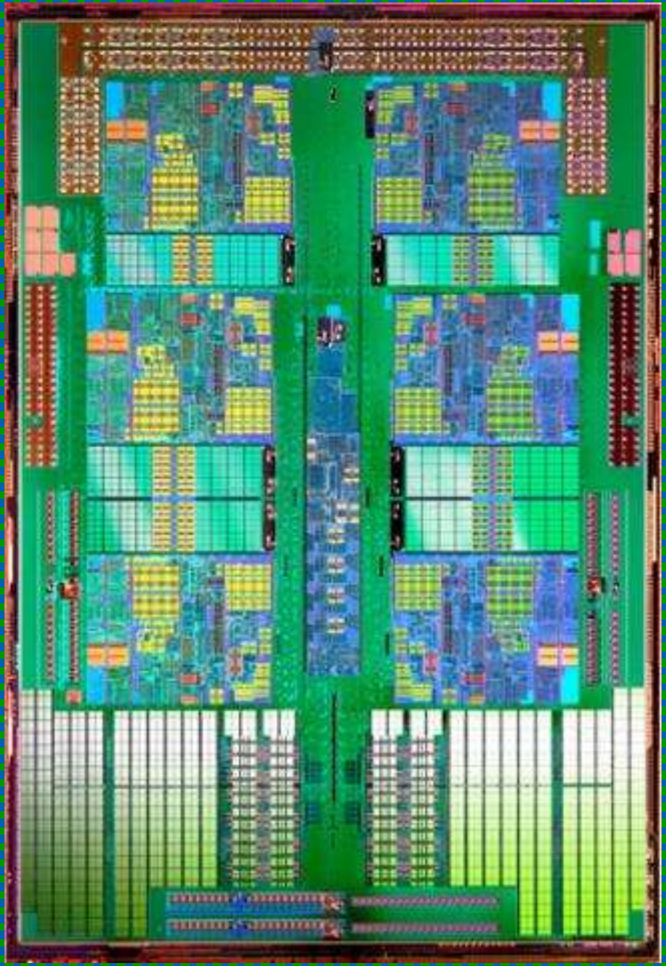
AMD Barcelona



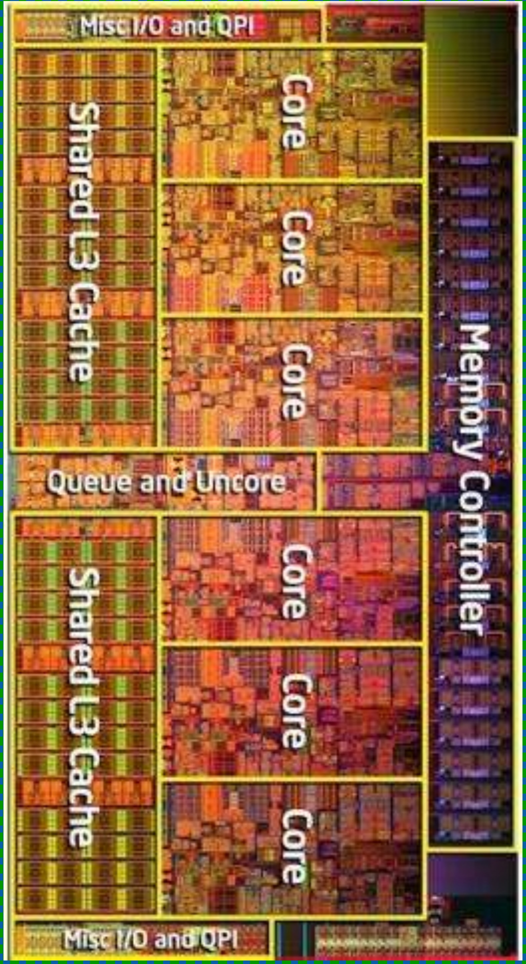
Intel Nehalem

Intel finally integrates Memory Controller and abandons shared Front Side Bus

Six Cores



2009: AMD Istanbul



2010: Intel Westmere

Mobile Computing Era

SIZE MATTERS

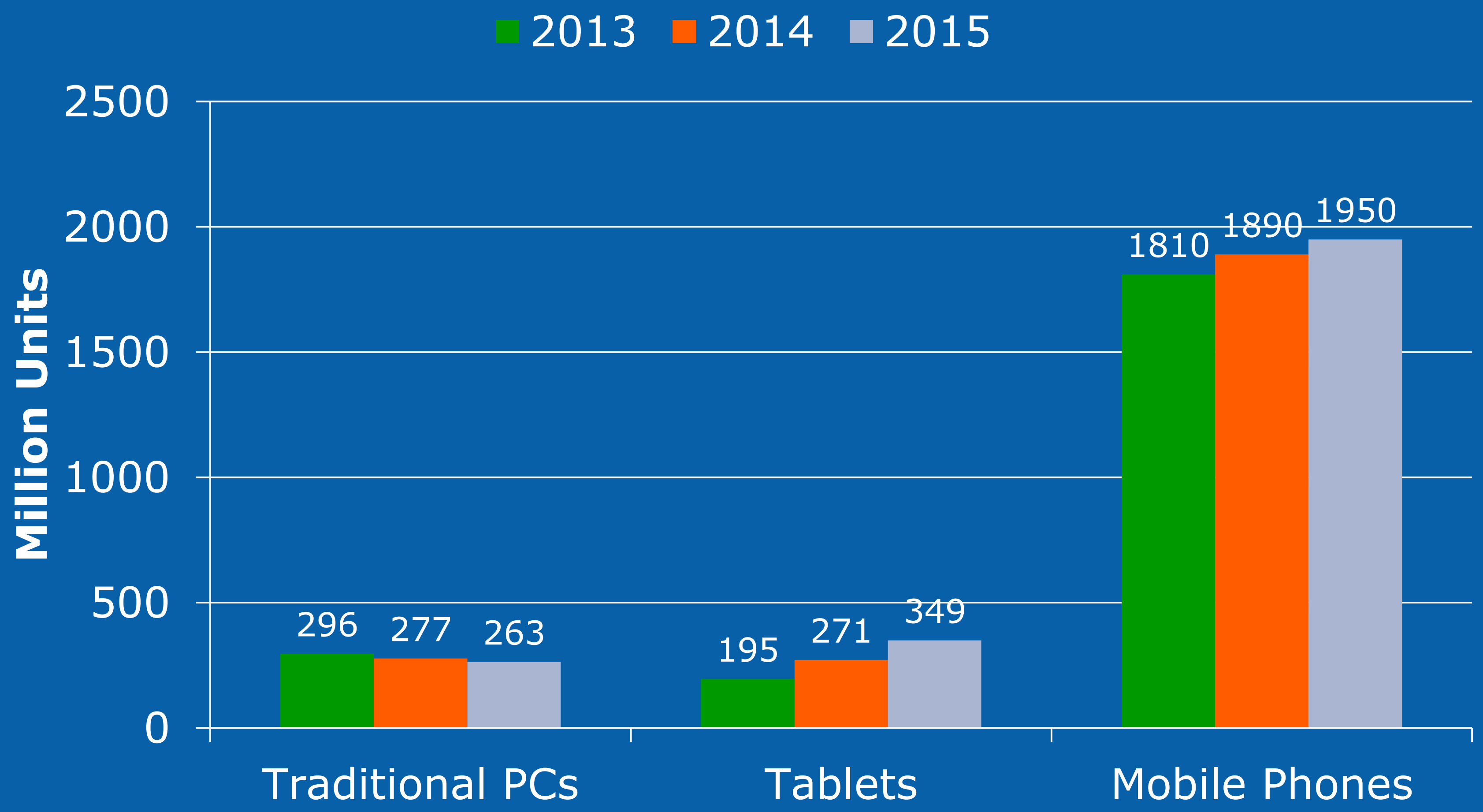
SMALL & LIGHT

LOW POWER

The Smart Phone Era
Is Redefining Computing

"The phone in your pocket will be as much of a computer as anyone needs".
– Dr. Irwin Jacobs, 2000

PC Market Shift



Source: www.pctoday.com

Continued smartphone momentum

>8B

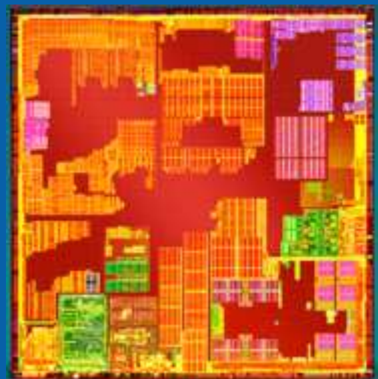
cumulative smartphone unit shipments forecast between 2014–2018

~2x

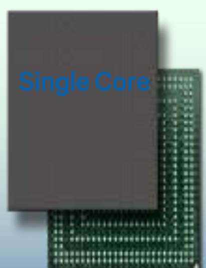
smartphone installed base 2018 vs. 2014

Source: Gartner Sep. '14

Qualcomm Processor Progression

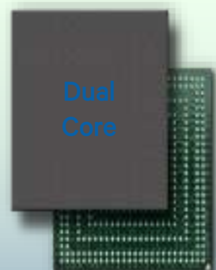


2011



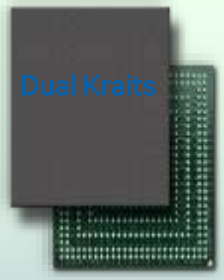
First 1GHz Single Core

2012

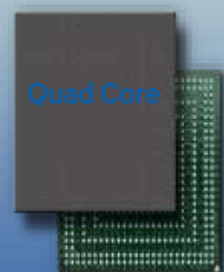


Dual Core

2013



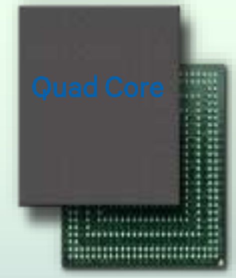
- Dual “Krait” CPUs
- Adreno GPU
- 28nm process
- Faster memory
- Industry leading modem
- Integrated Connectivity
- GPS



Quad Core

- Quad Core A5 CPUs
- Adreno GPU
- LPDDR2
- DSDS and DSDA
- 720p capture and playback
- Up to 8 Megapixel camera

2015



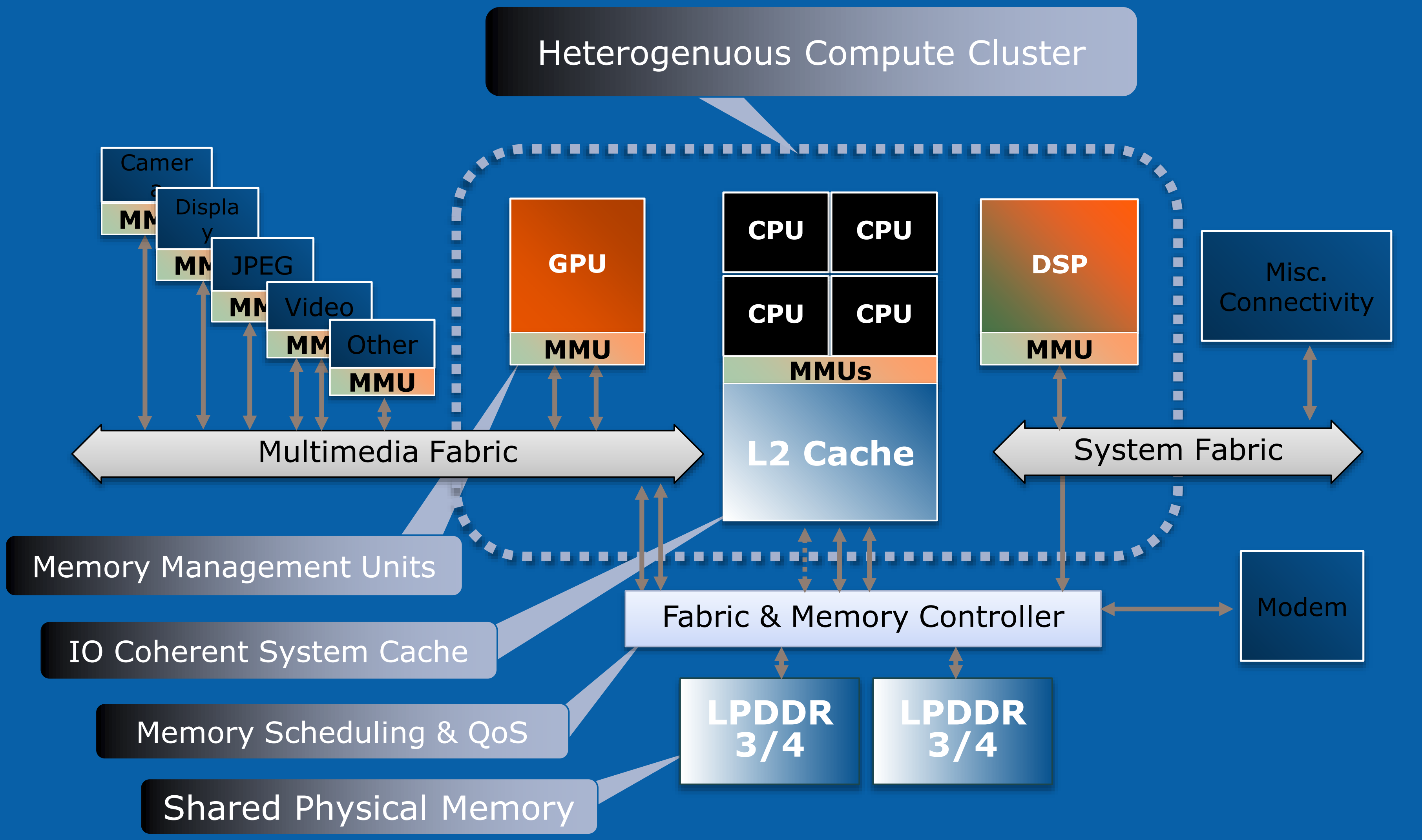
64 bit Quad + quad Core (20 nm)

- Quad Core A57 CPUs
- Quad Core A53 CPUs
- Adreno 430 GPU
- Hexagon™ V56 DSP
- Integrated X10 LTE
- DSDS and DSDA
- 4K capture and playback
- Up to 55 MP Dual ISP camera

The future is more about Heterogenuous Computing Cores

Qualcomm snapdragon

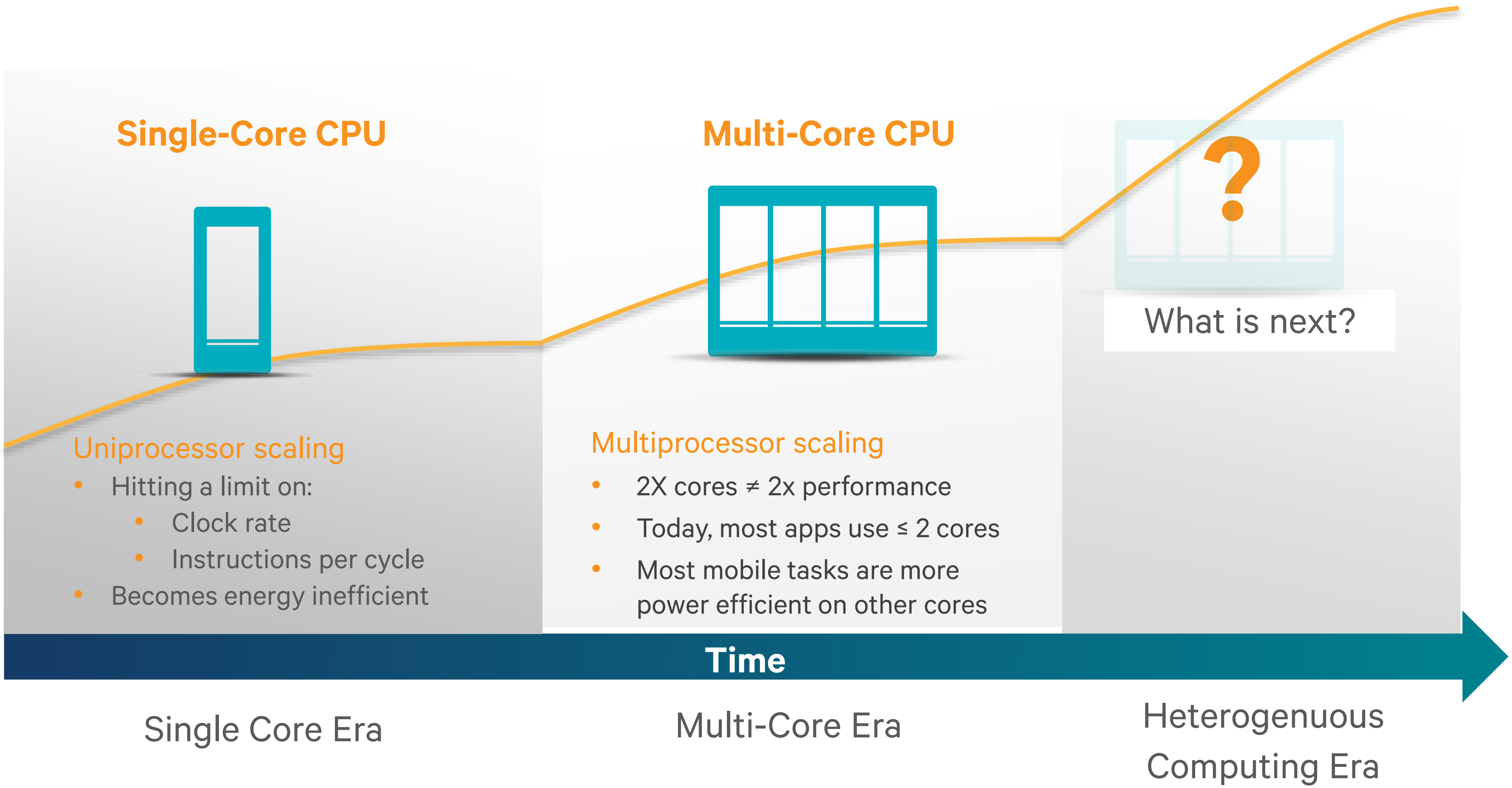
Representative System Architecture



Smartphones demand more processing horsepower

While consuming little power

CPU scaling is reaching diminishing returns

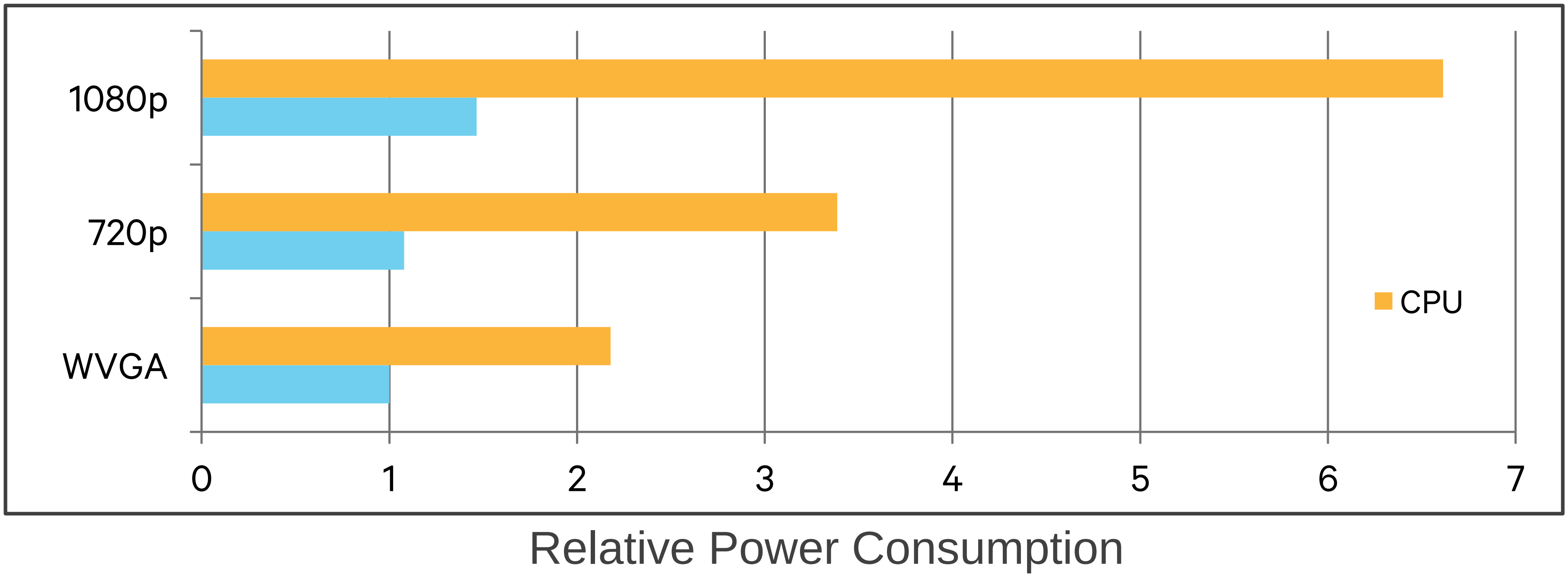


CPU takes a small area in modern mobile SoCs

15%

Most mobile tasks are more power efficient on other cores

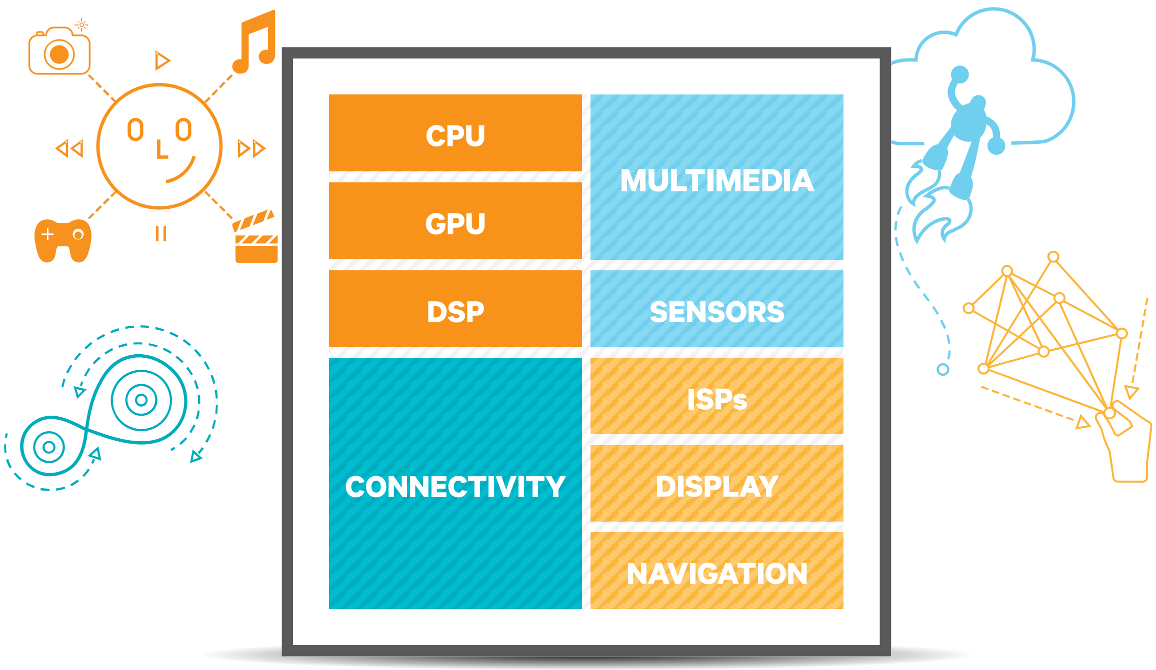
Specialized hardware can be an order of magnitude more power-efficient than the CPU



For all-day usage, video should be done on a dedicated video engine

Source: Qualcomm Technologies internal data

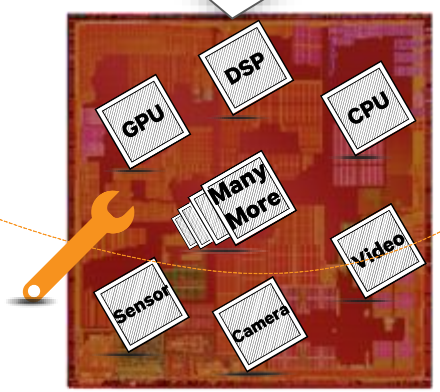
Mobile SoCs are made of many processing engines



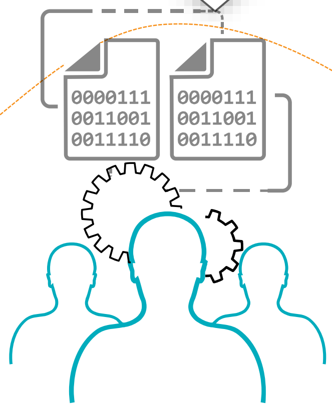
Mobile Heterogeneous Computing Architecture

Mobile heterogeneous computing

A computing approach that intelligently uses fundamentally different types of processing engines



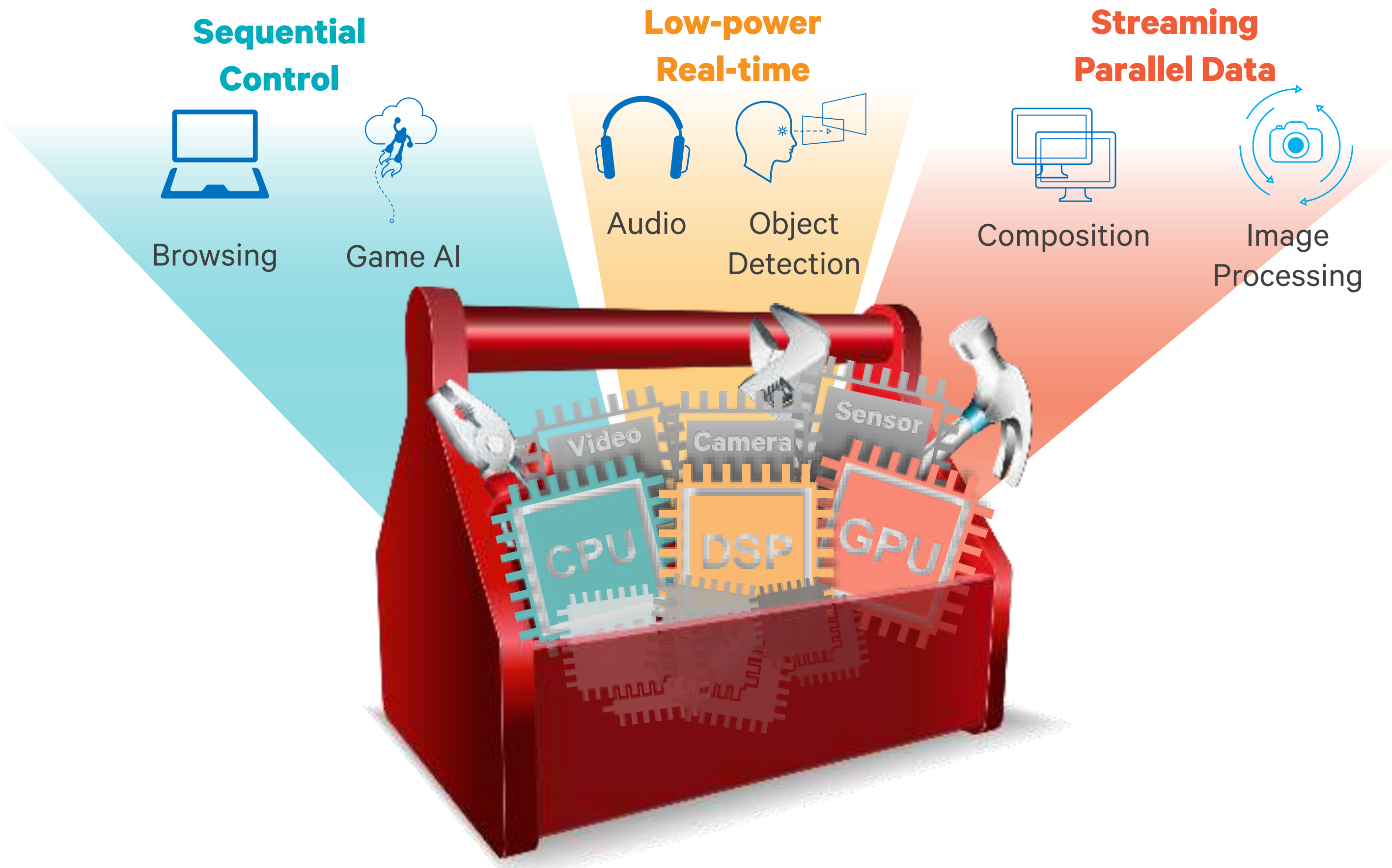
Assign right task for the right processing engine



Accessible & programmable processing engines

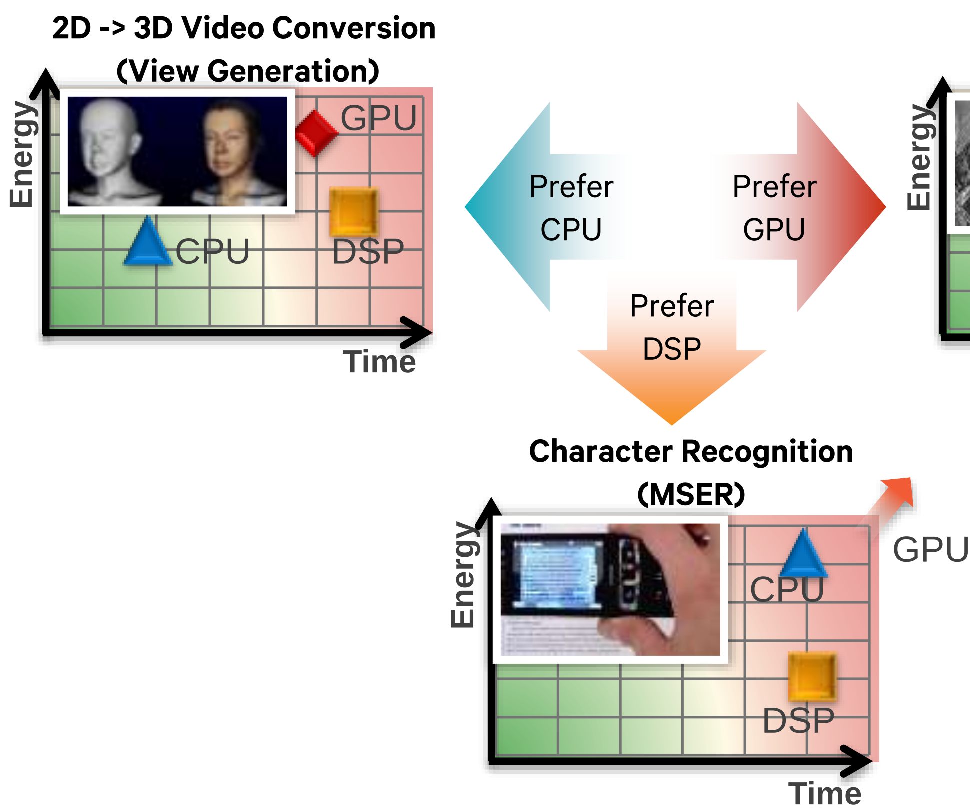
Specialization is key for mobile

Each processing engine has its own strengths



The performance and power benefits of heterogeneity

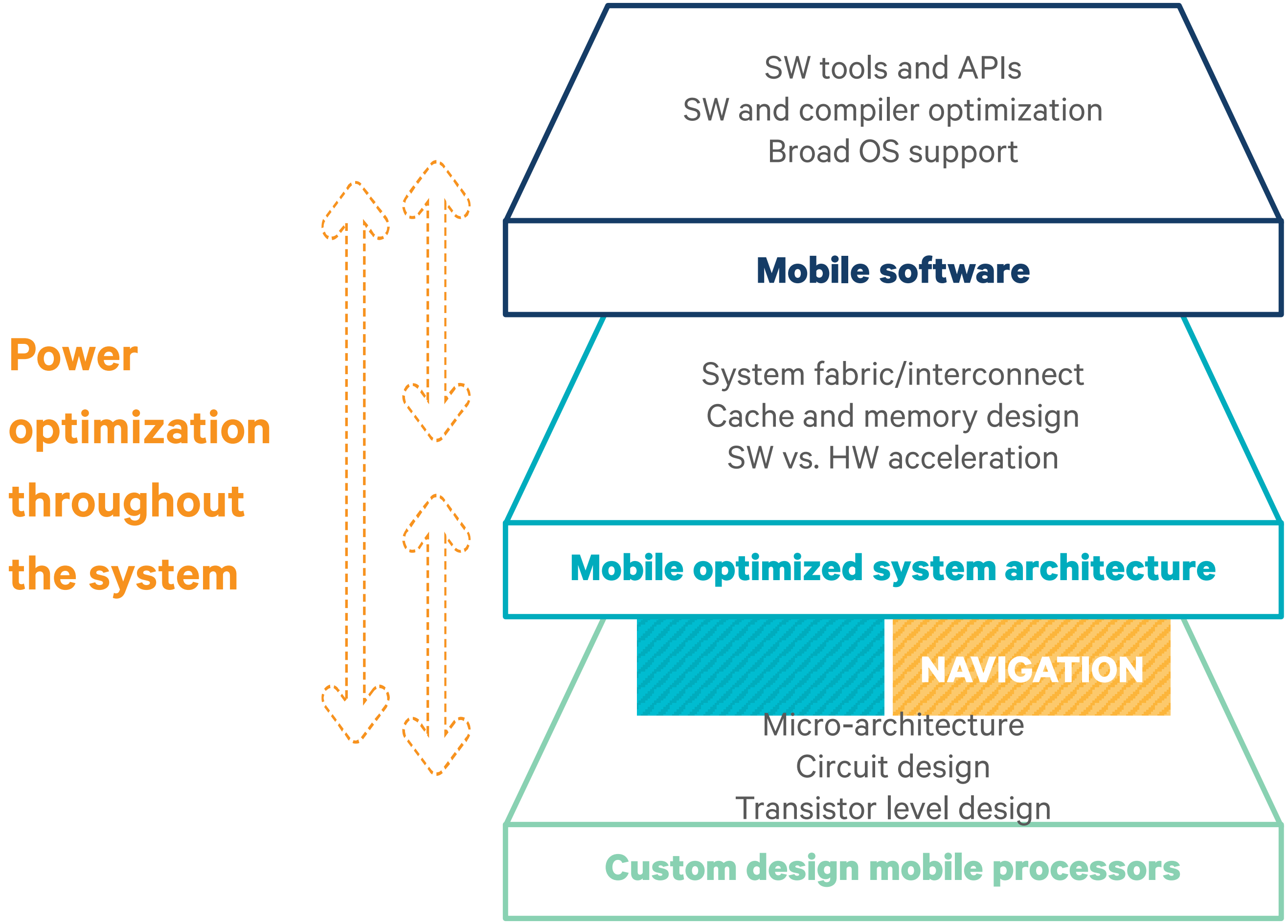
Right task on the right processing engine



Source: Internal Qualcomm technologies measurements on existing Snapdragon™ devices
Snapdragon is a product of Qualcomm Technologies, Inc.

Systems approach is needed for mobile solutions

High performance at low power and thermal



Mobile Cores Coming To Servers

Qualcomm to Build ARM-Based Server Chips

By Jeffrey Burt | Posted 2014-11-19 Email Print

eWEEK

The Data Centers of Tomorrow Will Use the Same Tech Our Phones Do

WIRED

By Peter Levine | Monday August 4, 2014

Share 0 Tweet 620 +1 0 Pin it 8

Today, the mobile phone industry is where so much innovation has been concentrated—resulting in an entirely new class of components created just for this smaller form factor: flash memory, smaller CPUs, networking hardware, and so on. Which means lightweight processors (such as ARM) and low-cost, low-power mobile components are now becoming the foundation of the next-generation datacenter.

MICROPROCESSOR report

Insightful Analysis of Processor Technology

BROADCOM BARES MUSCULAR ARM

Quad-Issue ARMv8 CPU Targets Xeon-Class Performance

By Linley Gwennap (October 21, 2013)

MICROPROCESSOR report

Insightful Analysis of Processor Technology

THUNDERX RATTLES SERVER MARKET

Cavium Develops 48-Core ARM Processor to Challenge Xeon

By Linley Gwennap (June 9, 2014)

CAVIUM

Applied Micro's X-Gene challenges for server processor market

Electronics Weekly.com

Applied Micro leads the charge to infiltrate the $12 billion server processor market with ARM-based ICs.

This is not a trivial task. The $54 billion gorilla standing in Applied Micro's way is Intel with a 90% plus share of the server processor market.

So what, if any, are Applied Micro's selling points compared to Intel's?

First and foremost there's the business model.

"Competition is what we're bringing," says Gaurav Singh, vp of technical strategy at Applied Micro, "in most other markets there is very healthy competition with multiple silicon customers."

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EUROPE WANTS A SMARTPHONE SUPERCOMPUTER

A consortium hopes to build exaflop supercomputers from mobile CPUs

Intel juices up microserver speeds with thrifty Avoton chip

intel Atom

Summary: *Intel is claiming to have made significant strides in performance and power efficiency in the microserver market with its new Avoton system on a chip.*

ZDNet

AMD Announces the Availability of 64-bit ARM Opteron Developer Kits

SUNNYVALE, Calif. 7/30/2014

AMD (NYSE: AMD) today announced the immediate availability of the AMD Opteron™ A1100-Series developer kit, which features AMD's first 64-bit ARM®-based processor, codenamed "Seattle." AMD is the first company to provide a standard ARM Cortex®-A57-based server platform for software developers and integrators. Software and hardware developers as well as early adopters in large datacenters are eligible and can apply on AMD's website.

"The journey toward a more efficient infrastructure for large-scale datacenters is taking a major step forward today with broader availability of our AMD Opteron A1100-Series development kit," said Suresh Gopalakrishnan, general manager and vice president, Server business unit, AMD. "After successfully sampling to major ecosystem partners such as Firmware, OS, and tools providers, we are taking the next step in what will be a collaborative effort across the industry to reimagine the datacenter based on the open business model of ARM innovation."

With this announcement, AMD becomes the only provider of 64-bit ARM server hardware with complete ARMv8 instruction support to foster the development of the ecosystem for efficient storage, Web applications and hosting. AMD is the only provider to offer the standard ARM Cortex-A57 technology.

Contact:

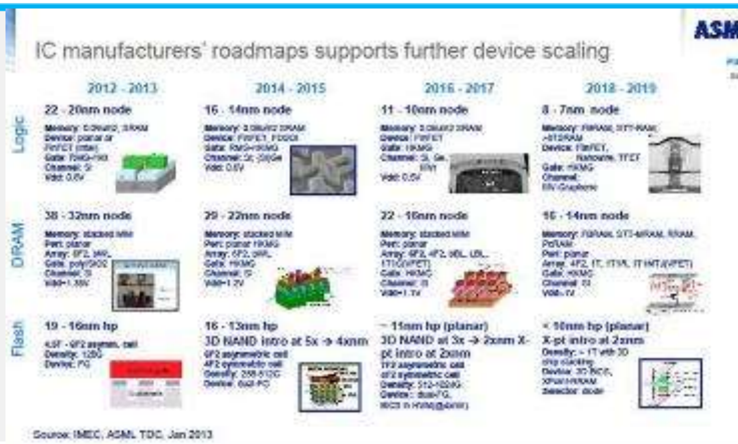
Kristen Lisa
AMD Public Relations
(512) 602-8630

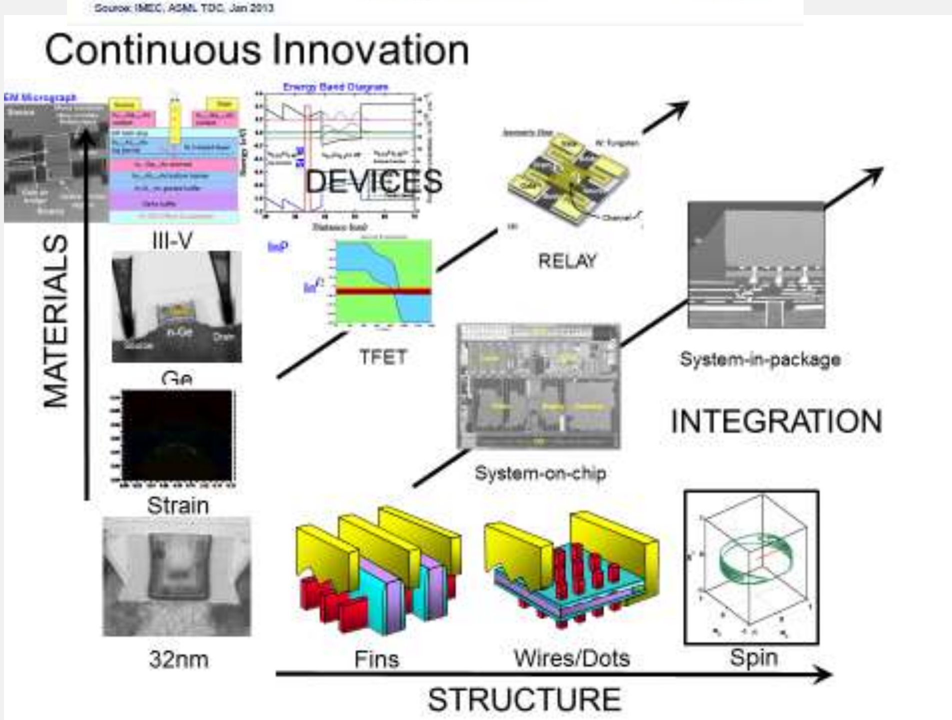
AMD

Where is The Industry Today?

- 14 nm is in production but ramping slower than previous generations
  - Future Generations will be even harder!
- Costs per wafer increasing
  - Capital, more process steps, increased mask costs
  - Cost per transistor decreasing
- PC sales slowing; Server volume growing
- Mobile computing (Smartphones & Tablets) & IoT are driving growth at lower price points
- Moore’s Law will slow down beyond 10 nm
  - Economics, Physics, Materials, Power
  - What is the best use for increased transistor density?
  - Heterogenuous Processing Engines Everywhere?

What is Needed





1999 - Copper Interconnect

200x - SOI Wafers

2003 - Low-k Interlayer Dielectric

2003 - SiGe Strained Silicon Transistors

2007 - High-k/Metal Gate Transistors

2009 - Immersion Lithography

2011 - Tri-Gate Transistors

2015 and beyond: EUV, New Devices, Structures, and Material

What happens beyond 5 nm?

Evolution of the Internet
Yesterday
Today
Tomorrow


Questions?
5 nm
7 nm
10 nm
14 nm
22 nm
32 nm
45 nm
65 nm
dbhandarkar@outlook.com