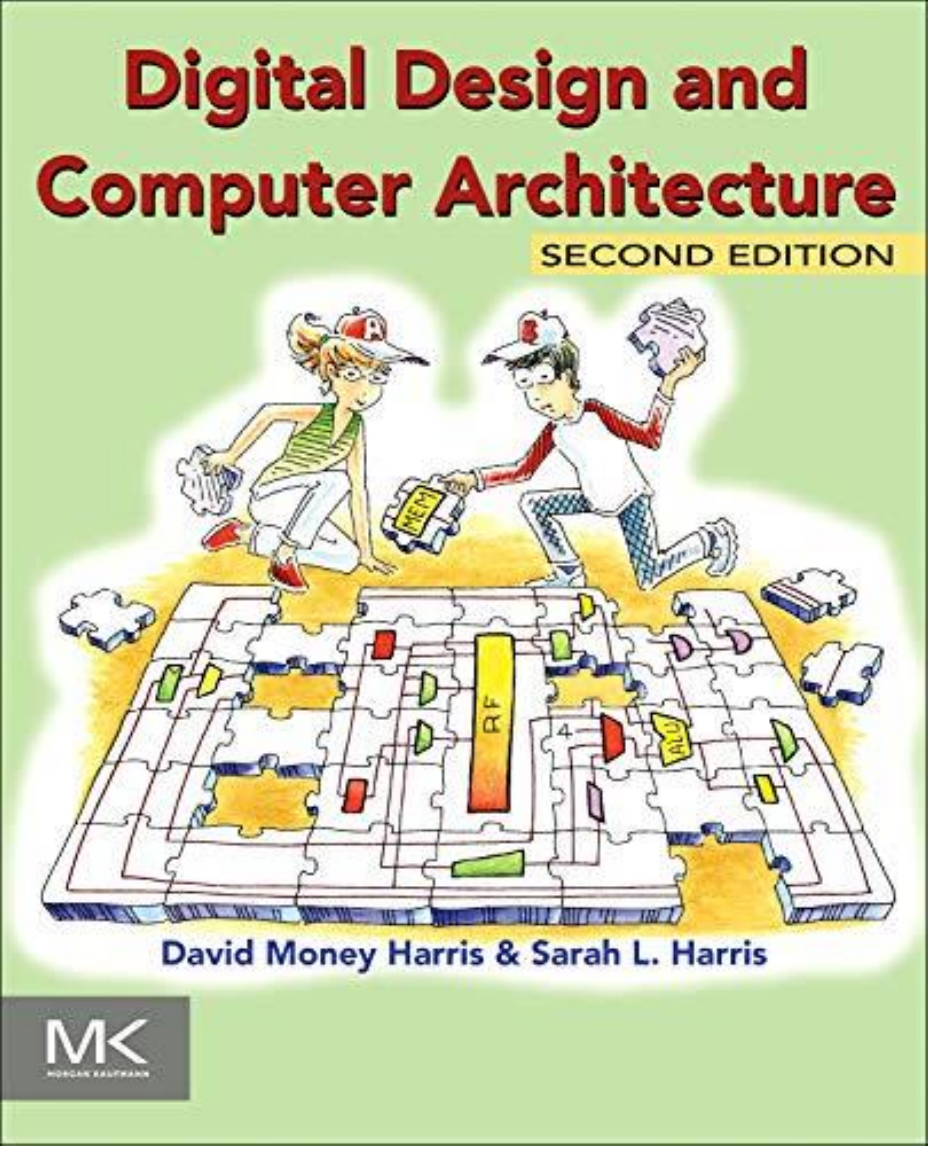
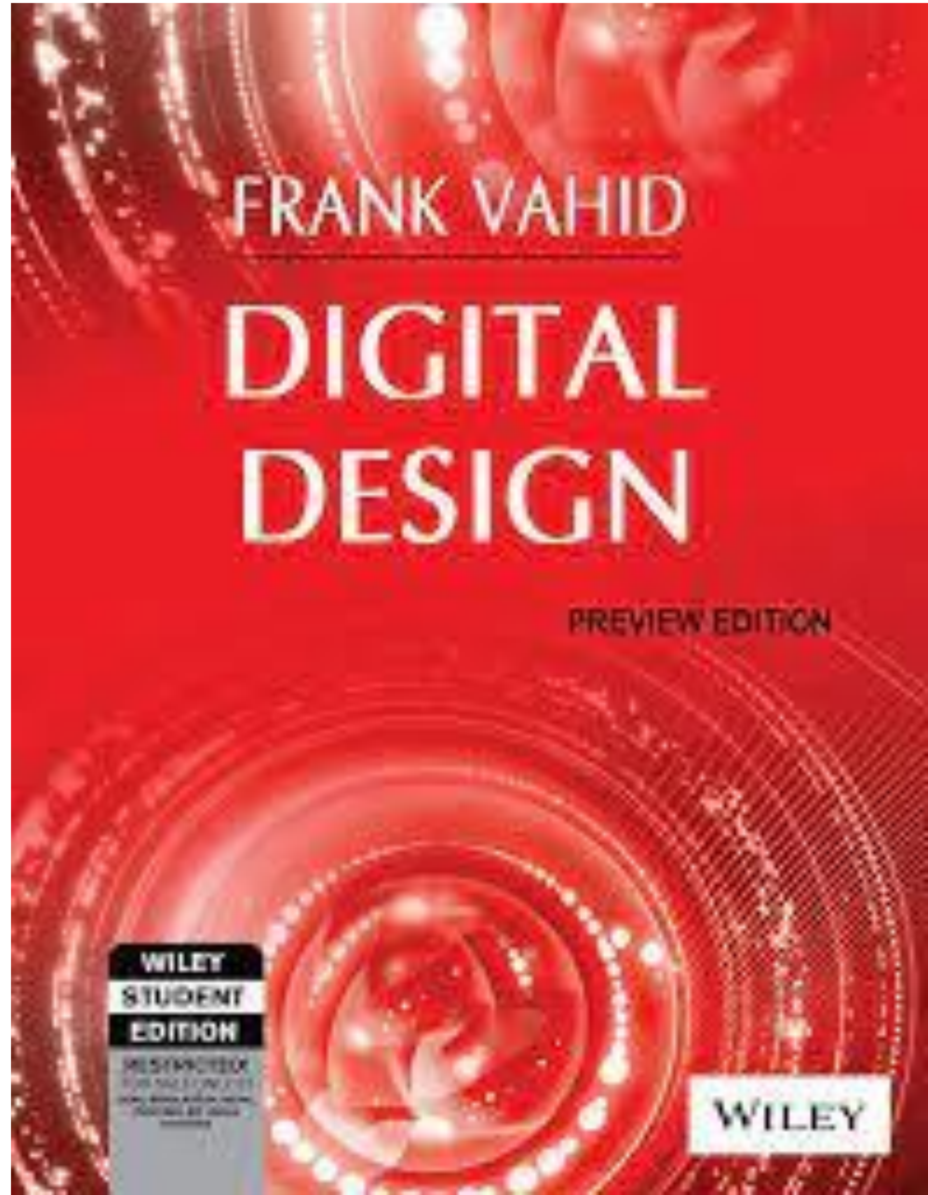
CS227: Digital Systems

Course files:

https://u.pcloud.link/publink/show?code=kZSEOAVZrjPRwQWbBo8xujOedEDXdfgSNaey


Digital Design and
Computer Architecture
SECOND EDITION
MEM
RF
ALU
4
David Money Harris & Sarah L. Harris


FRANK VAHID
DIGITAL
DESIGN
PREVIEW EDITION
WILEY
STUDENT
EDITION
WILEY

Evaluation Policy(CS227)

- Lab : 40%
- Theory : 60%
- Mini Project-important component

What’s Inside the Box?

Why learn this Course

- You want to call yourself a “computer engineer”
- You want to build software people use (need performance)
- You need to make a purchasing decision or offer “expert” advice

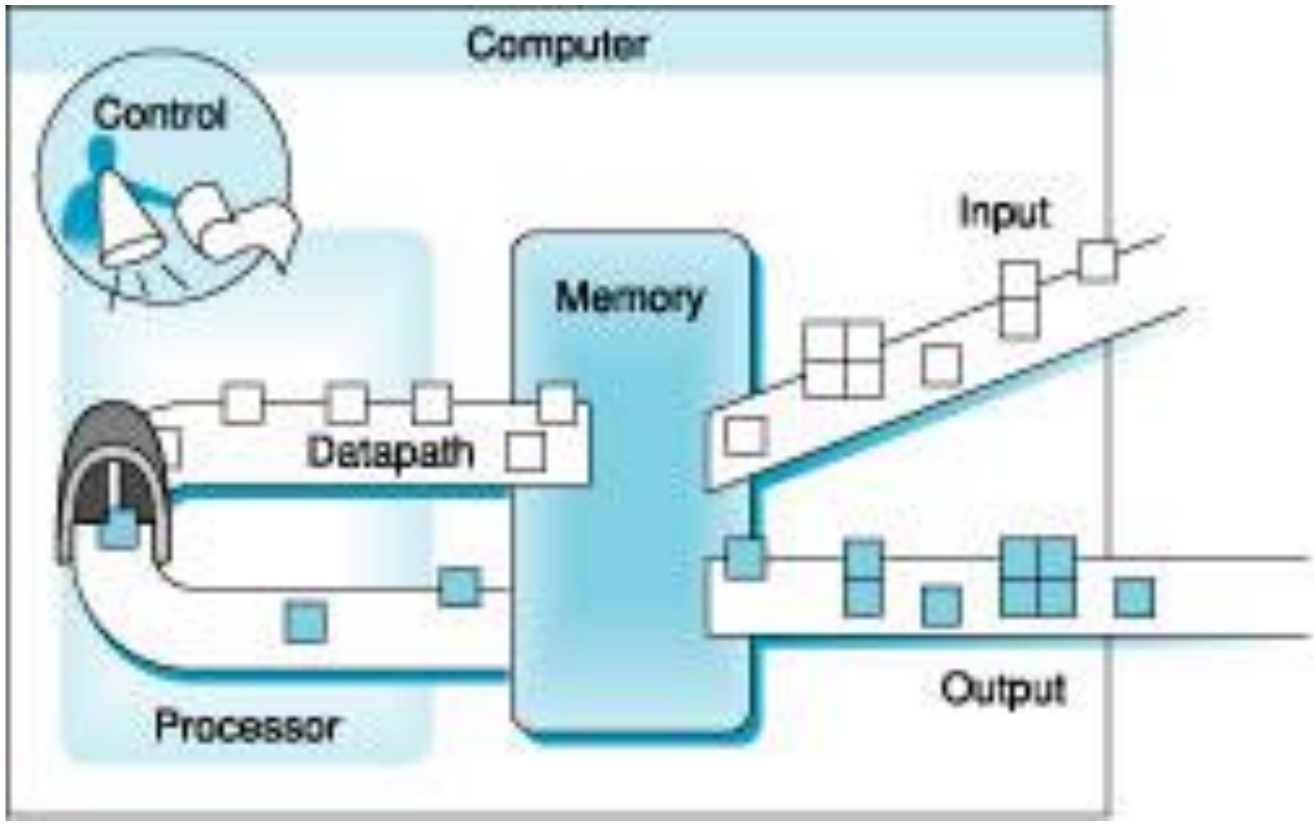
- Both Hardware and Software affect performance:
  - Algorithm determines number of source-level statements
  - Language/Compiler/Architecture determine machine instructions
    Processor/Memory determine how fast instructions are executed

  - I/O and Number_of_Cores determine overall system performance

Organization of a Computer

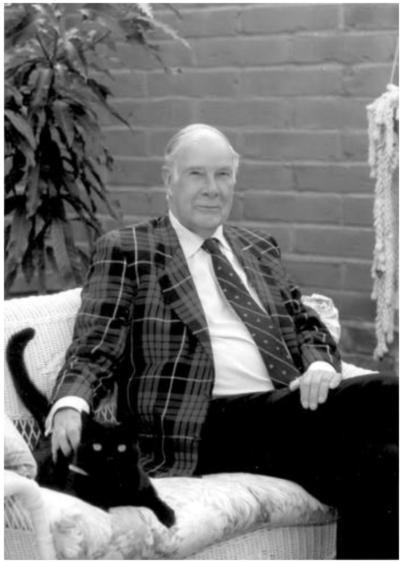
- Five classic components of a computer – input, output, memory, datapath, and control

- datapath + control = processor (CPU)



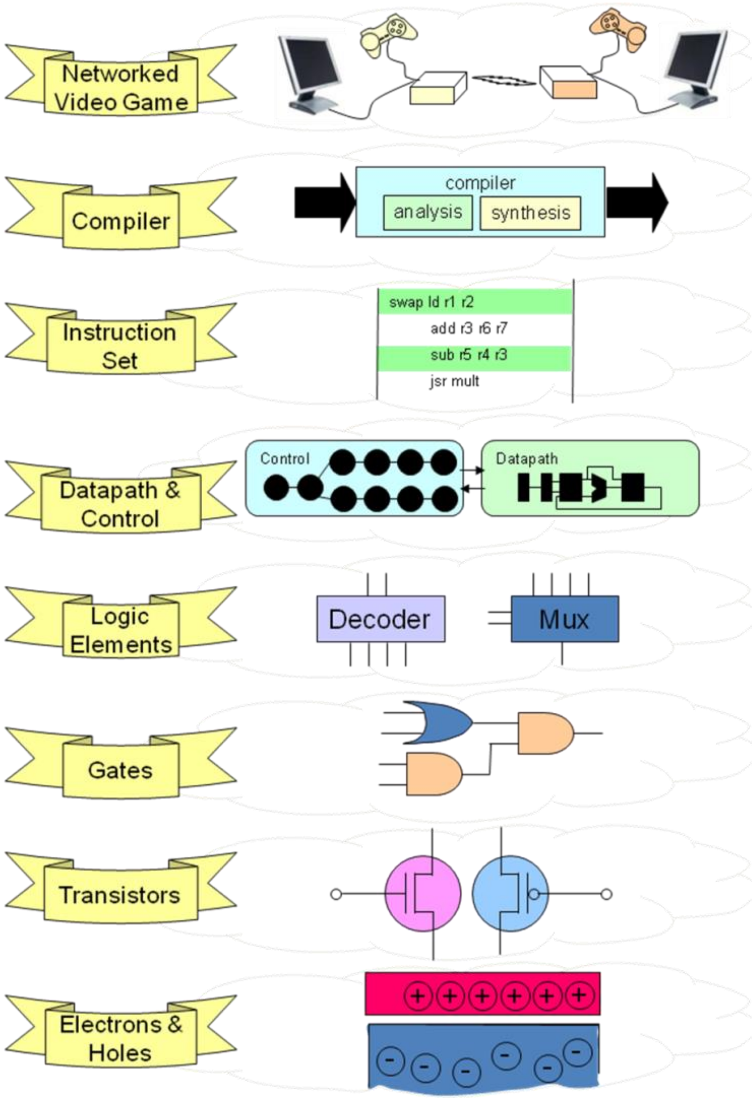
Levels of Transformation

"The purpose of computing is insight" (*Richard Hamming*)
We gain and generate insight by solving problems
How do we ensure problems are solved by electrons?

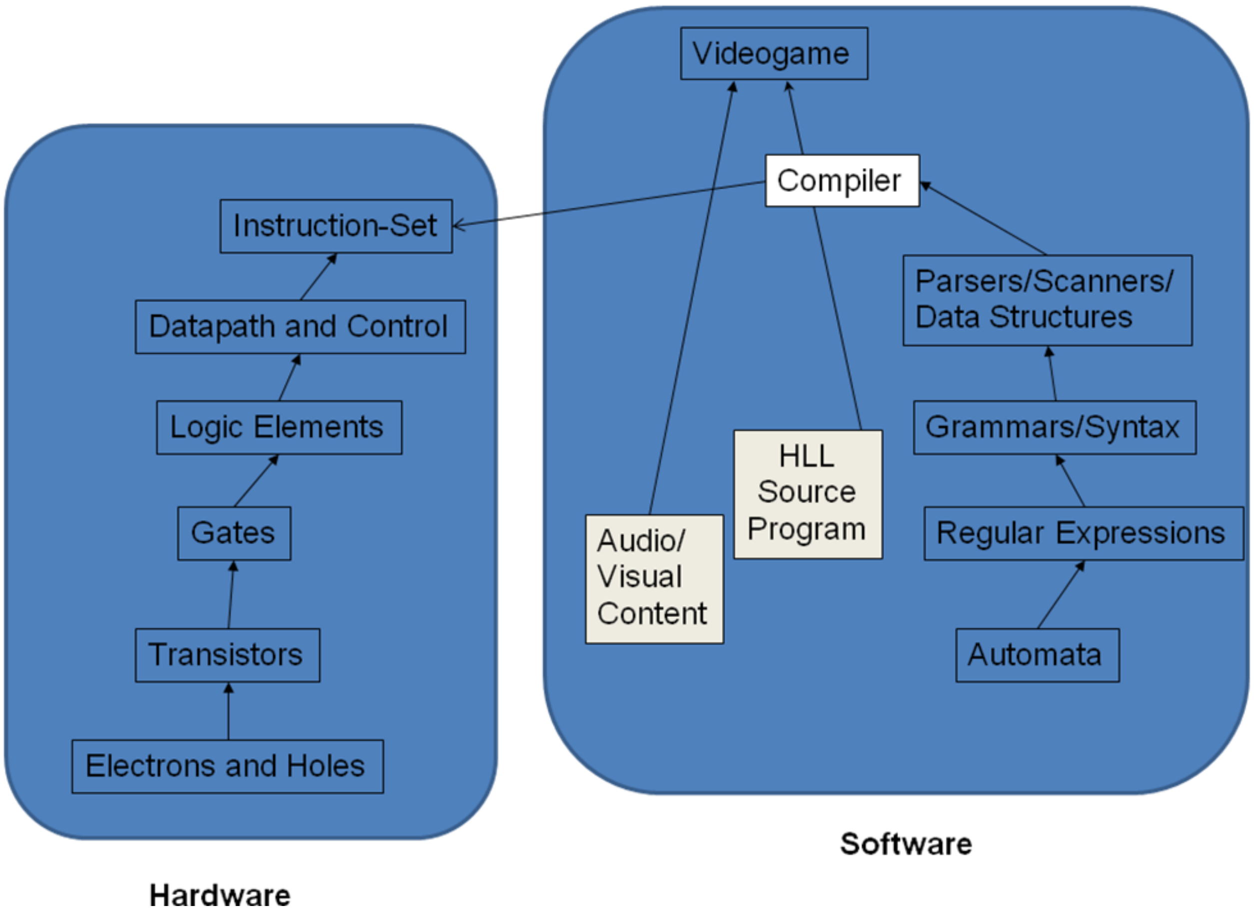


Problem
Algorithm
Program/Language
Runtime System (VM, OS, MM)
ISA (Architecture)
Microarchitecture
Logic
Circuits
Electrons

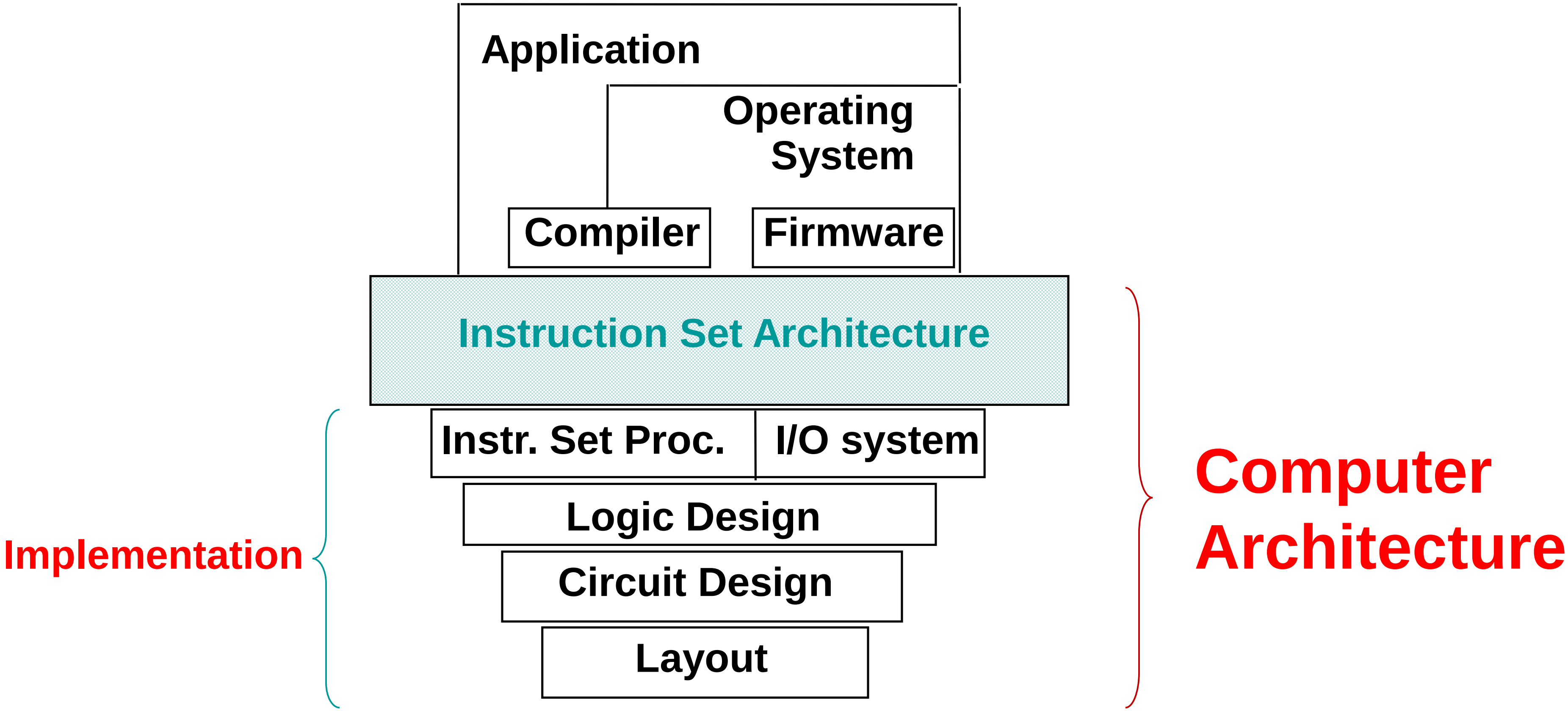
From Electrons & Holes to a Multiplayer Video Game



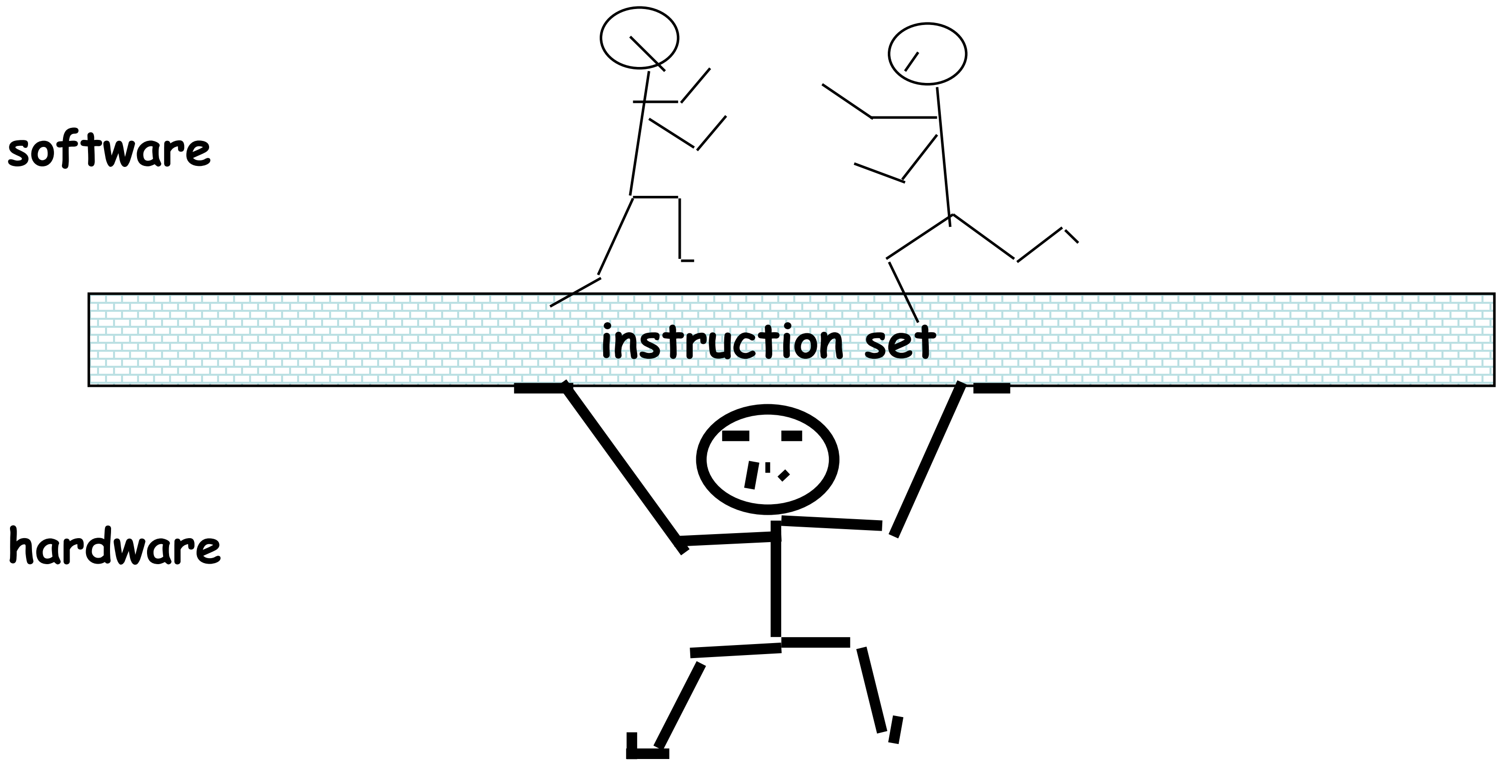
Hardware Software Interface



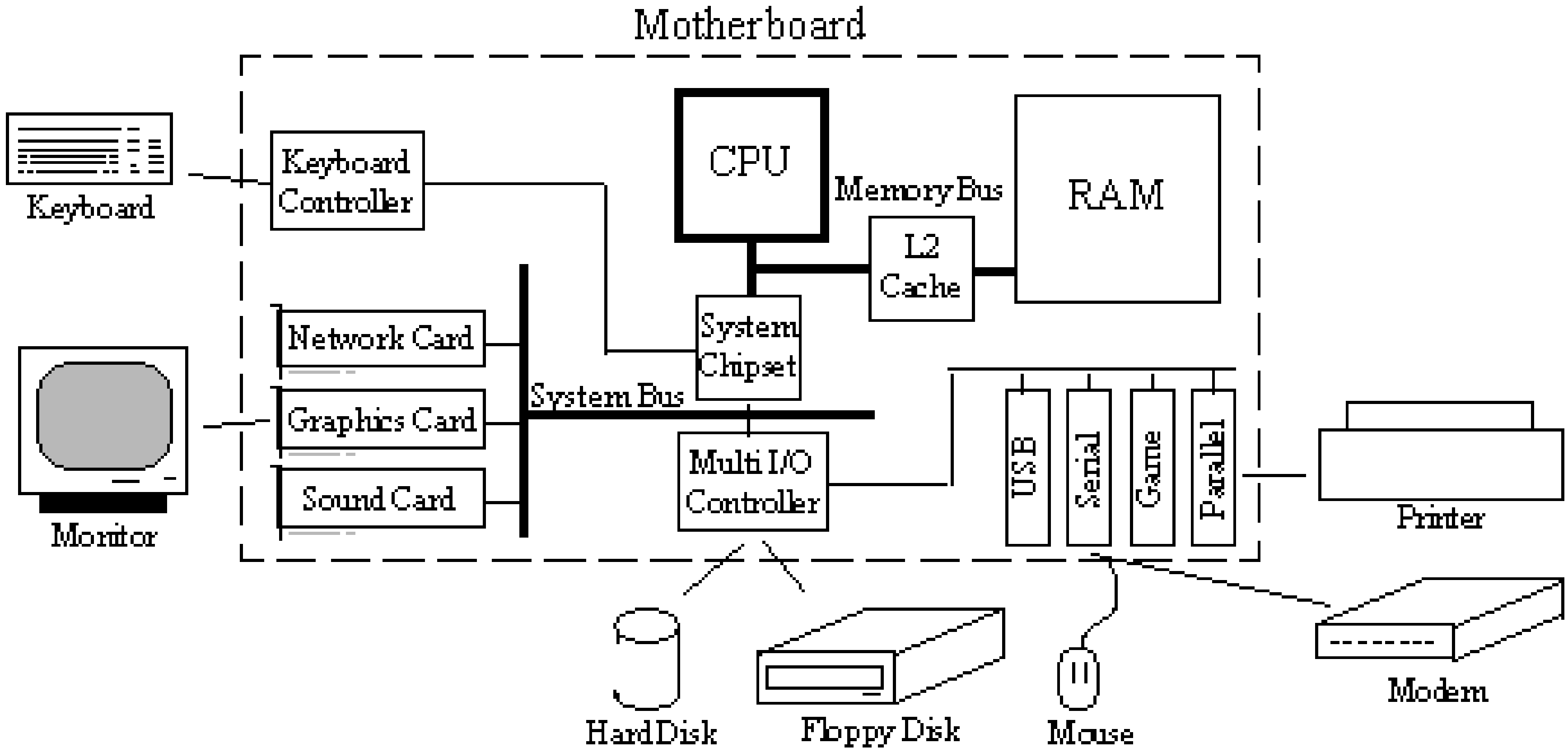
Computer Architecture



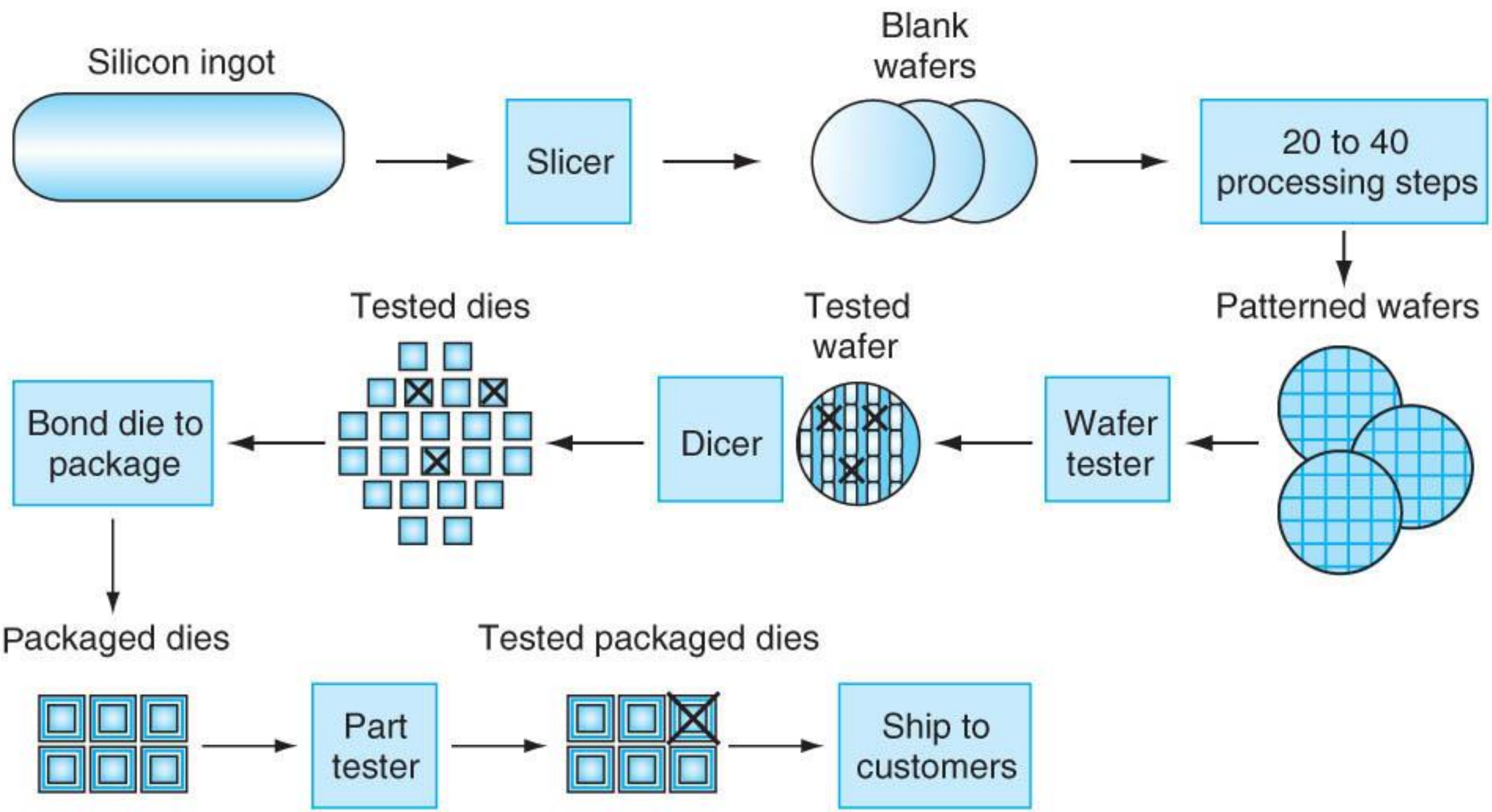
The Instruction Set: a Critical Interface



Function Units in a Computer

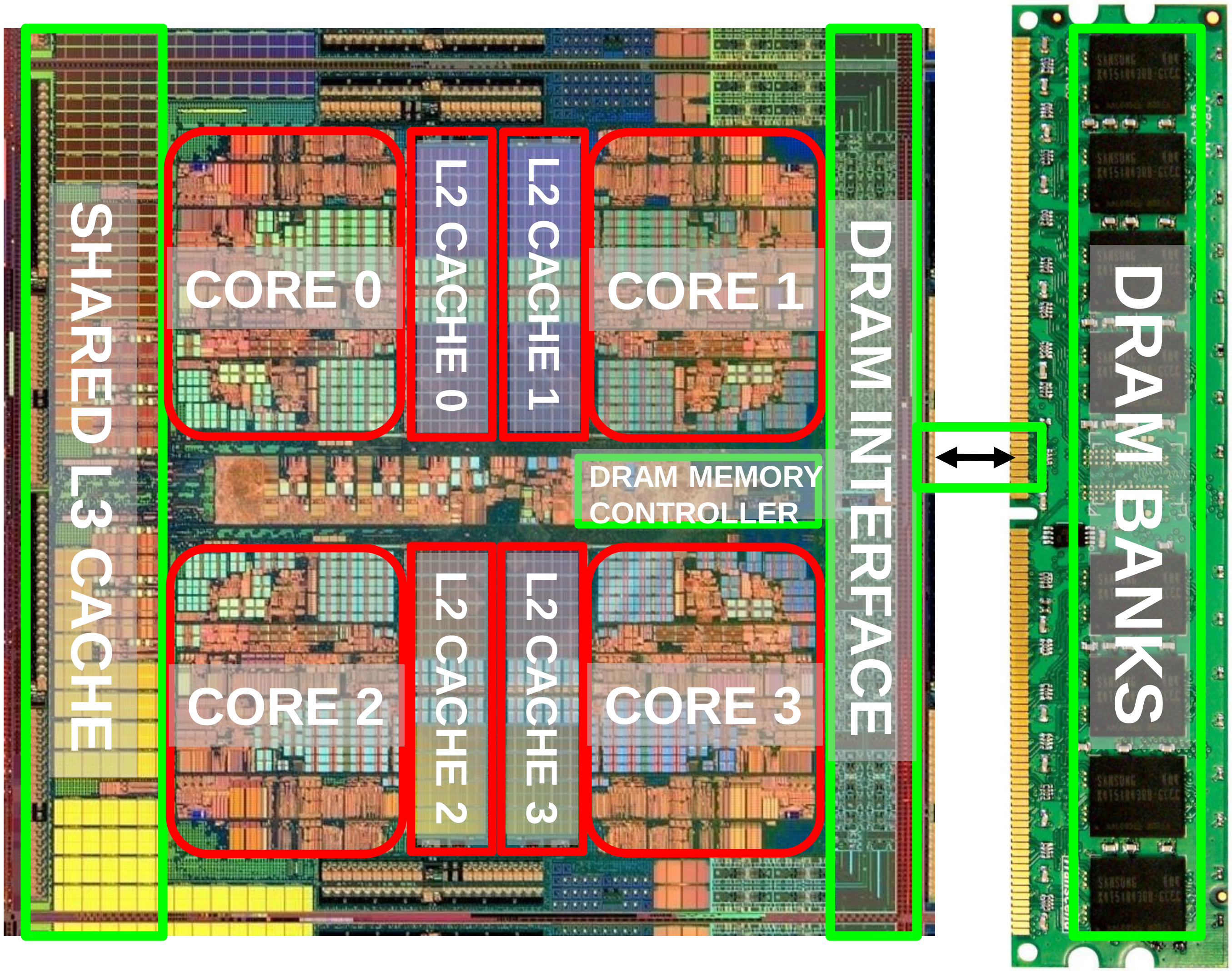


Semiconductor Manufacturing Process for Silicon ICs



An Example: Multi-Core Systems

Multi-Core Chip



*Die photo credit: AMD Barcelona

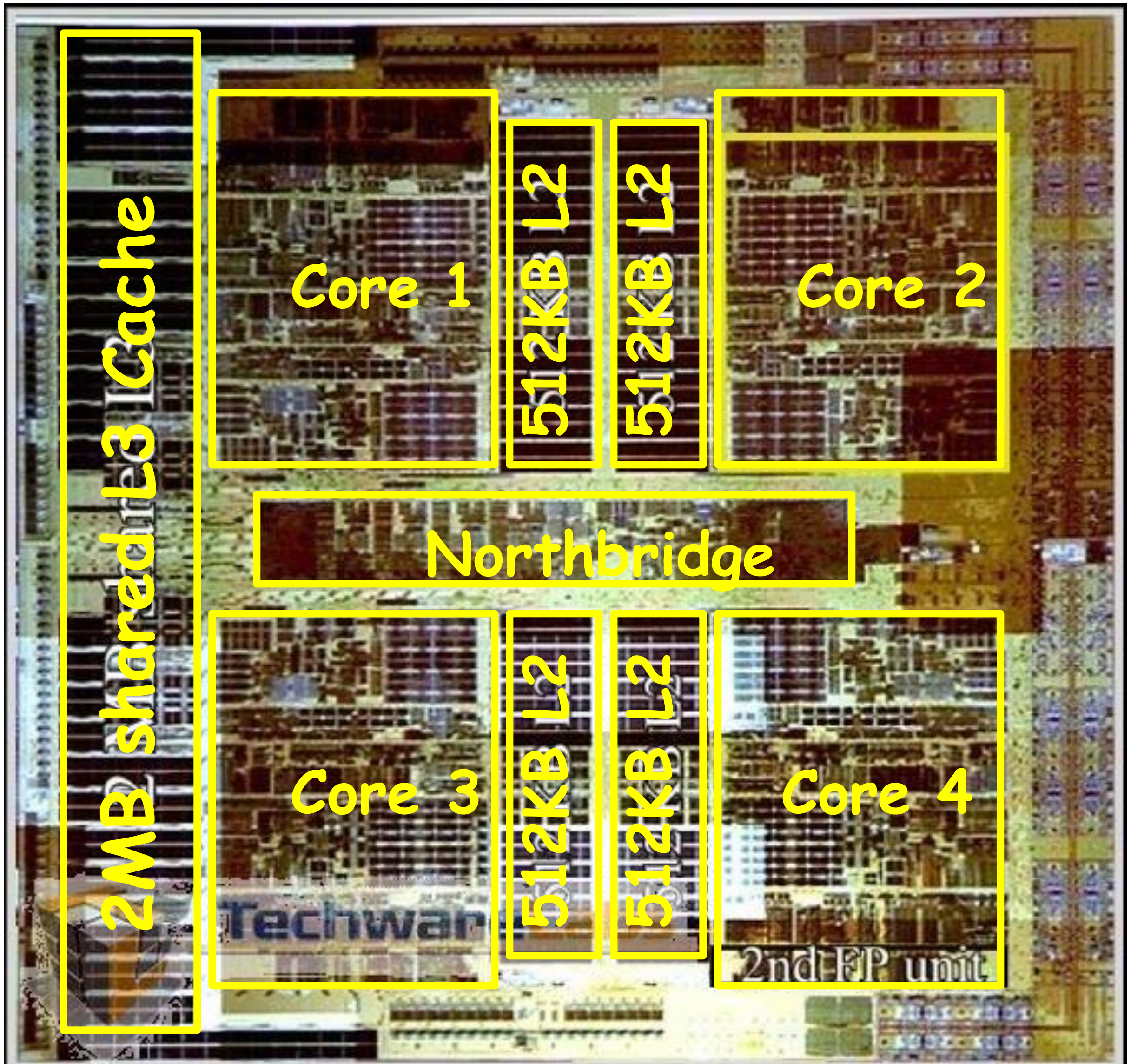
Multicores-Examples

- The power challenge has forced a change in the design of microprocessors
  - Since 2002 the rate of improvement in the response time of programs on desktop computers has slowed from a factor of 1.5 per year to less than a factor of 1.2 per year
- Since 2006, all desktop and server companies are shipping microprocessors with multiple processors – cores – per chip

Product	AMD Barcelona	Intel Nehalem	IBM Power 6	Sun Niagara 2
Cores per chip	4	4	2	8
Clock rate	2.5 GHz	~2.5 GHz?	4.7 GHz	1.4 GHz
Power	120 W	~100 W?	~100 W?	94 W

- The plan is to double the number of cores per chip per generation (about every two years)

AMD's Barcelona Multicore Chip



- Four out-of-order cores on one chip
- 1.9 GHz clock rate
- 65nm technology
- Three levels of caches (L1, L2, L3) on chip
- Integrated Northbridge

http://www.techwarelabs.com/reviews/processors/barcelona/

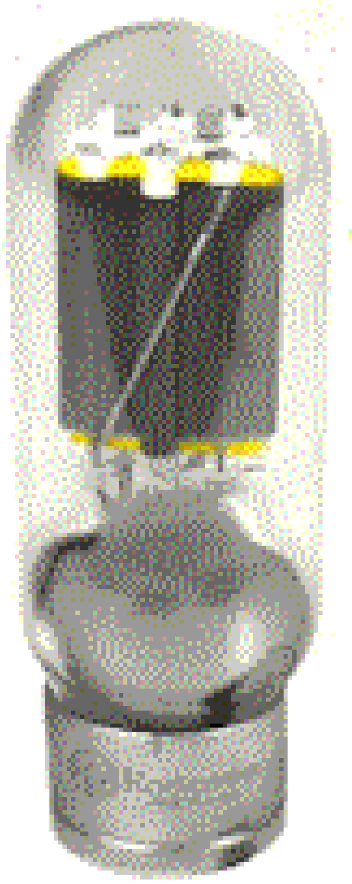
History

Generations of Computer

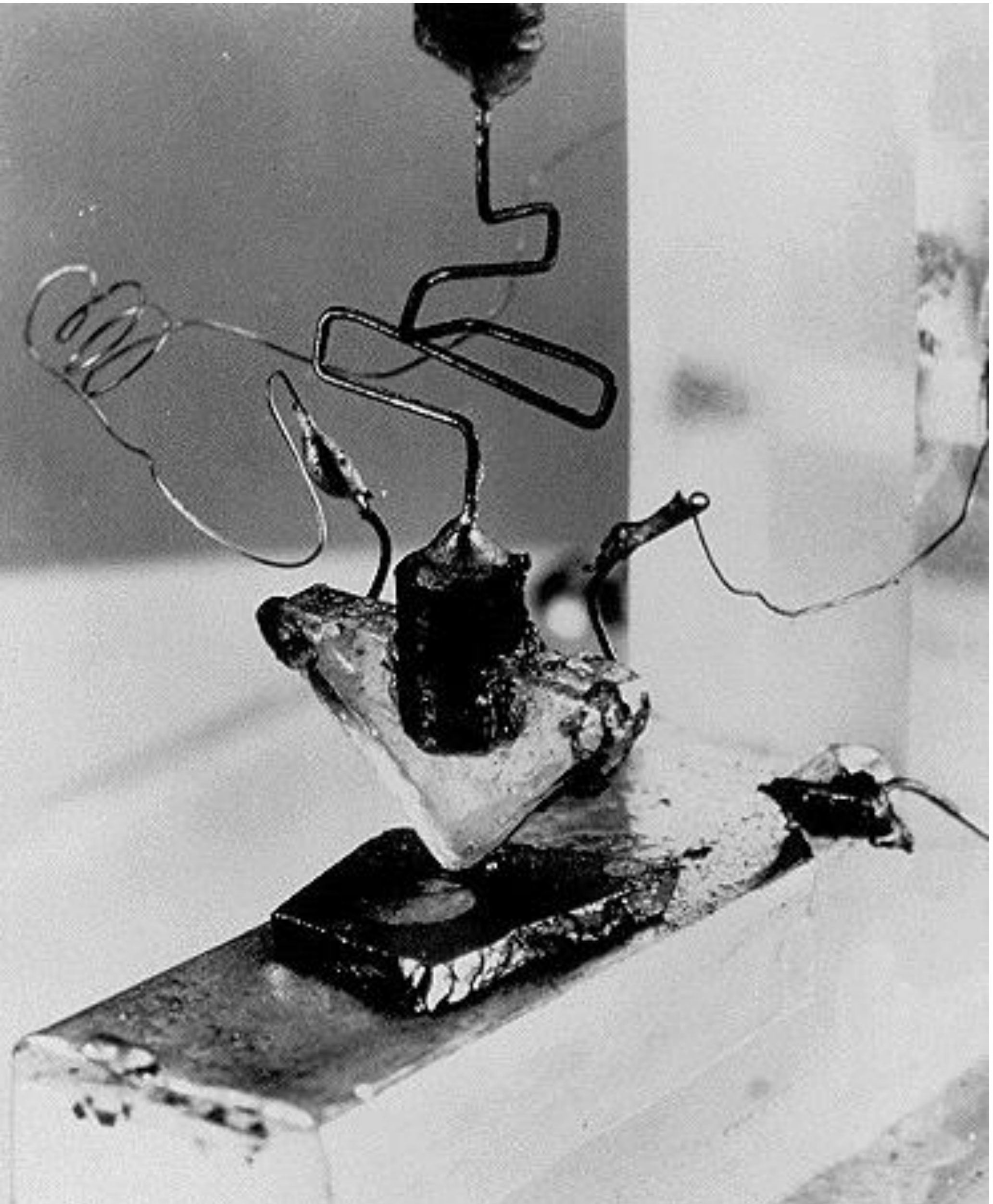
1. First Generation - 1940-1956: Vacuum Tubes
2. Second Generation - 1956-1963: Transistors
3. Third Generation - 1964-1971: Integrated Circuits
4. Fourth Generation - 1971-Present: Microprocessors
5. Fifth Generation - Present and Beyond: Artificial Intelligence

FIRST GENERATION 1940 - 1956

- First generation computers used Vacuum Tubes
- Vacuum tubes are glass tubes with circuits inside.
- The word vacuum indicates that they have no air inside, which protects the circuitry.
- Building a computer with these vacuum tubes would result in a very large machine occupying one full room.



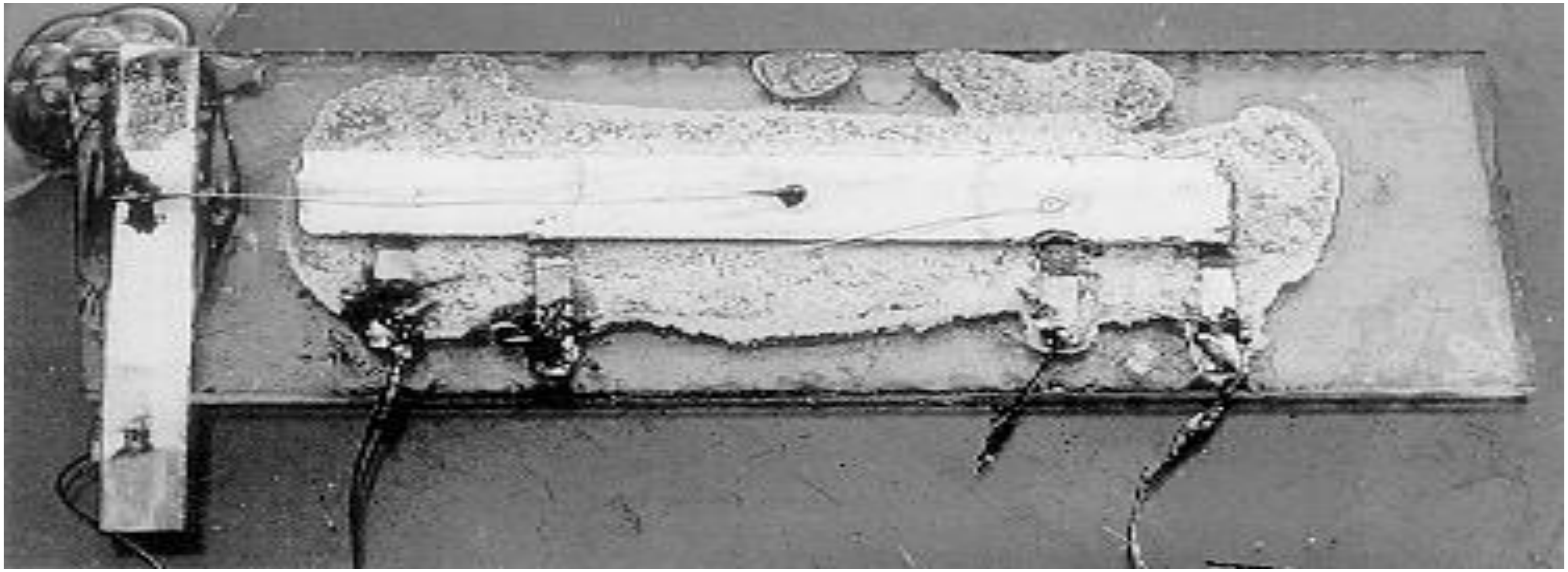
First Transistor



- Uses Silicon
- developed in 1948
- won a Nobel prize
- on-off switch

- Second Generation Computers used Transistors, starting in 1956

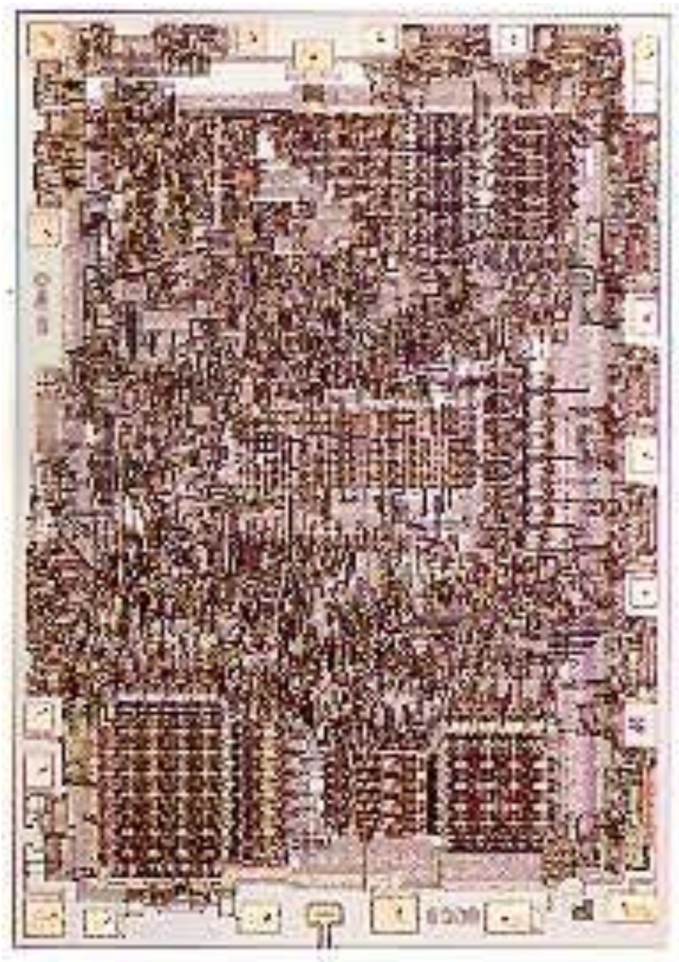
Integrated Circuits



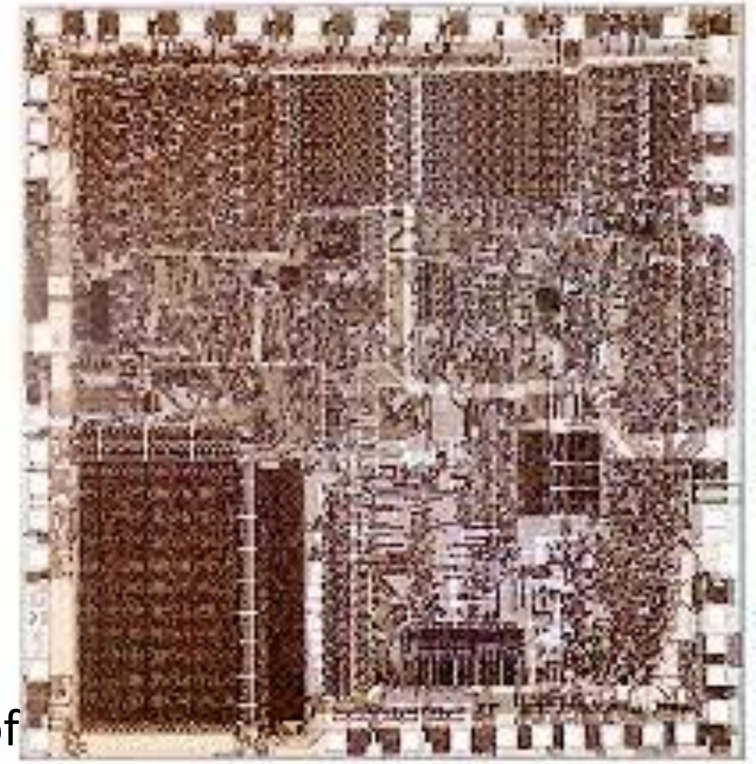
- Third Generation Computers used Integrated Circuits (chips).
- Integrated Circuits are transistors, resistors, and capacitors integrated together into a single “chip”
- Instead of punched cards and printouts, users started interacting with keyboards and mouse.

1972: 8008 Microprocessor

- The 8008 was twice as powerful as the 4004.
- According to the magazine *Radio Electronics*, Don Lancaster, a dedicated computer hobbyist, used the 8008 to create a predecessor to the first personal computer, a device *Radio Electronics* dubbed a "TV typewriter." It was used as a dumb terminal.

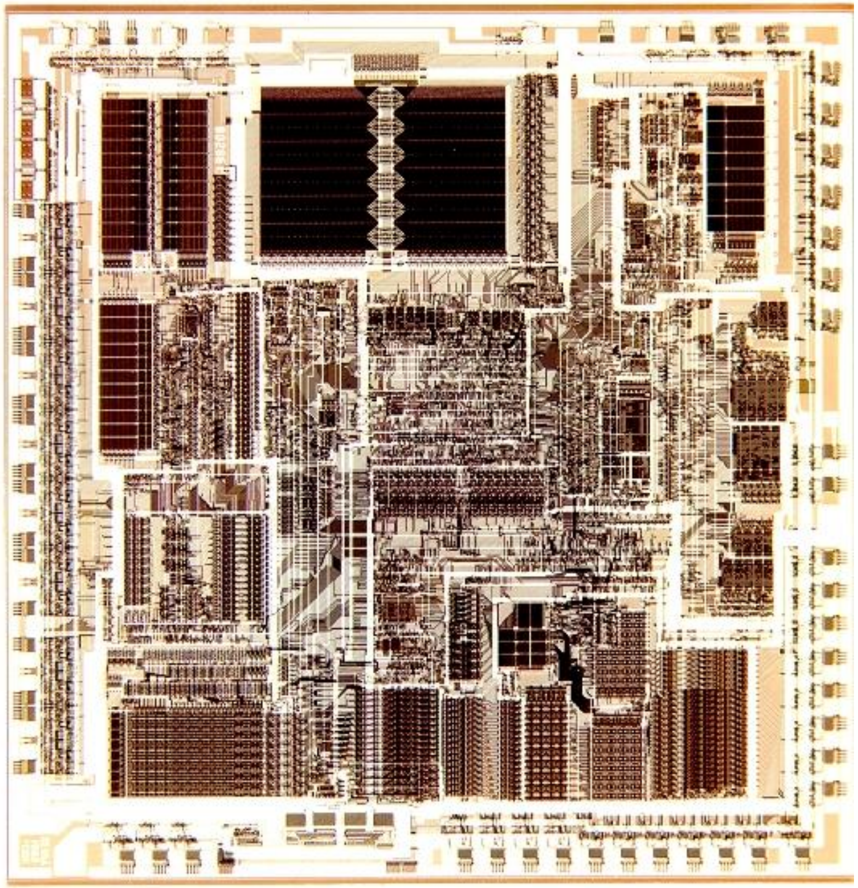


1978: 8086-8088 Microprocessor



- In 1978 intel
  - 16-bit
  - 2.5MIPS **millions of instructions per second**
  - **1M-byte memory**
  - 4- or 6-byte instruction (cache) queue that prefetch instructions
  - over 20,000 variations instructions.
  - 64K bytes of memory found in 8-bit microprocessors to execute efficiently
  - The 16-bit 8086 and 8088 provided 1M byte of memory for these applications
  - Popularity of Intel ensured in 1981 when IBM chose the 8088 in its personal computer

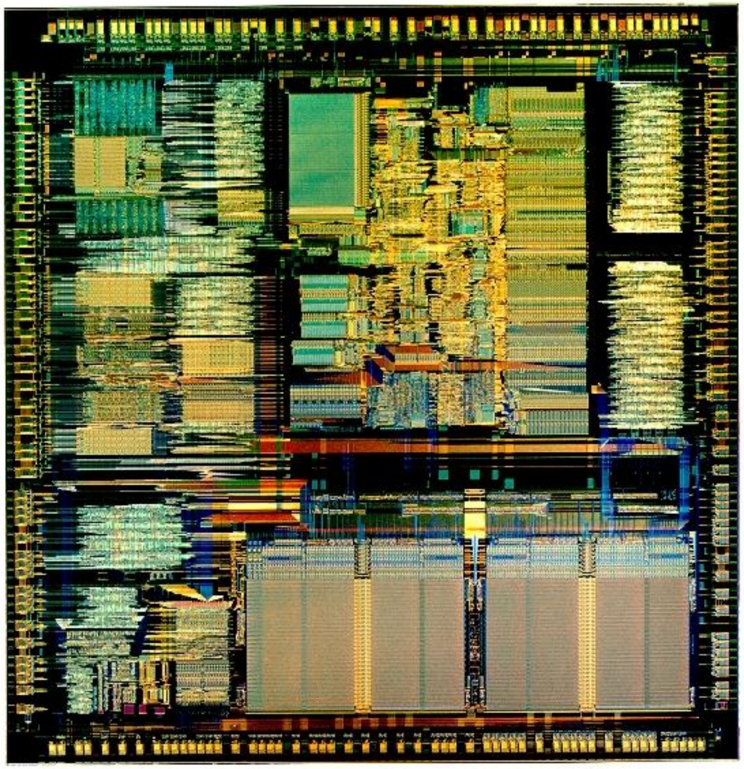
1982: 286 Microprocessor



- The 286, also known as the 80286, was the first Intel processor that could run all the software written for its predecessor.
- This software compatibility remains a hallmark of Intel's family of microprocessors.
- Within 6 years of it release, there were an estimated 15 million 286-based personal computers installed around the world.

- 80286: updated 8086
  - 16M byte memory addressing
  - Instructions identical to 8086 few more added
  - 4MIPS
  - 8Mhz clock speed

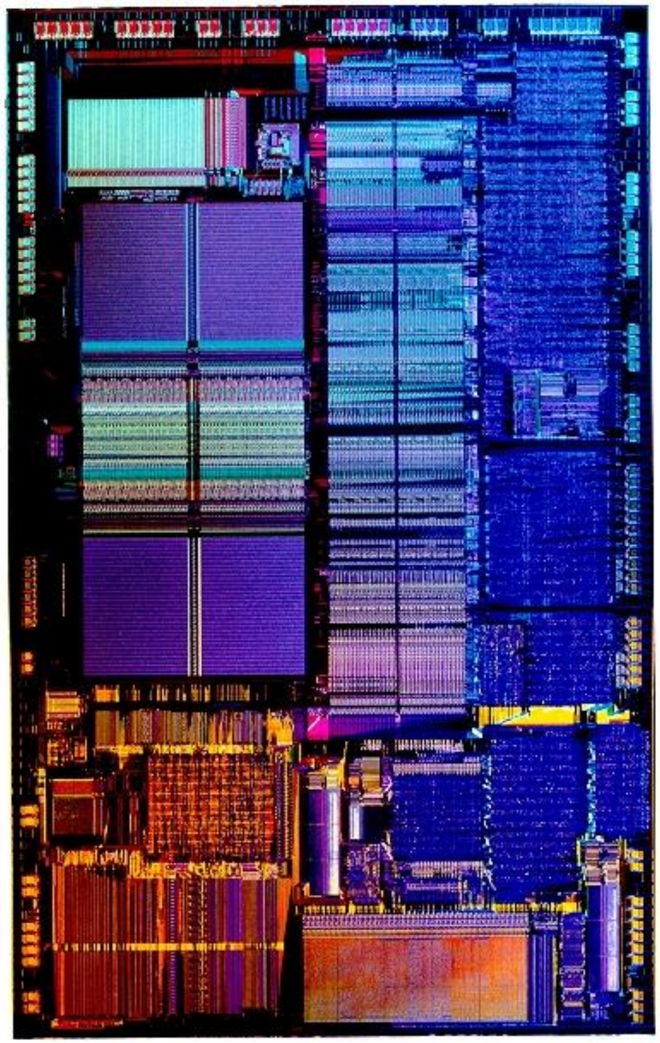
1985: Intel 386(TM) Microprocessor



- The Intel 386[TM] microprocessor featured 275,000 transistors--more than 100times as many as the original 4004.
- It was a 32-bit chip and was "multi tasking," meaning it could run multiple programs at the same time.
- 80386 (1985)
  - 32-bit address bus and 32-bit data bus
  - 4GB memory
  - Hardware circuitry for memory management
  - Additional instructions referenced 32-bit registers and managed the memory system
-

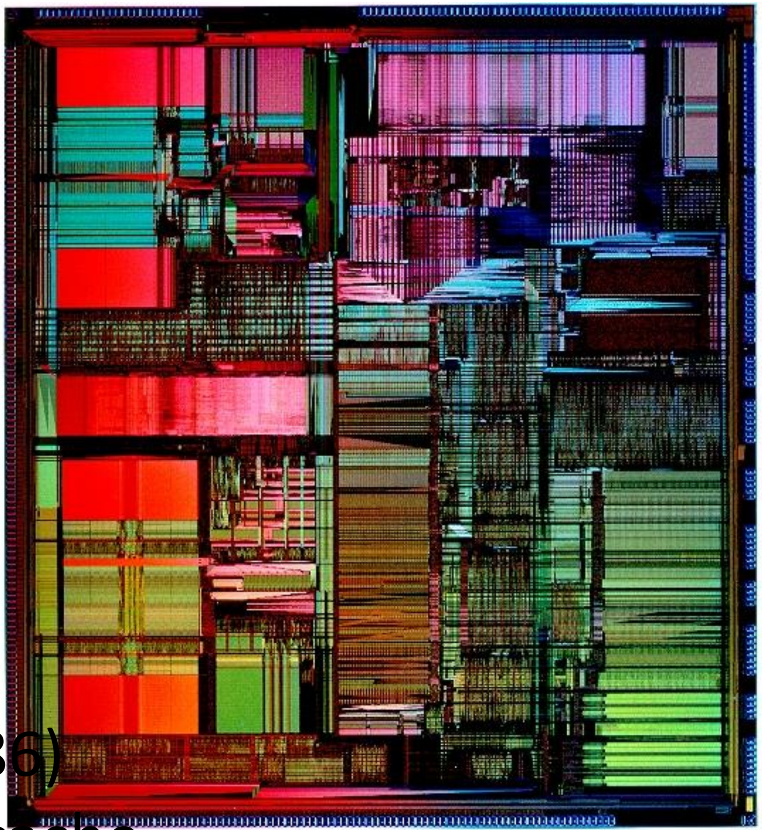
1989: Intel 486(TM) DX CPU

- The 486™ generation really allowed the shift from command-level computing into point-and-click computing.
- The Intel 486™ processor was the first to offer a built-in math coprocessor, which speeds up computing because it offers complex math functions from the central processor.



- 80486
  - Highly integrated package.
  - 80386-like microprocessor.80387-like numeric coprocessor.
  - 50 MIPS
  - 8K-byte cache memory system
  - Half of its instructions executed in 1 clock cycle rather 2

1993: Pentium® Processor

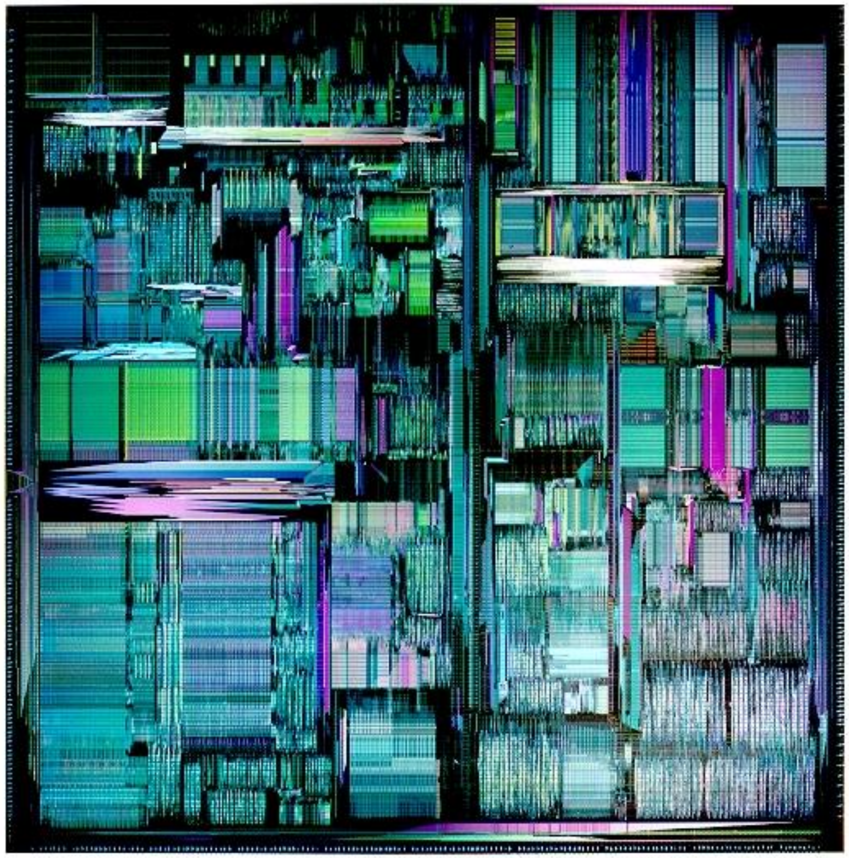


- The Pentium® processor allowed computers to more easily incorporate "real world" data such as speech, sound, handwriting and photographic images.
- The name Pentium®, mentioned in the comics and on television talk shows, became a household word soon after introduction.

  - Originally named P5 or 80586
  - Clock speed of 60MHZ
  - Executes 110MIPS
- Cache size: 16K bytes (8K cache in 80486)
  - 8K-byte instruction cache and data cache.
- Memory system up to 4G bytes.
- Data bus width increased to a full 64 bits.
- Data bus transfer speed 60 MHz or 66 MHz.
  - depending on the version of the Pentium

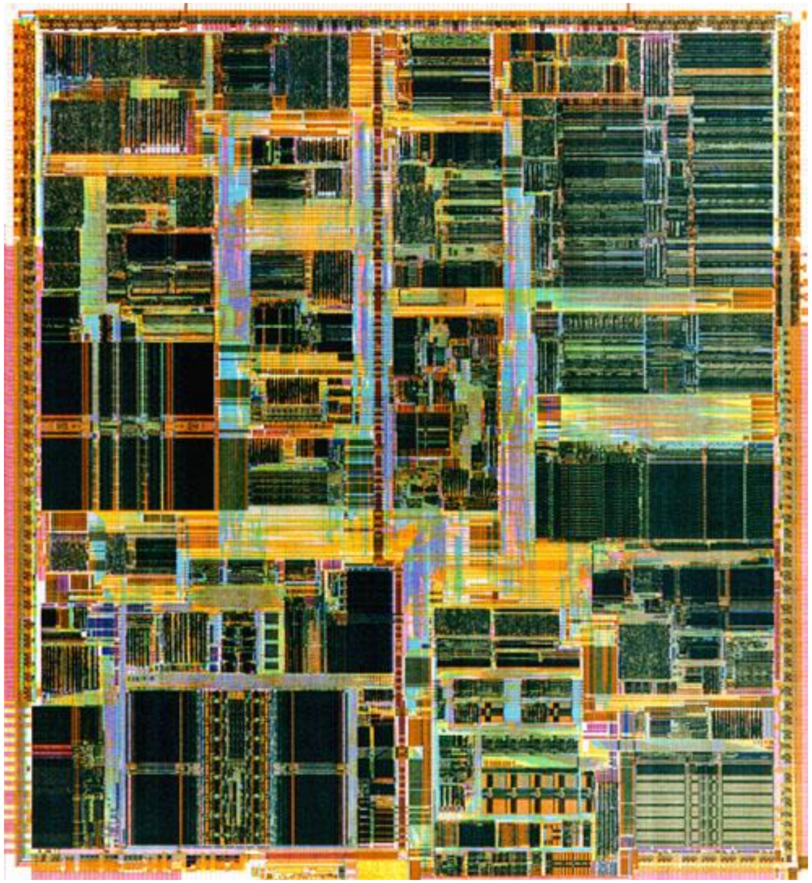
1995: Pentium® Pro Processor

- Released in the fall of 1995 the Pentium® Pro processor is designed to fuel 32-bit server and workstation-level applications, enabling fast computer-aided design, mechanical engineering and scientific computation.
- Each Pentium® Pro processor is packaged together with a second speed-enhancing cache memory chip.
- The powerful Pentium® Pro processor boasts 5.5 million transistors.



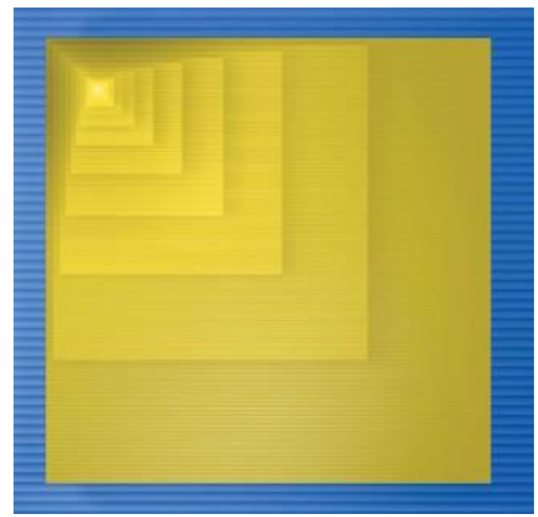
1997: Pentium® II Processor

- The 7.5 million-transistor Pentium® II processor incorporates Intel MMX™ technology, which is designed specifically to process video, audio and graphics data efficiently.
- It is packaged along with a high-speed cache memory chip in an innovative Single Edge Contact (S.E.C.) cartridge that connects to a motherboard via a single edge connector, as opposed to multiple pins.
- With this chip, PC users can capture, edit and share digital photos with friends and family via the Internet; edit and add text, music or between-scene transitions to home movies; and, with a video phone, send video over standard phone lines and the Internet.

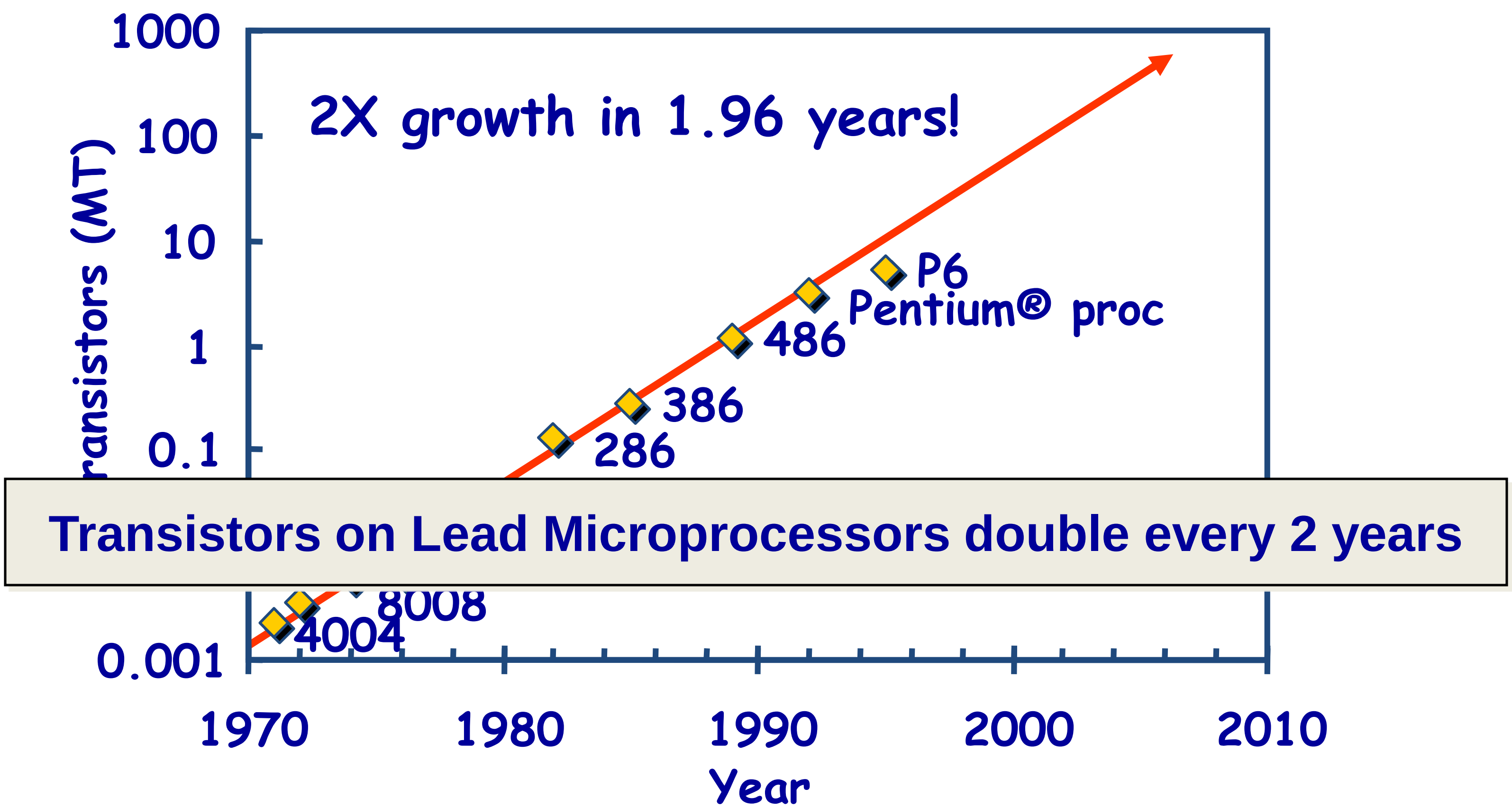


Moore’s Law

- **IC capacity doubling about every 18 months for several decades**
  - **Known as “Moore’s Law” after Gordon Moore, co-founder of Intel**
    - **Predicted in 1965 predicted that components per IC would double roughly every year or so**
  - **Picture depicts related phenomena**
    - **For a particular number of transistors, the IC shrinks by half every 18 months**
      - **Notice how much shrinking occurs in just about 10 years**
      - **Enables incredibly powerful computation in incredibly tiny devices**
  - **Today’s ICs hold *billions* of transistors**
    - **The first Pentium processor (early 1990s) needed only 3 million**

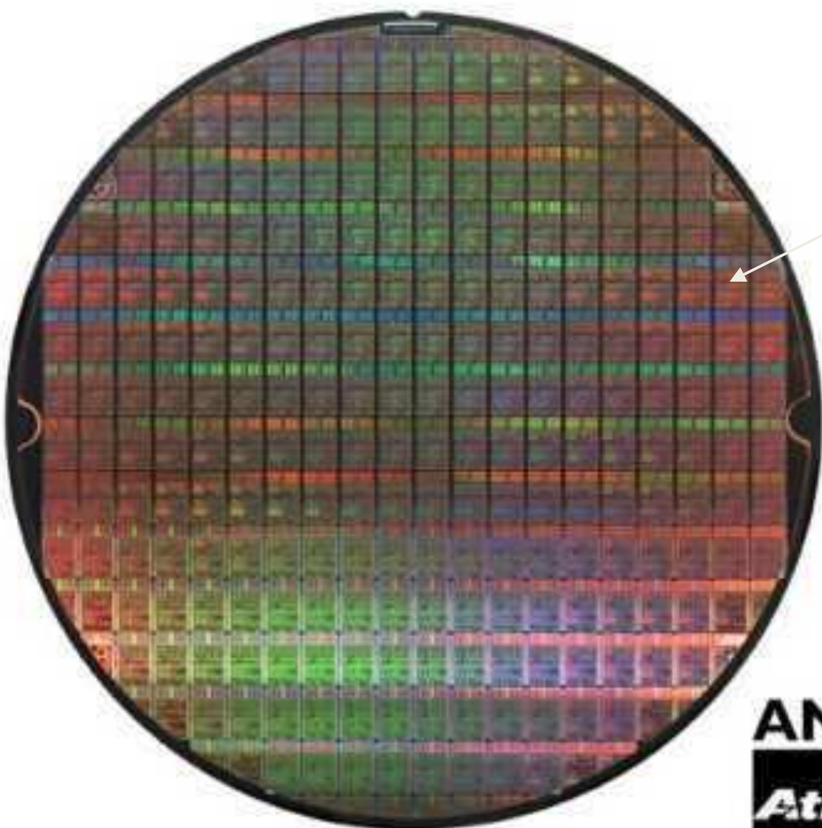


Moore’s law in Microprocessors



Courtesy, Intel

Die



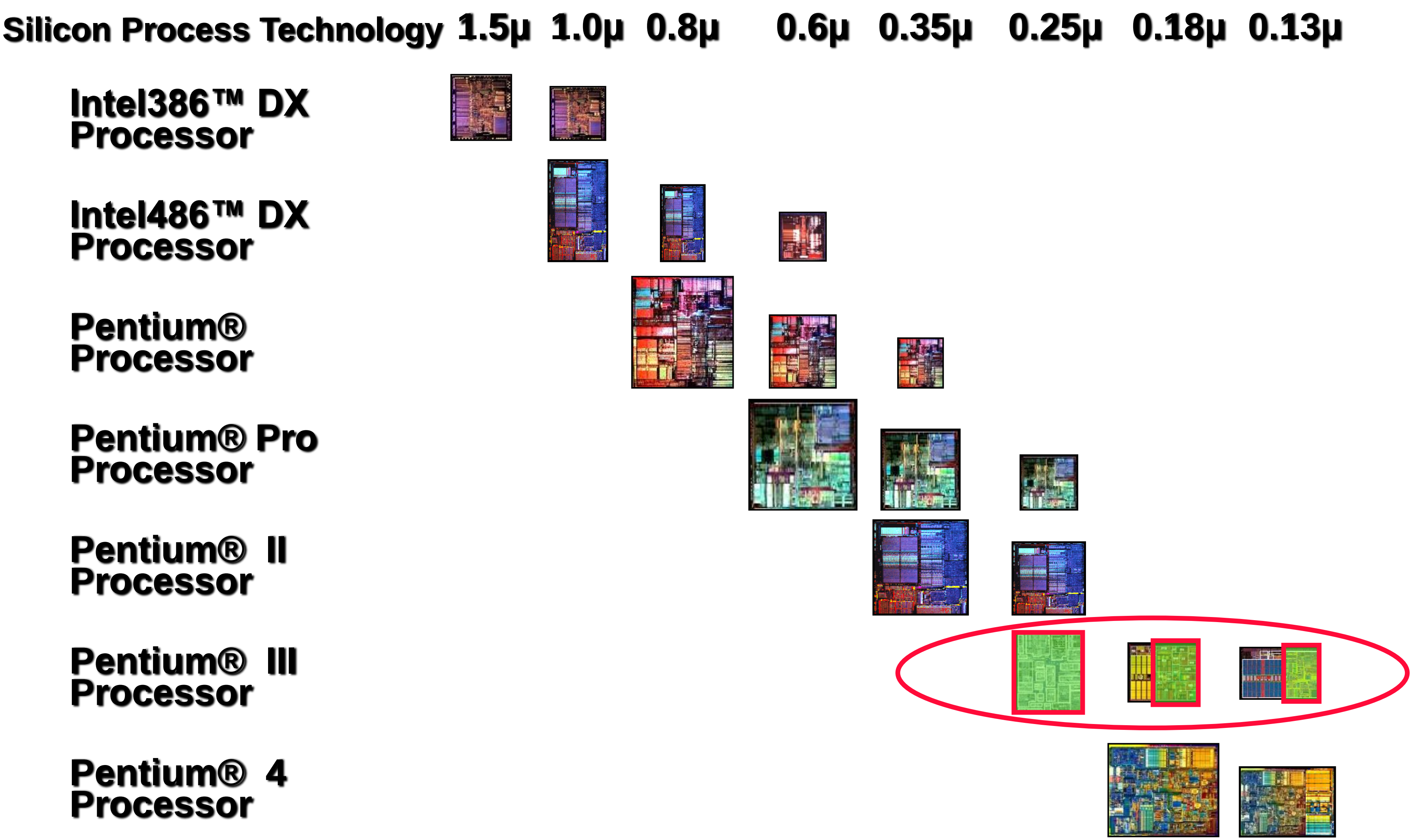
Single die

Wafer

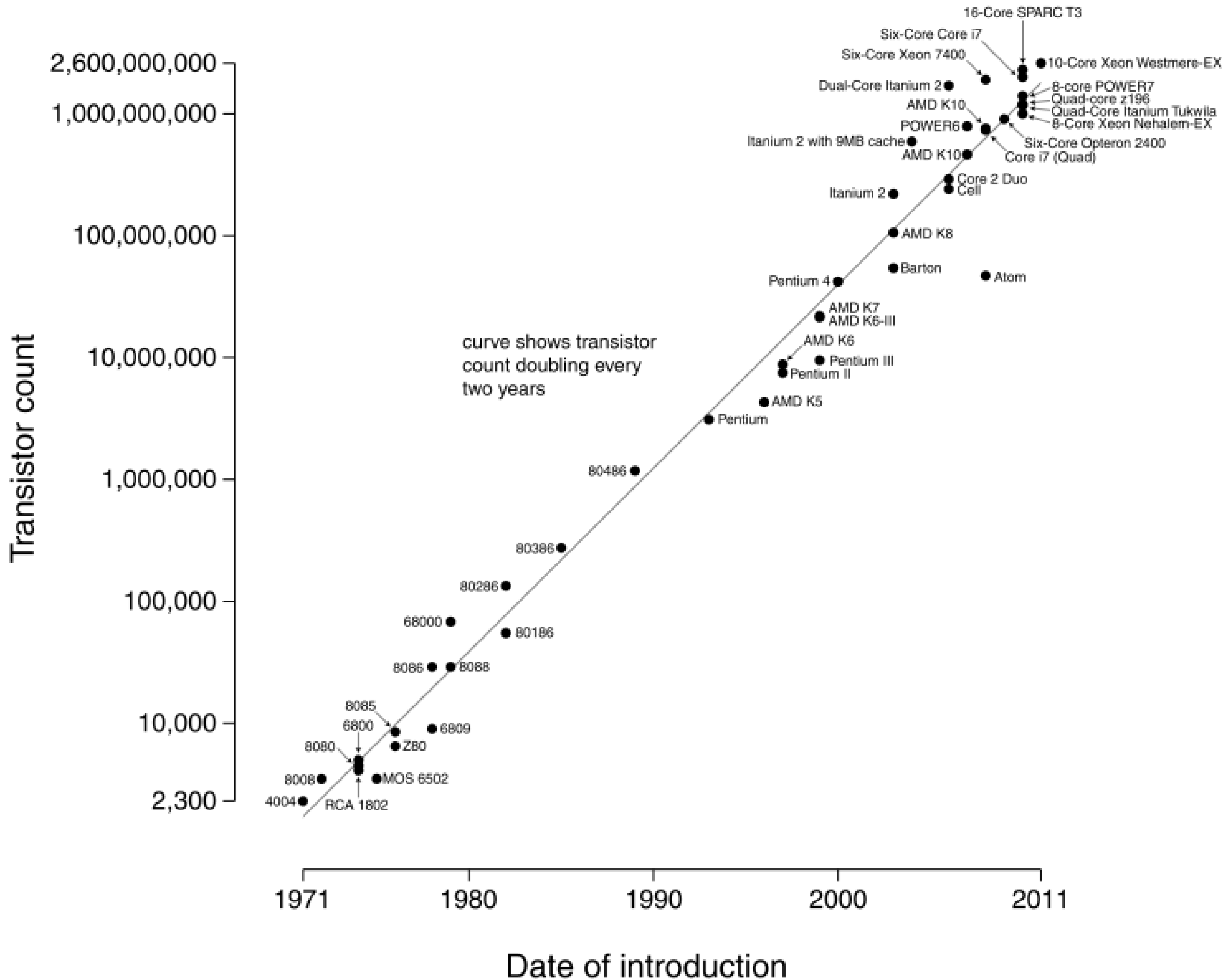
Going up to 12” (30cm)

From http://www.amd.com

In Picture…


Silicon Process Technology
1.5µ
1.0µ
0.8µ
0.6µ
0.35µ
0.25µ
0.18µ
0.13µ
Intel386™ DX Processor
Intel486™ DX Processor
Pentium® Processor
Pentium® Pro Processor
Pentium® II Processor
Pentium® III Processor
Pentium® 4 Processor

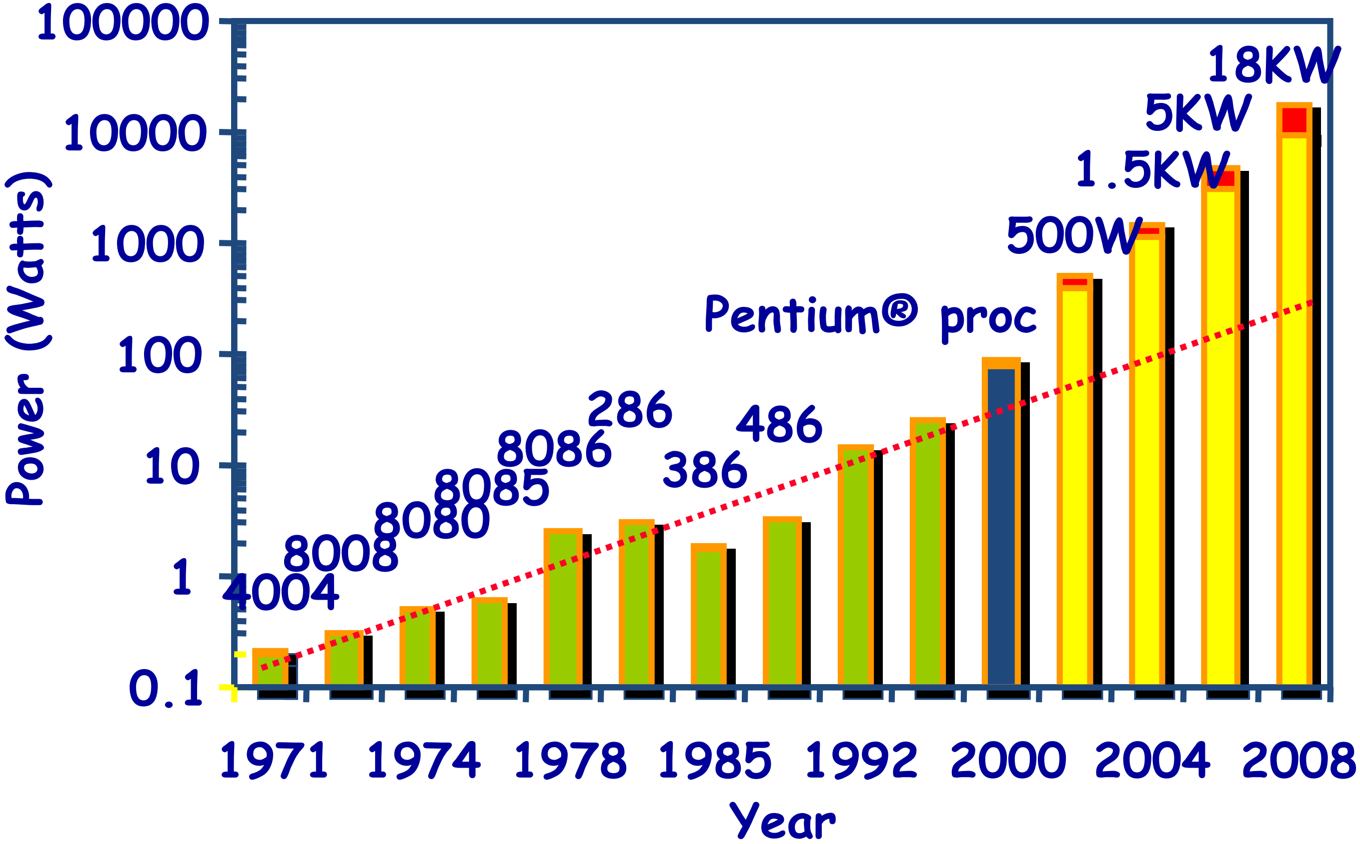
Microprocessor Transistor Counts 1971-2011 & Moore's Law



Number of transistors on an integrated circuit doubles ~ every two years

Image source: Wikipedia

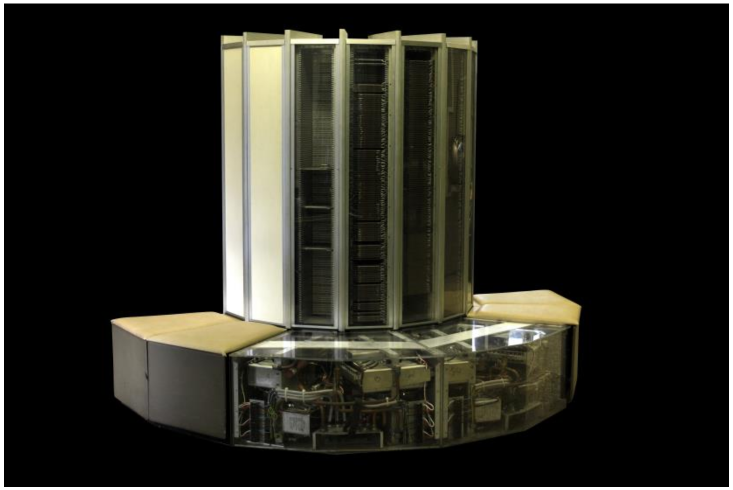
Power will be a major problem

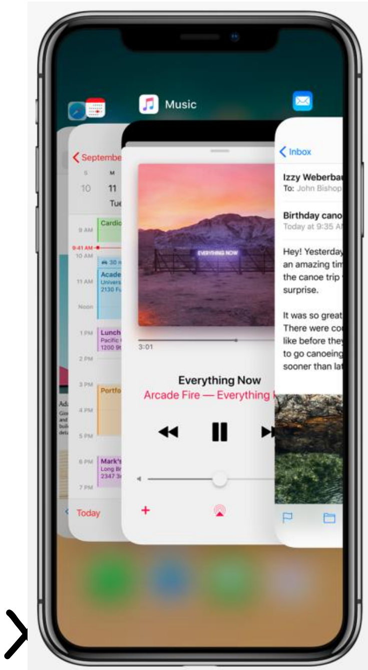


Power delivery and dissipation will be prohibitive

Courtesy, Intel

What Moore's Law Has Meant



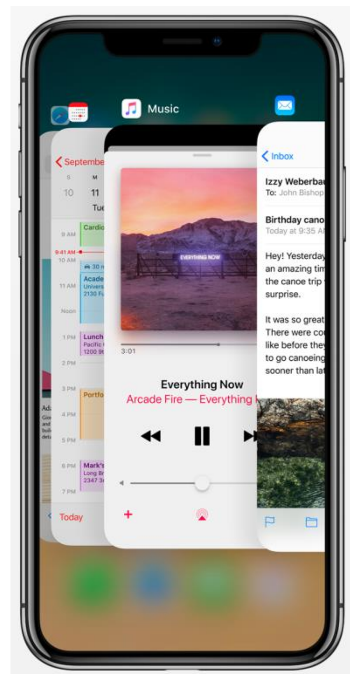


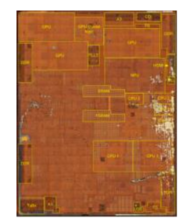
- 1976 Cray 1
  - 250 M Ops/second
  - ~170,000 chips
  - 0.5B transistors
  - 5,000 kg, 115 KW
  - $9M
  - 80 manufactured
- 2017 iPhone X
  - > 10 B Ops/second
  - 16 chips
  - 4.3B transistors (CPU only)
  - 174 g, < 5 W
  - $999
  - ~3 million sold in first 3 days

What Moore’s Law Has Meant

- 1965 Consumer Product

- 2017 Consumer Product



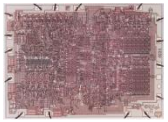


Apple A11 Processor
4.3B transistors

Visualizing Moore's Law to Date

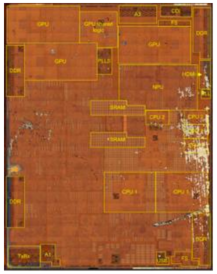
If transistors were the size of a grain of sand

Intel 4004
1970
2,300 transistors



0.1 g

Apple A11
2017
4.3 B transistors



189 kg

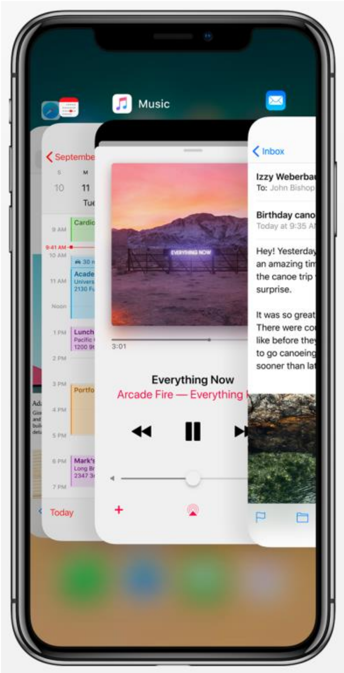
What Moore's Law Has Meant

iPhone	Released with	Release date
iPhone (1st Gen.)	iPhone OS 1.0	June 29, 2007
iPhone 3G	iPhone OS 2.0	July 11, 2008
iPhone 3GS	iPhone OS 3.0	June 19, 2009
iPhone 4	iOS 4.0	June 21, 2010
iPhone 4S	iOS 5.0	October 14, 2011
iPhone 5	iOS 6.0	September 21, 2012
iPhone 5C	iOS 7.0	September 20, 2013
iPhone 5S	iOS 7.0	September 20, 2013
iPhone 6 (Plus)	iOS 8.0	September 19, 2014
iPhone 6S (Plus)	iOS 9.0	September 25, 2015
iPhone SE	iOS 9.3	March 31, 2016
iPhone 7 (Plus)	iOS 10.0	September 16, 2016
iPhone 8 (Plus)	iOS 11.0	September 22, 2017
iPhone X	iOS 11.0.1	November 3, 2017

12 generations of iPhone since 2007

What Moore's Law Could Mean

- 2017 Consumer Product

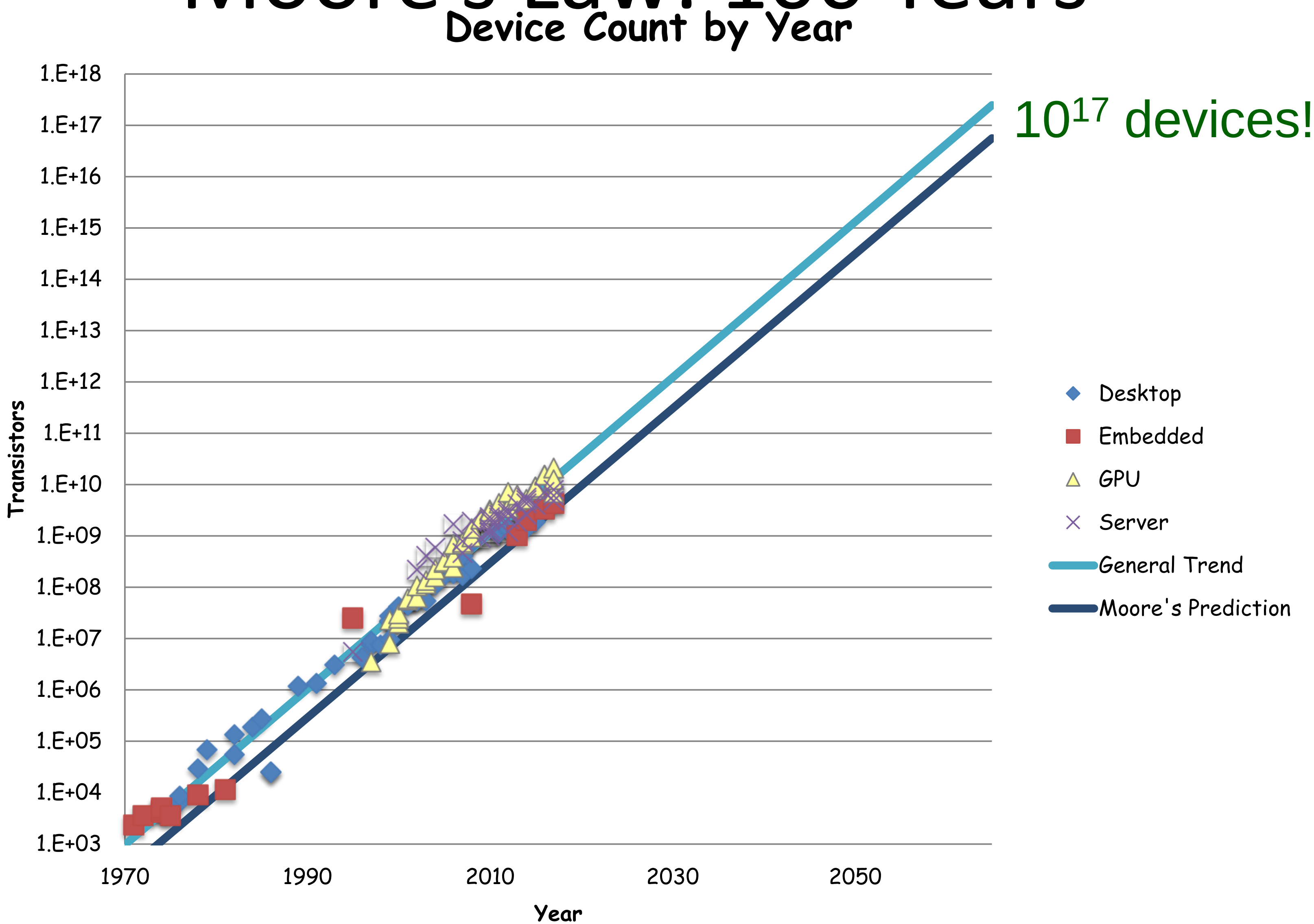


- 2065 Consumer Product

?

  - Portable
  - Low power
  - Will drive markets & innovation

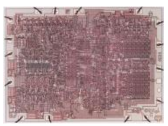
Moore's Law: 100 Years

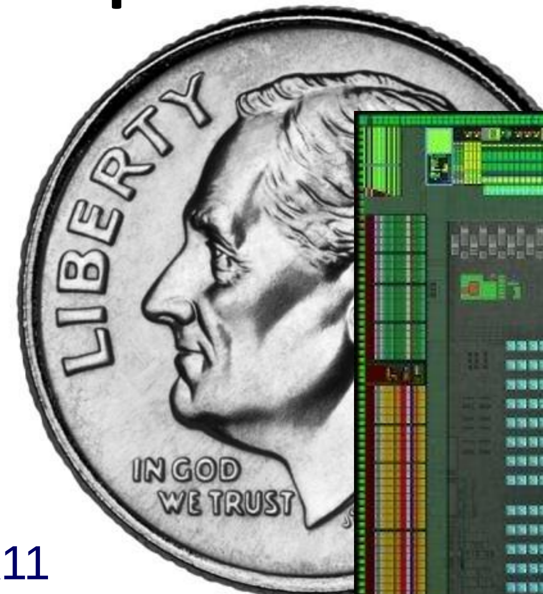


Chips Have Gotten Bigger

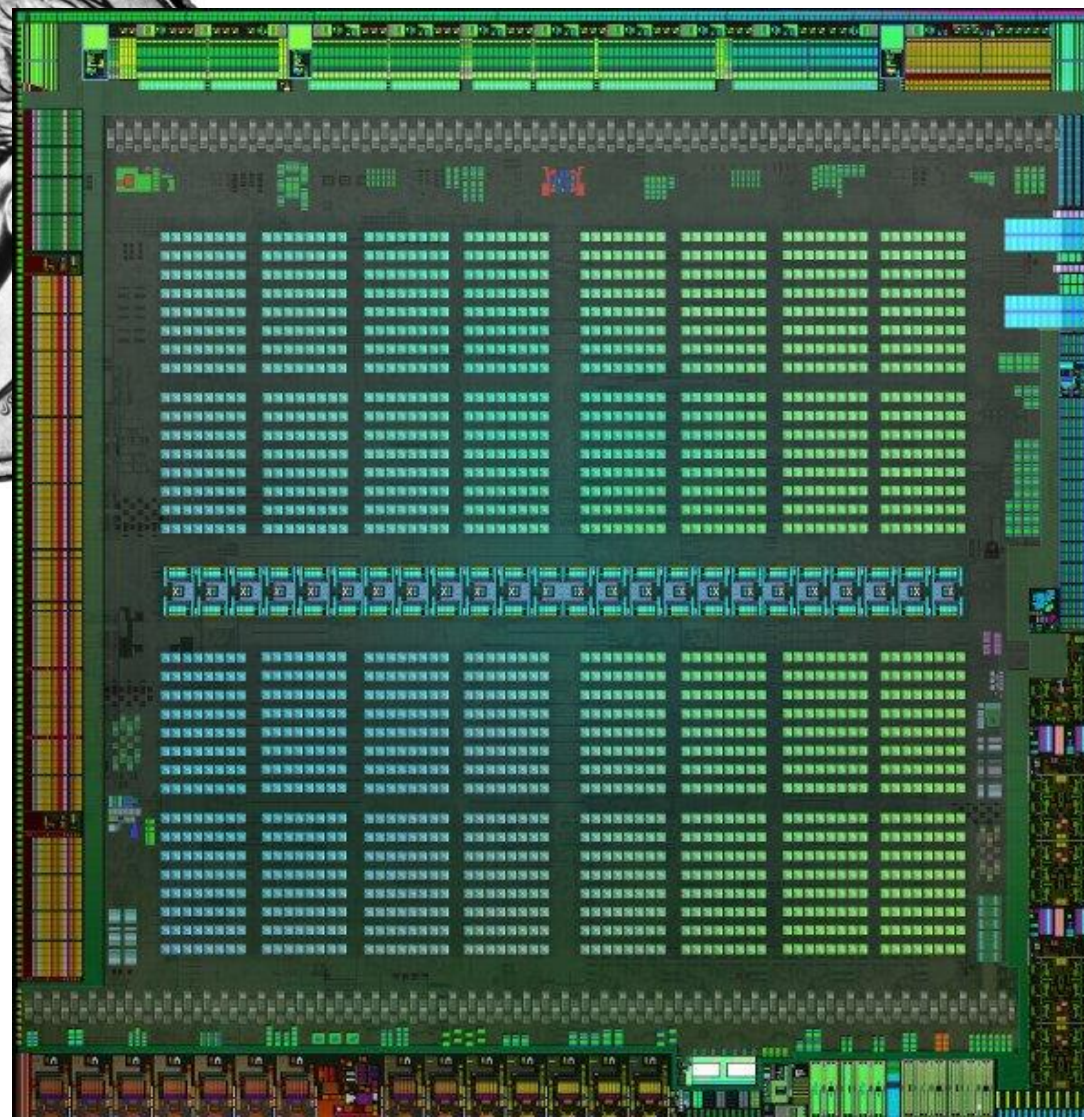
NVIDIA GV100 Volta
2017
21.1 B transistors
815 mm^2

Intel 4004
1970
2,300 transistors
12 mm^2





Apple A11
2017
4.3 B transistors
89 mm^2



NVIDIA GPU Specification Comparison			
	GV100	GP100	GK110
CUDA Cores	5376	3840	2880
Tensor Cores	672	N/A	N/A
SMs	84	60	15
CUDA Cores/SM	64	64	192
Tensor Cores/SM	8	N/A	N/A
Texture Units	336	240	240
Memory	HBM2	HBM2	GDDR5
Memory Bus Width	4096-bit	4096-bit	384-bit
Shared Memory	128KB, Configurable	24KB L1, 64KB Shared	48KB
L2 Cache	6MB	4MB	1.5MB
Half Precision	2:1 (Vec2)	2:1 (Vec2)	1:1
Double Precision	1:2	1:2	1:3
Die Size	**815mm2**	610mm2	552mm2
Transistor Count	21.1B	15.3B	7.1B
TDP	300W	300W	235W
Manufacturing Process	TSMC 12nm FFN	TSMC 16nm FinFET	TSMC 28nm
Architecture	Volta	Pascal	Kepler

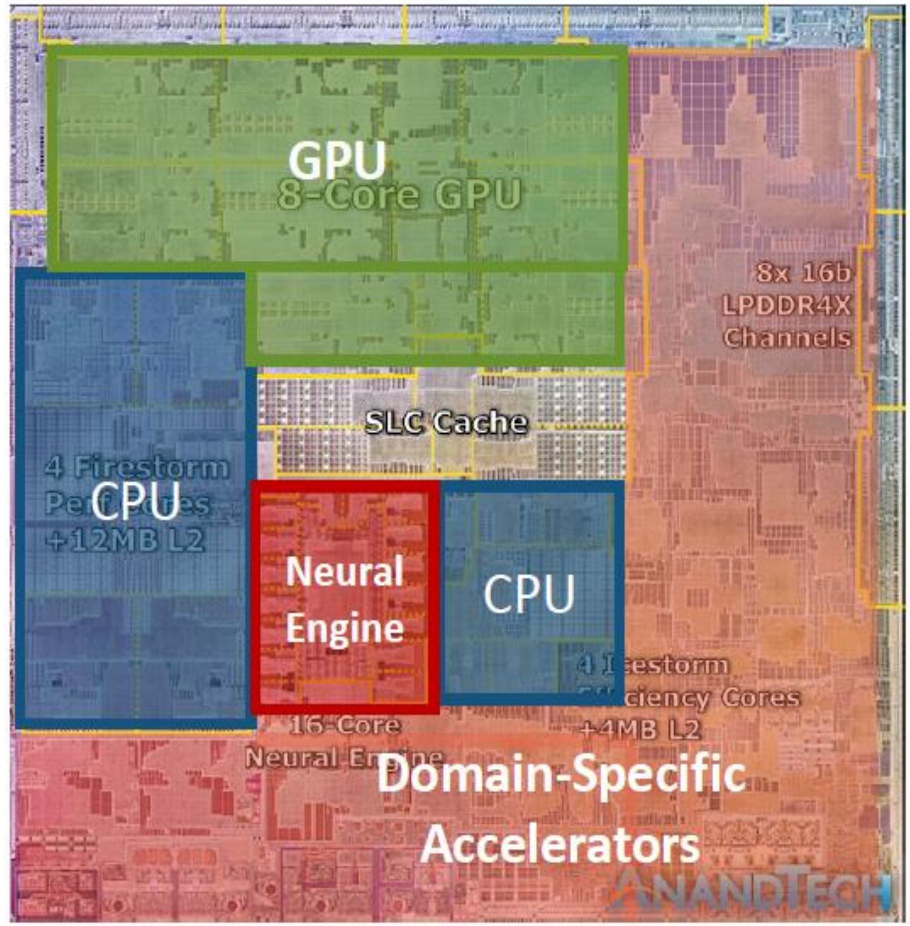
NVIDIA DGX A100

GPUs	**8x NVIDIA A100**
GPU Memory	320 GB total
Peak performance	5 petaFLOPS AI 10 petaOPS INT8
NVSwitches	6
System Power Usage	**6.5kW max**
CPU	**Dual AMD Rome 7742** 128 cores total, 2.25 GHz(base), 3.4GHz (max boost)
System Memory	**1TB**
Networking	**8x Single-Port Mellanox ConnectX-6** 200Gb/s HDR Infiniband (**Compute Network**) 1x (or **2x***) **Dual-Port Mellanox ConnectX-6** 200GB/s HDR Infiniband (**Storage Network** also used for Eth*)
Storage	OS: 2x 1.92TB M.2 NVME drives Internal Storage: 15TB (4x 3.86TB) U.2 NVME drives
Software	Ubuntu Linux OS (**5.3+ kernel**)
System Weight	271 lbs (123 kgs)
Packaged System Weight	315 lbs (143 kgs)
Height	**6U**
Operating temperature range	**5C to 30C** (41F to 86F)

Domain-Specific Accelerators

- Customized hardware designed for a domain of applications.

Apple M1 Chip
2020



This Course

- Micro-architecture: how to implement an architecture in hardware
- Processor:
  - Datapath: functional blocks
  - Control: control signals