CS361: Computer Architecture

Introduction to Computer Architecture

Computer Science

Course Contents

- Introduction to Computer Architecture
- Basics of Computer Architecture
- Detailed Study of different Instruction set types
  - RISC
  - CISC
- CPU Design
- Memory Organization
- Input / Output Design
- And more ..

Resources

- Computer Organization and Design, (4th Edition) by David A. Patterson and John L. Hennessy (5th if available)

Other

- Computer Architecture: A Quantitative Approach, (5th) by John L. Hennessy and David A. Patterson
- Computer Organization and Architecture, (8th Edition) by William Stallings
- Slides uploaded on:
- https://drive.google.com/folderview?id=0B9hP8iqYxX10cGx6N1dWU3pXODQ&usp=sharing

Assessment

- Final – 50
- Midterm – 30
- Assignments and quizzes – 20
- All material from the slides including assignments and anything that is suggested for further reading.

Lecture Outline

- Introduction to Computer Architecture
  - What is Computer Architecture?
  - Basic Operations of a Computer
  - Structure of a Computer
- Evolution
  - Mechanical Systems
  - Electro-Mechanical Systems
  - Electronic Systems
    - Generations
    - Moore's Law
    - The x86 Family
    - Semi-Conductor Memory
    - Performance Gap Between Processors and Memory
    - I/O Devices
- Latest Trends in Chip Organization and Architecture
  - Past: The Single Processor Chip
  - Present : Multicores
  - Future: Manycores

What is Computer Architecture?

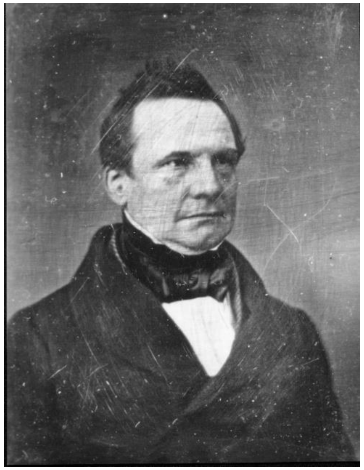
- Architecture is those attributes visible to the programmer
  - These attributes have a direct impact on program execution: instruction set, number of bits used for data representation, I/O mechanisms, addressing techniques.
- More broadly Computer Architecture = Instruction set + hardware + organization

Lecture Outline

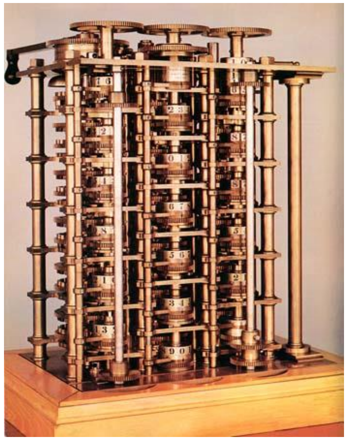
- Introduction to Computer Architecture
  - What is Computer Architecture?
  - Basic Operations of a Computer
  - Structure of a Computer
- Evolution
  - Mechanical Systems
  - Electro-Mechanical Systems
  - Electronic Systems
    - Generations
    - Moore's Law
    - The x86 Family
    - Semi-Conductor Memory
    - Performance Gap Between Processors and Memory
    - I/O Devices
- Latest Trends in Chip Organization and Architecture
  - Past: The Single Processor Chip
  - Present : Multicores
  - Future: Manycores

Mechanical Systems

- The Difference Engine (Paper -1823)
  - A machine to compute mathematical tables.
  - Built later by Swiss engineer Scheutz and his son (displayed in Paris in 1855).
- The Analytical Engine (Paper-1833)
  - First concept of a general purpose computer.
  - Could not be implemented.



Charles Babbage

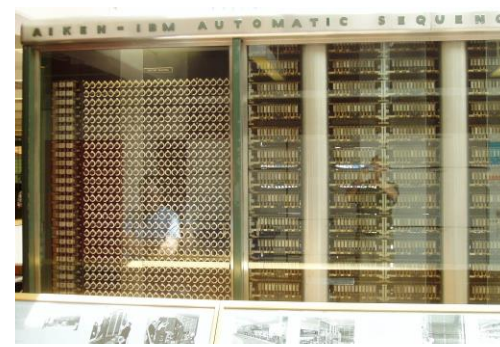


Difference Engine

Electro-Mechanical Systems

- Linear Equation Solver (sometime in 1930s)
  - Tubes and Electromechanical relays
- Harvard Mark I (1944 at IBM laboratories)
  - Howard Aiken – Professor of Physics at Harvard
  - Essentially mechanical but had some electro-magnetically controlled relays and gears
  - Weighed 5 tons and had 750,000 components
  - A synchronizing clock that beat every 0.015 seconds (66Hz)

Linear Equation Solver



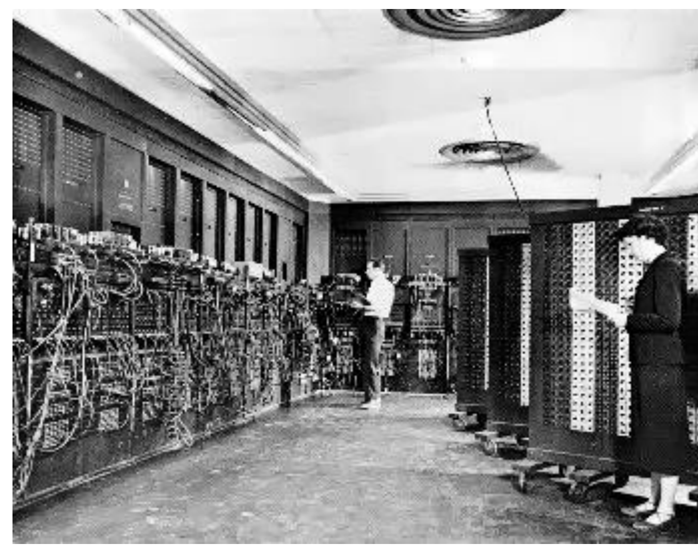
Harvard Mark I

Electronic Computers -- Generations

- 1st Generation– Vacuum Tubes 1946-1957
- 2nd Generation – Transistors 1958-1964
- 3rd Generation – Integrated Circuits Small Scale Integration (SSI) 1965-1971
  - upto 3,000 devices per chip
- 4th Generation – Large Scale Integration (LSI) 1971-1977
  - 3,000 - 100,000 devices on a chip
- 5th Generation – VLSI 1978 -1991
  - 100,000 - 100,000,000 devices on a chip
- 6th Generation – ULSI 1991 onwards
  - Over 100,000,000 devices on a chip

ENIAC - background

- Electronic Numerical Integrator And Computer
- Eckert and Mauchly
- University of Pennsylvania
- Trajectory tables for weapons
- Started 1943 and finished 1946 which was too late for war effort



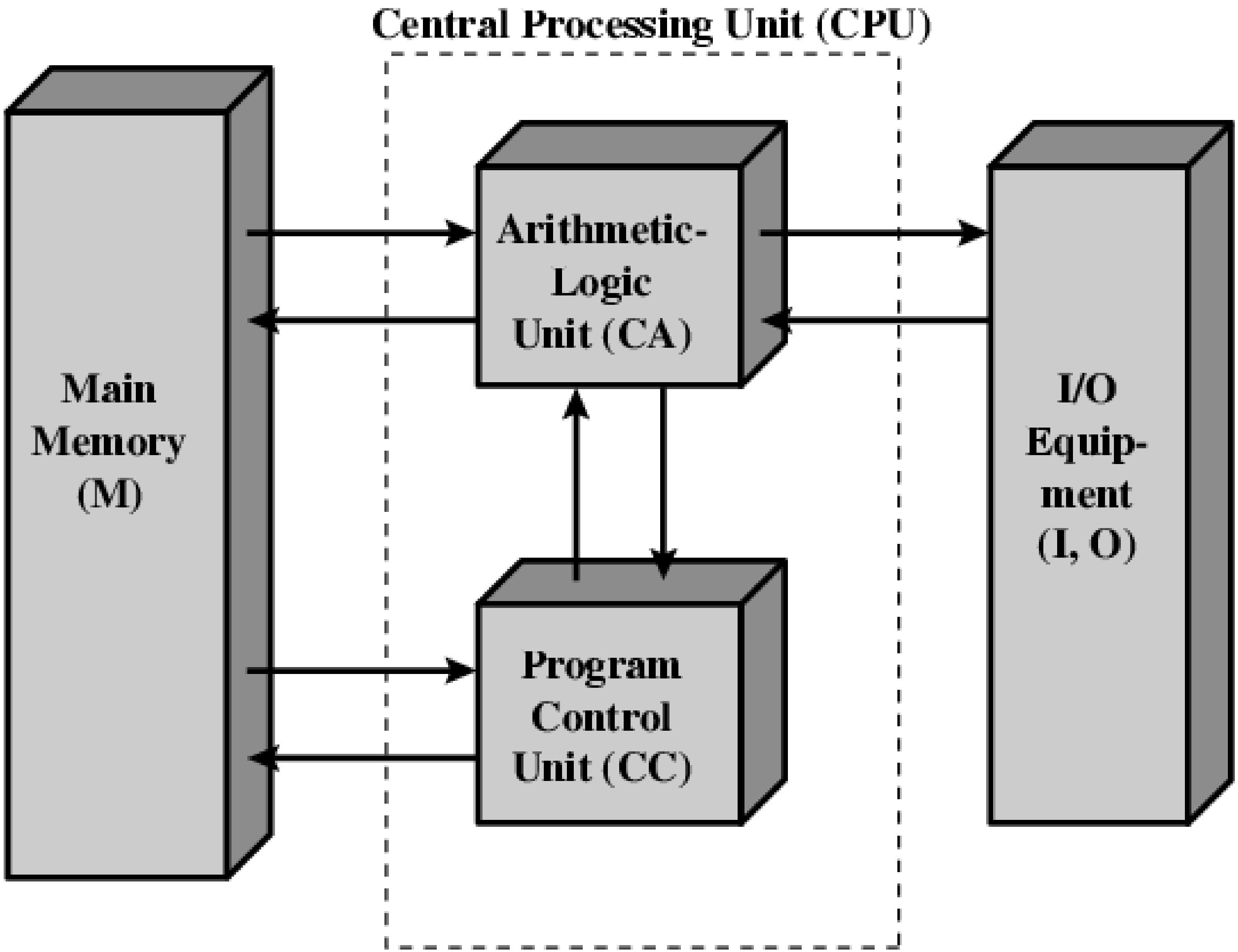
ENIAC - details

- Programmed manually by switches
- Huge: 18,000 vacuum tubes, 30 tons, 15,000 square feet
- 140 kW power consumption
- 5,000 additions per second and used Decimal (not binary)

Electronic Discrete Variable Automatic Computer (EDVAC)

- ENIAC's programming system was external (manually by switches)
- Eckert, Mauchly, John von Neumann and others designed EDVAC to solve this problem
  - Solution was the *stored program computer*
- *First Draft of a report on EDVAC* was published in 1945

Structure of von Neumann machine



Commercial Computers

- 1947 - Eckert-Mauchly Computer Corporation
- UNIVAC I (Universal Automatic Computer)
  - US Bureau of Census 1950 calculations
- Late 1950s - UNIVAC II
  - Faster
  - More memory

Second Generation: Transistors

- Replaced vacuum tubes, made from silicon (sand) Invented 1947 at Bell Labs
- Smaller, Cheaper, Less heat dissipation
- IBM 7000
- DEC - 1957
  - Produced PDP-1

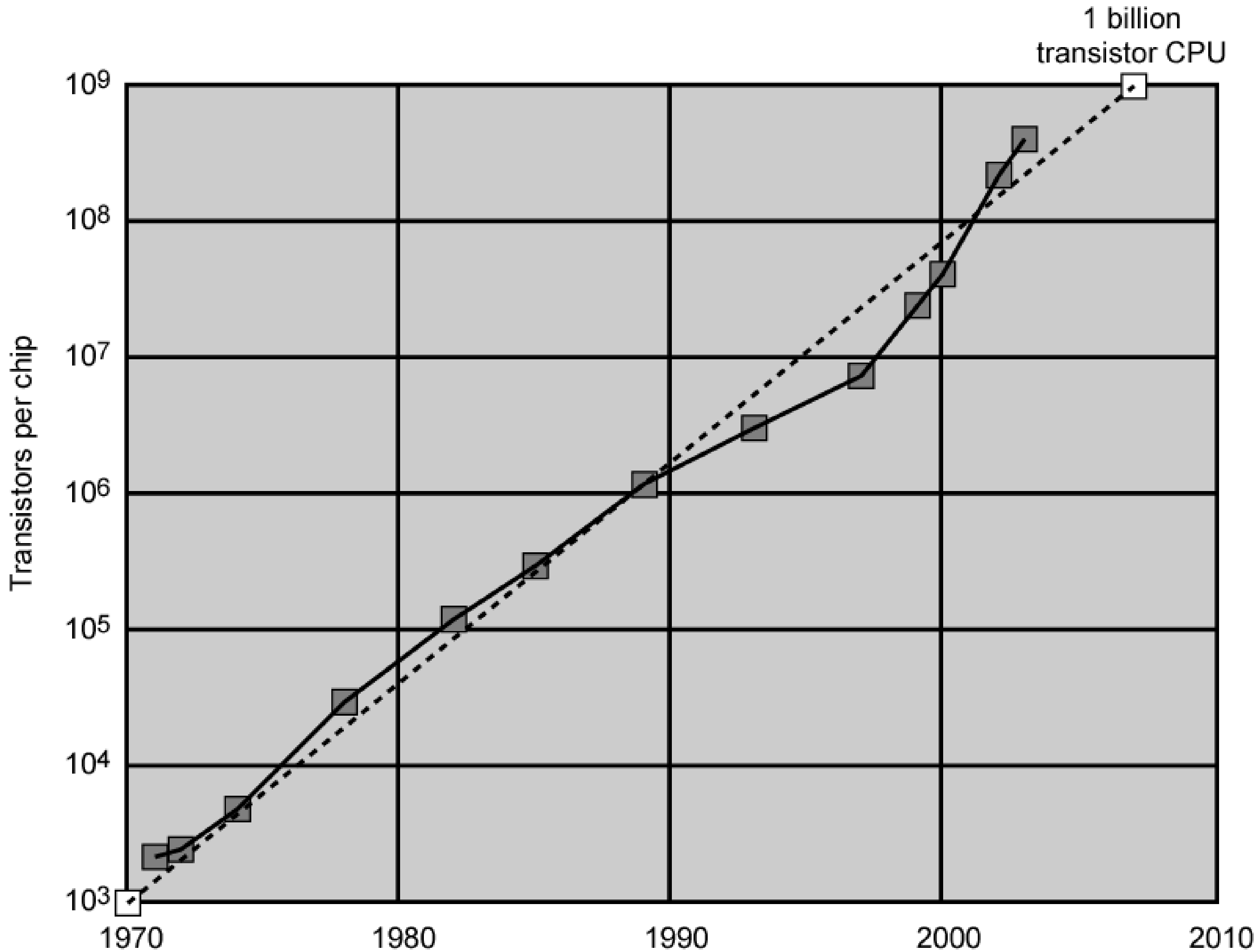
3rd Generation Integrated Circuits: Microelectronics

- Literally - “small electronics”
- A computer is made up of gates, memory cells and interconnections
- These can be manufactured on a semiconductor
- e.g. silicon wafer

Moore's Law

- Increased density of components on chip
- Gordon Moore – co-founder of Intel
- Number of transistors on a chip will double every year
- Since 1970's development has slowed a little
  - Number of transistors doubles every 18 months
- Cost of a chip has remained almost unchanged
- Higher packing density means shorter electrical paths, giving higher performance
- Smaller size gives increased flexibility
- Reduced power and cooling requirements
- Fewer interconnections increases reliability

Growth in CPU Transistor Count

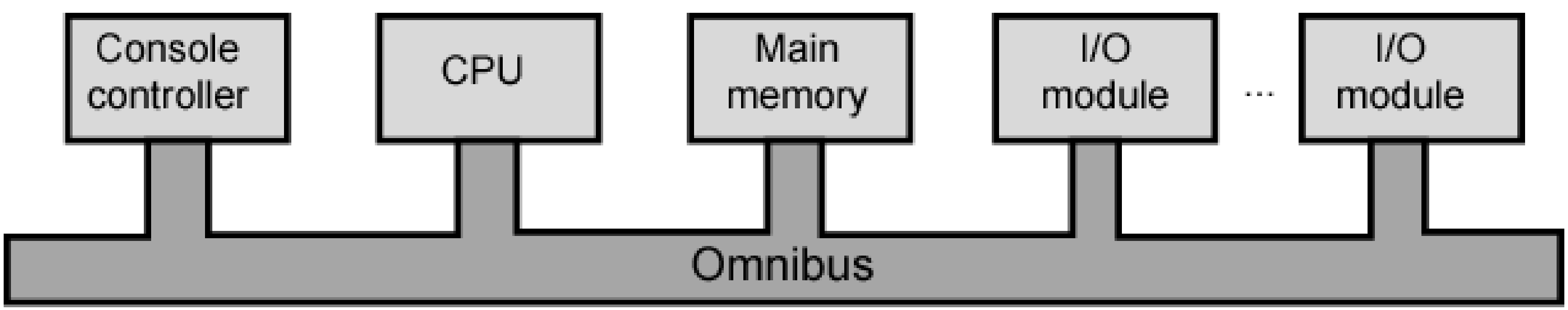


IBM 360 series (started in 1964)

- Replaced (& not compatible with) 7000 series
- First planned "family" with a range of computers
  - Similar or identical instruction sets
  - Similar or identical O/S
  - More expensive systems had a higher speed, increasing number of I/O ports, increased memory size and vice versa.

DEC PDP-8(Developed in 1964)

- First minicomputer
- Did not need air conditioned room and Small enough to sit on a lab bench
- $16,000 as compared to 100k+ for IBM 360
- Embedded applications
- Introduced BUS STRUCTURE



Intel

- 1971 - 4004
  - First microprocessor
  - All CPU components on a single chip
  - 4 bit
- Followed in 1972 by 8008
  - 8 bit
  - Both designed for specific applications
- 1974 - 8080
  - Intel's first general purpose microprocessor

x86 Evolution (1)

- 8080
  - first general purpose microprocessor
  - 8 bit data path
  - Used in first personal computer – Altair
- 8086 – 5MHz – 29,000 transistors
  - much more powerful
  - 16 bit
  - instruction cache, prefetch few instructions
- 80286
  - 16 Mbyte memory addressable
  - up from 1Mb
- 80386
  - 32 bit
  - Support for multitasking
- 80486
  - sophisticated powerful cache and instruction pipelining
  - built in maths co-processor

x86 Evolution (2)

- Pentium
  - Superscalar
  - Multiple instructions executed in parallel
- Pentium Pro
  - Increased superscalar organization
  - branch prediction
  - data flow analysis
  - speculative execution
- Pentium II
  - MMX technology for graphics, video & audio processing
- Pentium III
  - Additional floating point instructions for 3D graphics

x86 Evolution (3)

- Pentium 4
  - Further floating point and multimedia enhancements
- Core
  - First x86 with dual core
- Core 2
  - 64 bit architecture
- Core 2 Quad – 3GHz – 820 million transistors
  - Four processors on chip

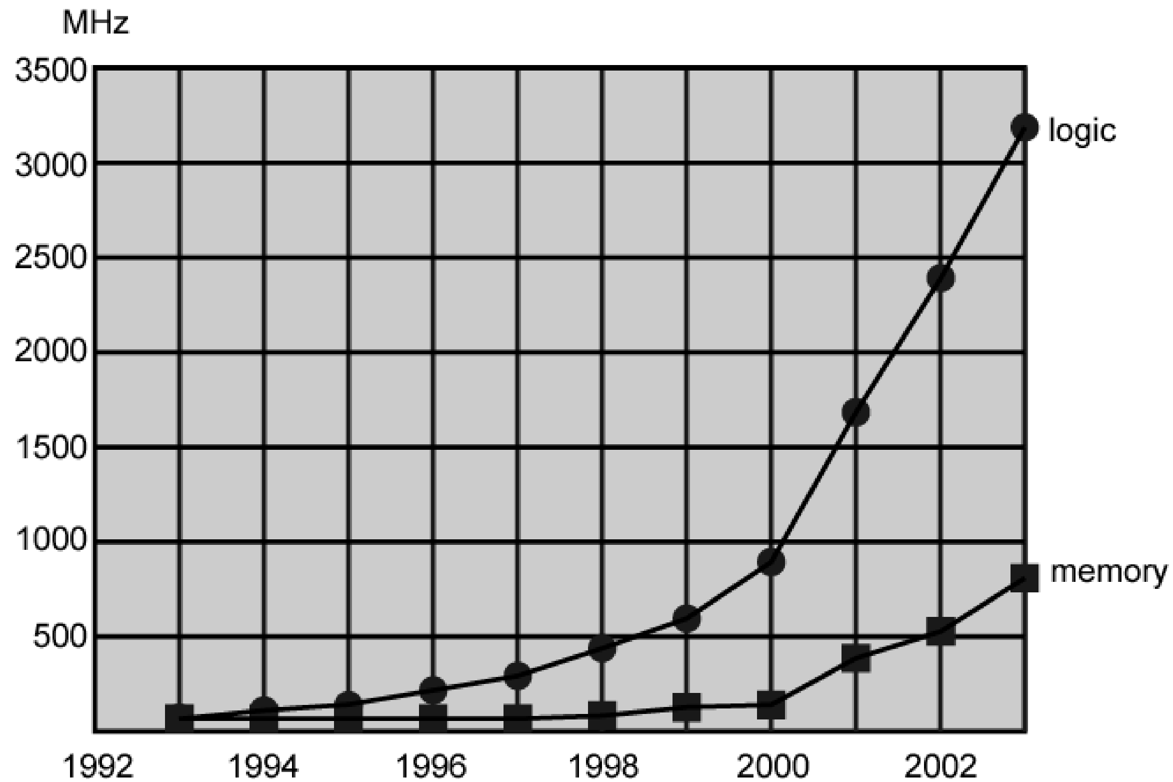
- x86 architecture dominant outside embedded systems
- Organization and technology changed dramatically
- Instruction set architecture evolved with backwards compatibility
- ~1 instruction per month added

Semiconductor Memory

- In 1950s and 1960s magnetic core (or core was used) which was expensive and bulky.
- Introduced in 1970
- Holds 256 bits
- Non-destructive read
- Much faster than magnetic core
- Capacity approximately doubles each year

Processor and Memory Performance Gap

- Processor speed increased
- Memory capacity increased
- Memory speed lags behind processor speed



Solutions

- Increase number of bits retrieved at one time
  – Make DRAM “wider” rather than “deeper”
- Change DRAM interface
  – Cache
- Reduce frequency of memory access
  – More complex cache and cache on chip
- Increase interconnection bandwidth
  – High speed buses
  – Hierarchy of buses

I/O Devices

- Peripherals with intensive I/O demands
- Large data throughput demands
- Processors can handle this
- Problem moving data
- Solutions:
  - Caching
  - Buffering
  - Higher-speed interconnection buses
  - More elaborate bus structures
  - Multiple-processor configurations

Lecture Outline

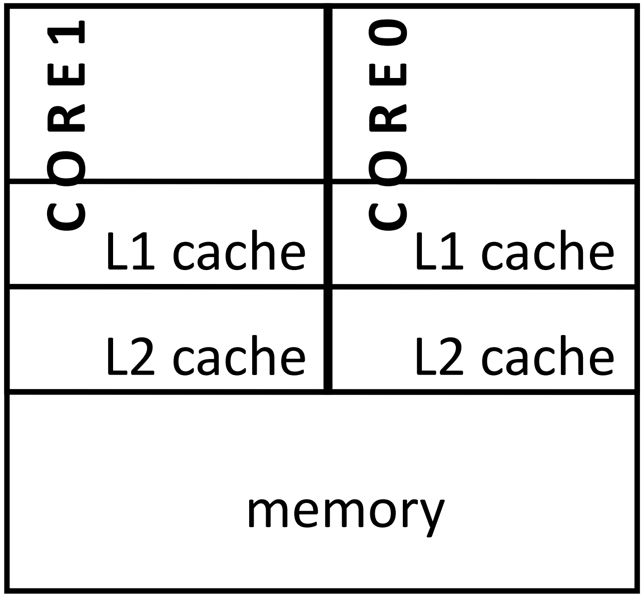
- Introduction to Computer Architecture
  - What is Computer Architecture?
  - Basic Operations of a Computer
  - Structure of a Computer
- Evolution
  - Mechanical Systems
  - Electro-Mechanical Systems
  - Electronic Systems
    - Generations
    - Moore's Law
    - The x86 Family
    - Semi-Conductor Memory
    - Performance Gap Between Processors and Memory
    - I/O Devices
- Latest Trends in Chip Organization and Architecture
  - Past: The Single Processor Chip
  - Present : Multicores
  - Future: Manycores

Past: The Single Processor Chip

- Performance scaling accomplished largely through increases in clock frequency
- Problems
  - Heat Dissipation
  - Power Consumption
  - High Cost required for little gain in performance
- In 2004, Intel cancelled two single core processors because of these problems. Other vendors had already moved onto multicores at the time.

Present: Multicores

- Number of cores per chip: 2-10s
- Heterogeneous processors: Specialized processors
- Features relevant for programming are
  - Shared and coherent caches
  - Single Address space
  - Non-uniform memory access (NUMA): Memory can be distributed for more scalability.



Future: Manycores

- Massively parallel architectures
- Number of cores: 1000s and more
- Memory
  - Multiple address spaces
  - Non-uniform access timings
  - On-chip addressable memory
  - Islands of cache coherence