Computer Organization & Architecture

Chapter 1 – Basic Structure of Computers

Zhang Yang 张杨

cszyang@scut.edu.cn

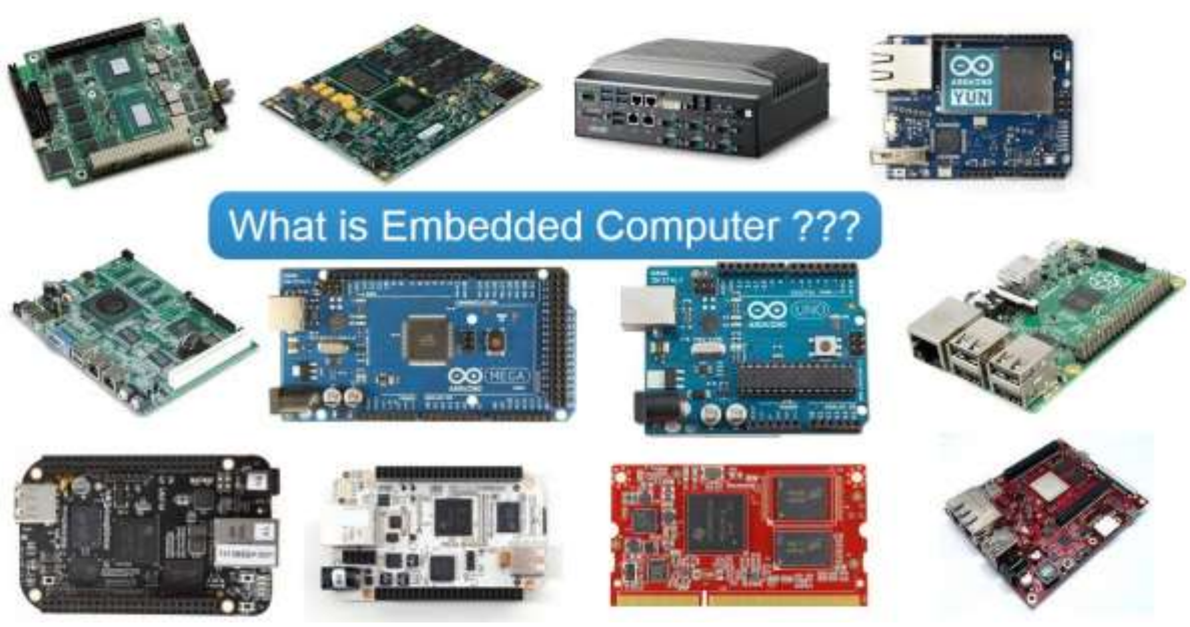
Autumn 2025

Content of this lecture

- 1.1 Computer Types
- 1.2 Functional Units
- 1.3 Basic Operational Concepts
- 1.6 Performance
- 1.7 Historical Perspective
- Summary

Embedded Computers (1)

- An embedded computer is a microcontroller or micro-processor based system devised for specific function.
- Integrated into a larger device or system.



Embedded Computers (2)

- Characteristics
  - Small Form-factor: generally single high-density PCB
  - Lower power components
  - Minimal upgradeability/expansions
  - Low hardware cost
- Examples
  - Calculators
  - Digital Cameras
  - Vending Machines
  - Elevators
  - Copiers
  - Printers
  - GPS
  - …

Personal Computers (1)

- A small, general purpose computer that is created to be utilized by one person at a time.
- General computation, document preparation, computer-aided design, audiovisual entertainment, …
- Types
  - Workstation
    - High-end personal computer.
    - Used for tasks such as computer-aided design, drafting and modeling, computation-intensive scientific and engineering calculations, image processing, architectural modeling, and computer graphics for animation and motion picture visual effects.

Personal Computers (2)

- Types (ctd.)
  - Desktop computers
    - Gaming computers
    - Single-unit PC(all-in-one PC)
  - Portable computers
    - Notebook (Laptop) computers
    - Netbook computers
      - Small, lightweight and inexpensive laptop computers suited for general computing tasks and accessing web-based applications.
    - Tablet

Servers and Enterprise Systems

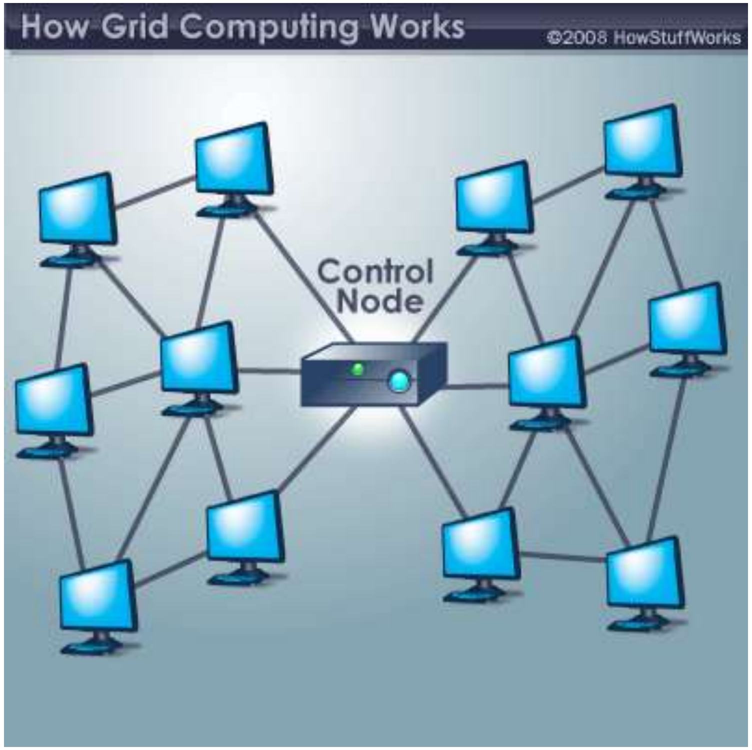
- Servers
  - A server is a computer that provides data to other computers.
  - Shared by a potentially large number of users.
  - Contain sizable database storage units.
  - Capable of handling large volumes of requests to access the data.
  - Business data processing in medium to large corporations, information processing for a government agency,…

Supercomputers and Grid Computers (1)

- Supercomputers
  - A supercomputer is a type of computer that has the architecture, resources and components to achieve massive computing power.
  - FLOPS
    - Floating-point operations per second
    - Measurement of computing performance of a supercomputer.
  - Used in large-scale numerical calculations required in applications weather forecasting, aircraft design and simulation.

Supercomputers and Grid Computers (2

- Grid Computers
  - Grid computers combine a large number of personal computers and disk storage units in a physically distributed high-speed network.



A group of computers that coordinate to solve a problem together.

Cloud Computing

- Cloud Computing is nothing but the delivery of computing services such as a database, networking, software, storage, servers, and many more.

- Question: Grid Computing vs. Cloud Computing

Content of this lecture

- 1.1 Computer Types
- 1.2 Functional Units
- 1.3 Basic Operational Concepts
- 1.6 Performance
- 1.7 Historical Perspective
- Summary

Basic Functions of a Computer

- Data Processing
  - Data means any digital information
- Control
- Data Storage
- Data Movement

Basic Functional Units of a Computer (1)

- Arithmetic and Logic Unit (ALU)
- Control Unit
- Memory
- Input Unit
- Output Unit

Processor

I/O Unit

Basic Functional Units of a Computer (2)

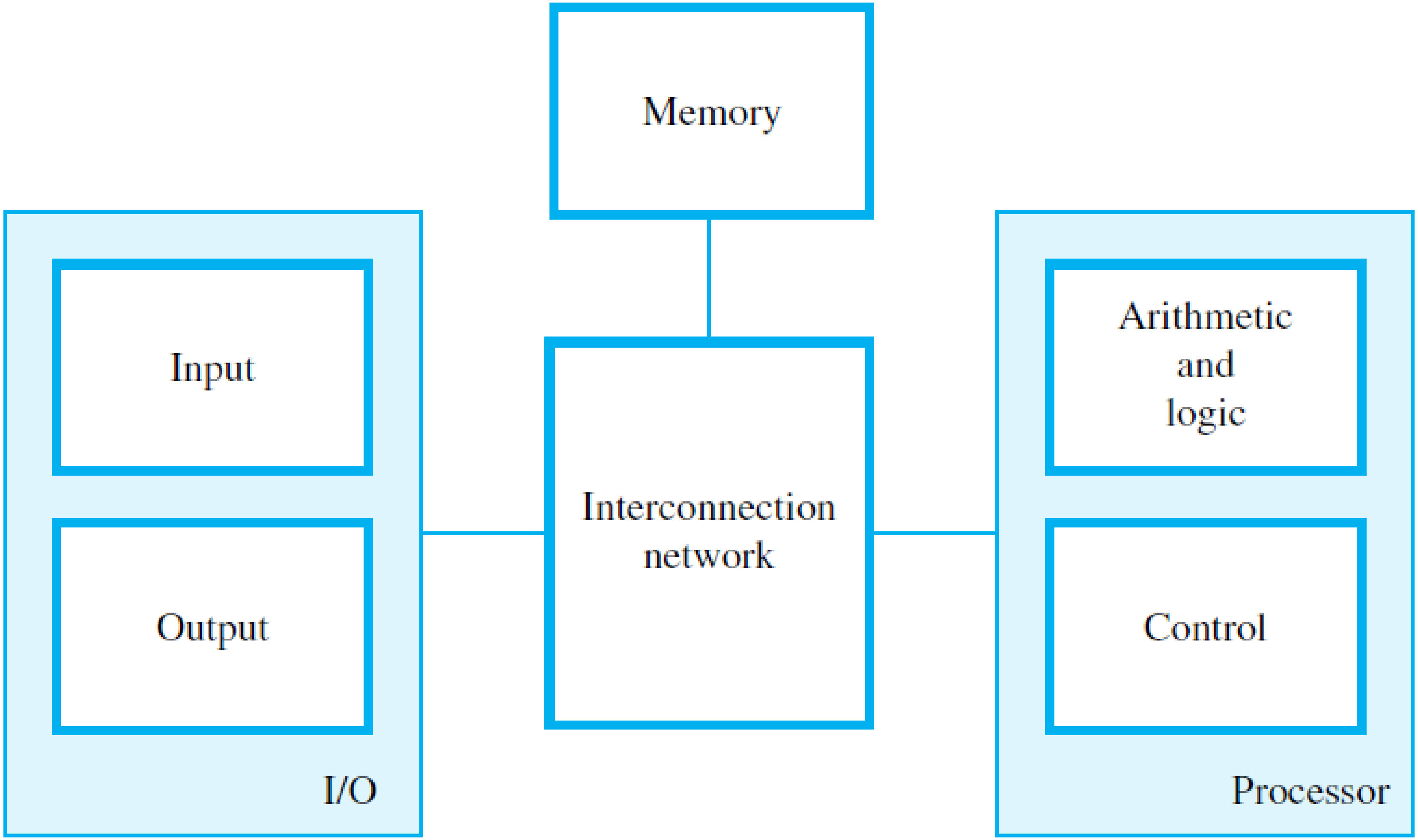


Figure 1.1 Basic functional units of a computer.

Information Handled by a Computer

- Instructions/Machine Instructions
  - Govern the transfer of information within a computer as well as between the computer and its I/O devices
  - Specify the arithmetic and logic operations to be performed
  - Program: a list of instructions which performs a task
- Data
  - Numbers and characters used as operands by the instructions
- Encoded in binary code – 0 and 1

Input Units

- Function
  - Make computer accept coded information
- Input Devices
  - Keyboard
  - Mouse
  - Digital camera
  - Joystick
  - Touchpad
  - Scanner
  - Microphone
  - …

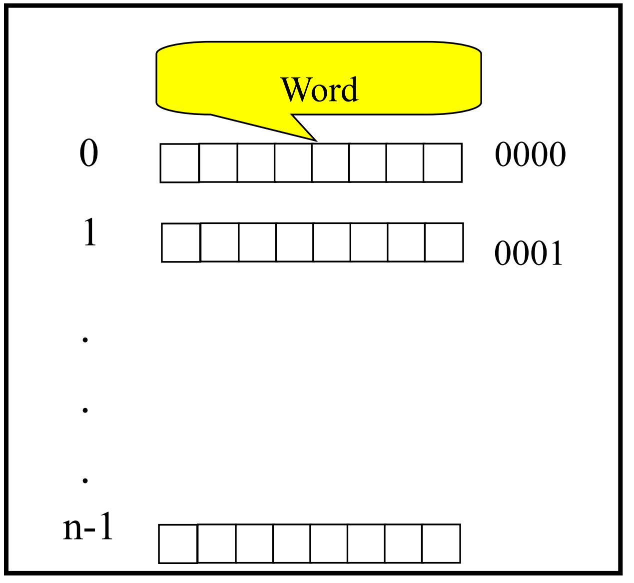
Output Units

- Function
  - Send processed results to the outside world
- Output Devices
  - Video display
  - Printer
  - …

Memory Unit (1)

- Function
  - Store instructions and data
- Classification of Storage
  - Primary Storage
    - Store programs when they are being executed
    - Related terms
      - Word
      - Word length
      - Address
      - Memory access time
      - Random access memory

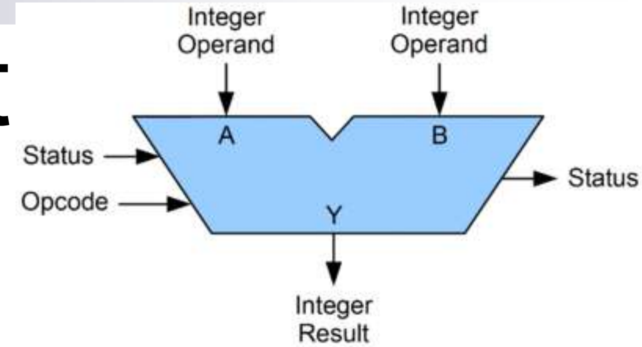
Memory



Memory Unit (2)

- Classification of Storage
  - Secondary Storage
    - Store many programs and large amounts of data, particularly for information that is accessed infrequently.
    - Magnetic Disks and Tapes
      - Hard disks
      - Floppy disks
    - Optical Disks
      - CD-ROMs

Arithmetic and Logic Unit



- Function
  - Execute arithmetic or logic operations
- Most computer operations are executed in ALU of the processor.
- Load the operands into memory – bring them to the processor – perform operation in ALU – store the result back to memory or retain in the processor.
- Registers in ALU: high speed storage elements

Control Unit

- Functions
  - Controls sequential instruction execution
  - Interprets instructions
  - Guides data flow through different computer areas
  - Regulates and controls processor timing
  - Sends and receives control signals from other computer devices
  - Handles multiple tasks, such as fetching, decoding, execution handling and storing results

Operations of a Computer

- The computer accepts information in the form of programs and data through an input unit and stores it in the memory.
- Information stored in the memory is fetched, under program control, into an arithmetic and logic unit where it is processed.
- Processed information leaves the computer through an output unit.
- All activities inside the machine are directed by the control unit.

Content of this lecture

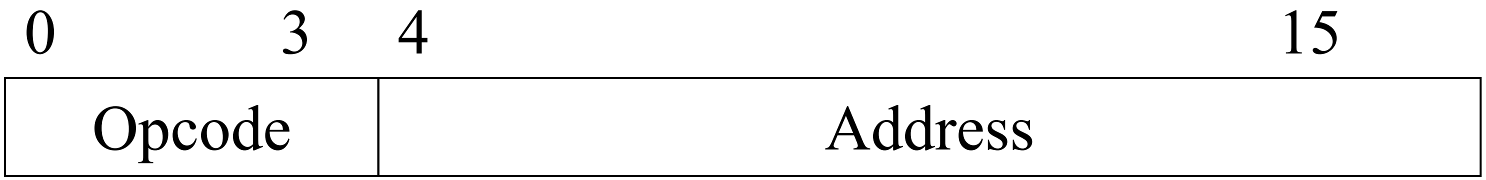
- 1.1 Computer Types
- 1.2 Functional Units
- 1.3 Basic Operational Concepts
- 1.6 Performance
- 1.7 Historical Perspective
- Summary

Program Execution (1)

- Execute instructions specified in the program.
  - 1.The processor reads (fetch) instructions from memory one at a time, and executes each instruction.
  - 2.Program execution consists of repeating the process of instruction fetch and instruction execution.

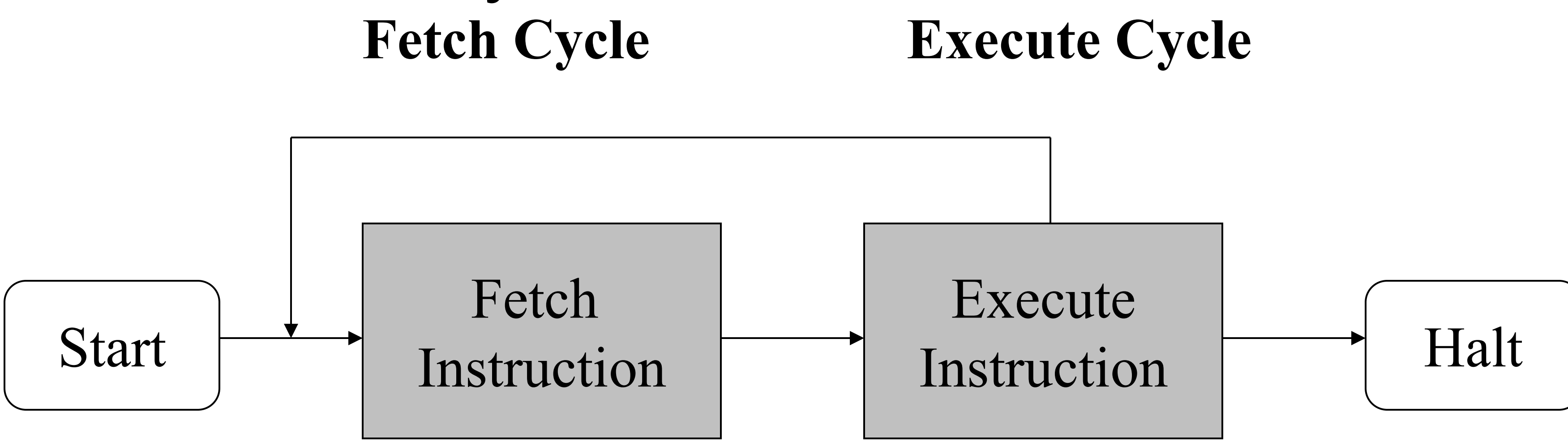
Program Execution (2)

- Example of an Instruction



An Example Instruction Format

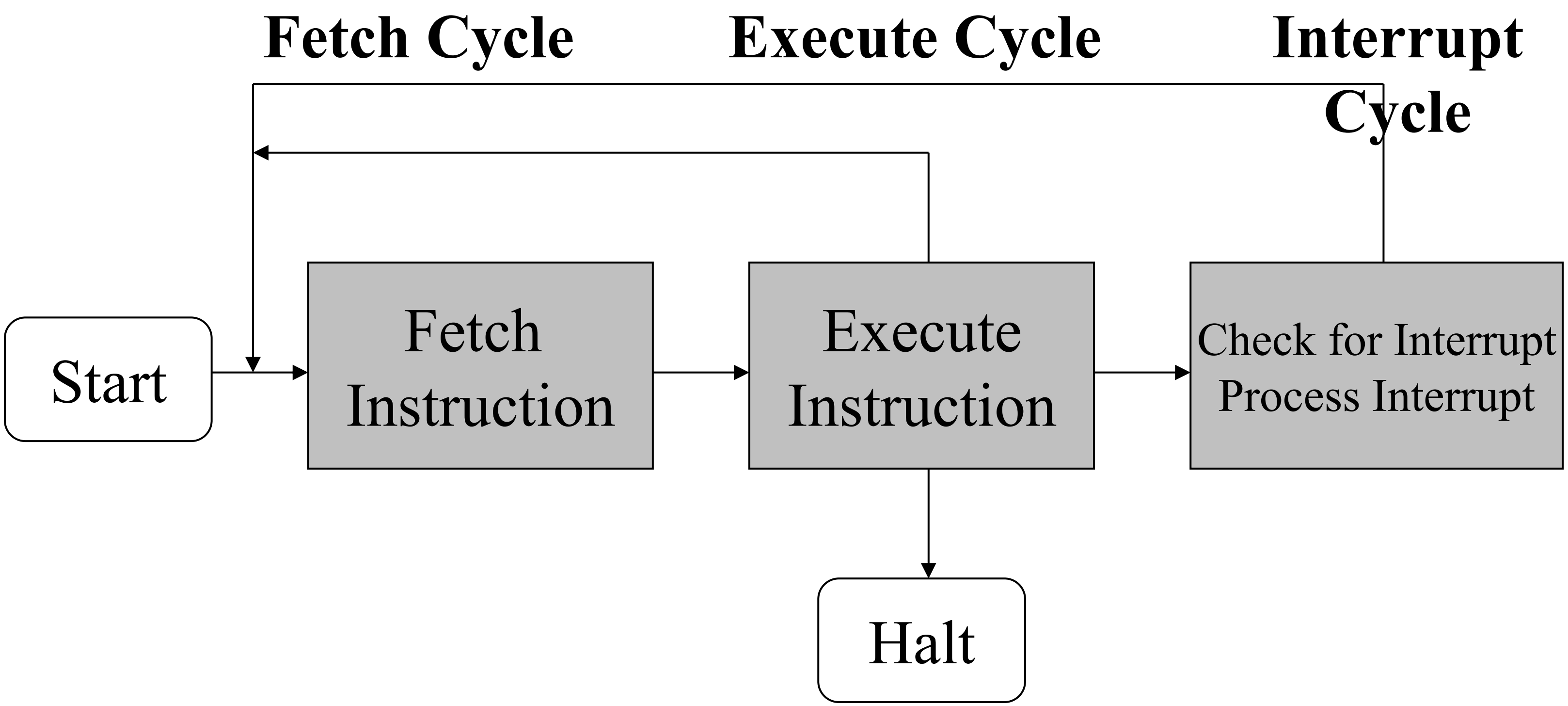
- Instruction Cycle



Instruction Cycle

Program Execution (3)

- Instruction Cycle with Interrupt
  - Interrupt: An interrupt is a request from an I/O device for service by the processor.



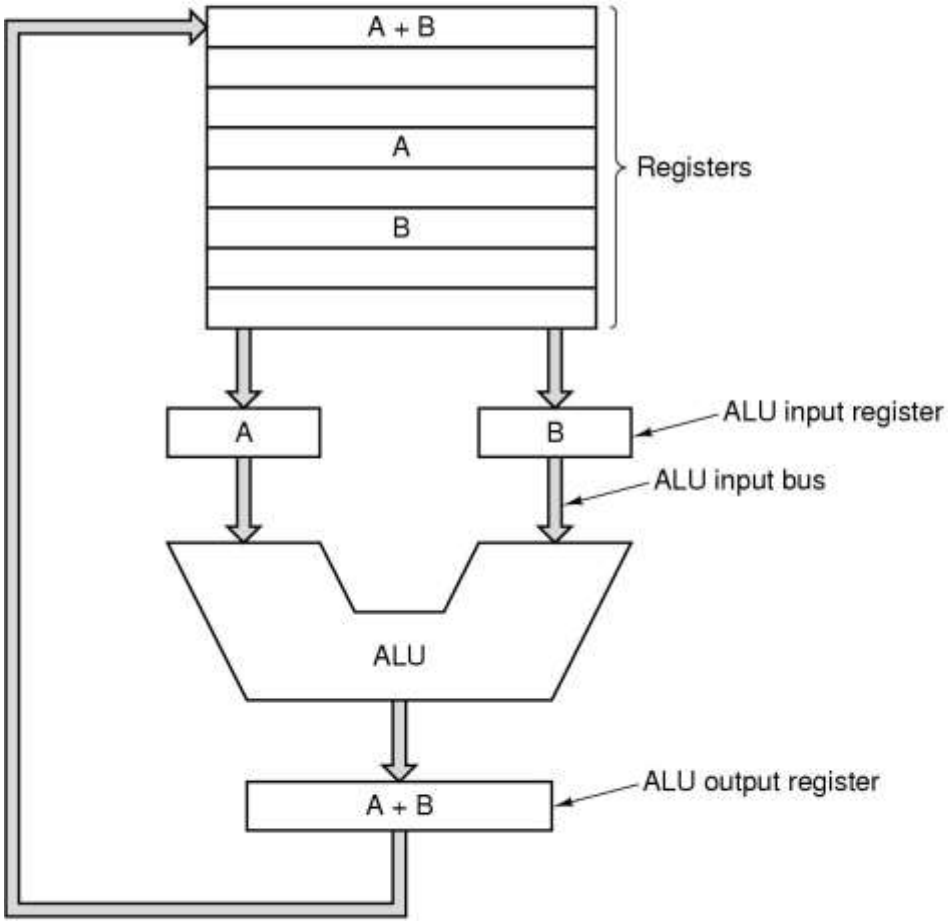
Program Execution (4)

- Instruction Execution
  - Example: *Add LOCA, R0*
    - Fetch the instruction from the memory into the processor(IR).
    - The operand at LOCA is fetched and added to the contents of R0 (combines a memory access operation with an ALU operation).
    - The resulting sum is stored in register R0.
  - *Load LOCA, R1*

    Add R1, R0

Program Execution (5)

- Instruction Execution (ctd.)



Program Execution (6)

- Instruction Execution (ctd.)

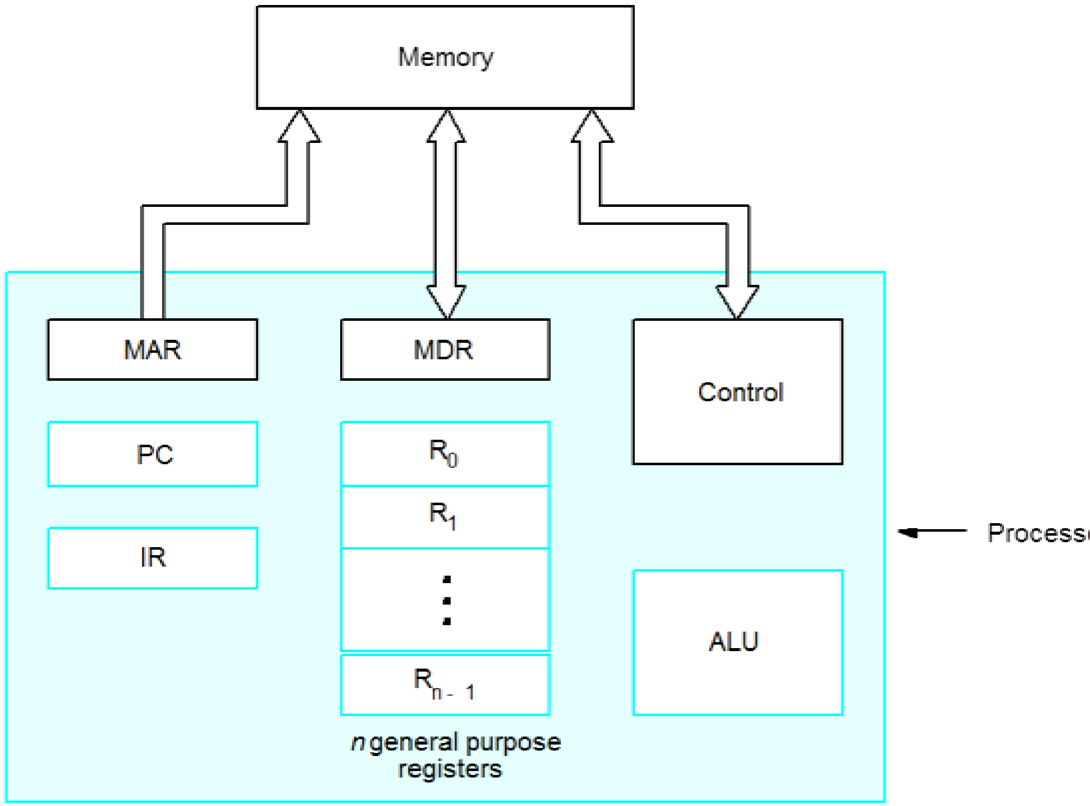


Figure 1.2. Connections between the processor and the memory.

Program Execution (7)

- Instruction Execution (ctd.)
  - Normal steps of an instruction execution

(1) MAR← [PC], Read

(2) WMFC, MDR← [MAR], PC← [PC]+1

(3) IR← [MDR]

If an operand resides in the memory, then (4) to (7), else (7)

(4) MAR← LOCA, Read

(5) WMFC, MDR← [MAR]

(6) Ri← [MDR]

(7) ALU performs operation

If the result is to be stored in the memory, then (8)

(8) MDR← result

(9) MAR← address, Write

Content of this lecture

- 1.1 Computer Types
- 1.2 Functional Units
- 1.3 Basic Operational Concepts
- 1.6 Performance
- 1.7 Historical Perspective
- Summary

Measuring Performance

- Measuring Performance
  - How quickly a computer can execute programs
    - Design of instruction set
    - Hardware, including hardware fabrication technology
    - Software, including OS
    - Compiler
  - MIPS
    - Millions of Instructions per second
    - Measurement of how many machine instructions a processor execute per second.

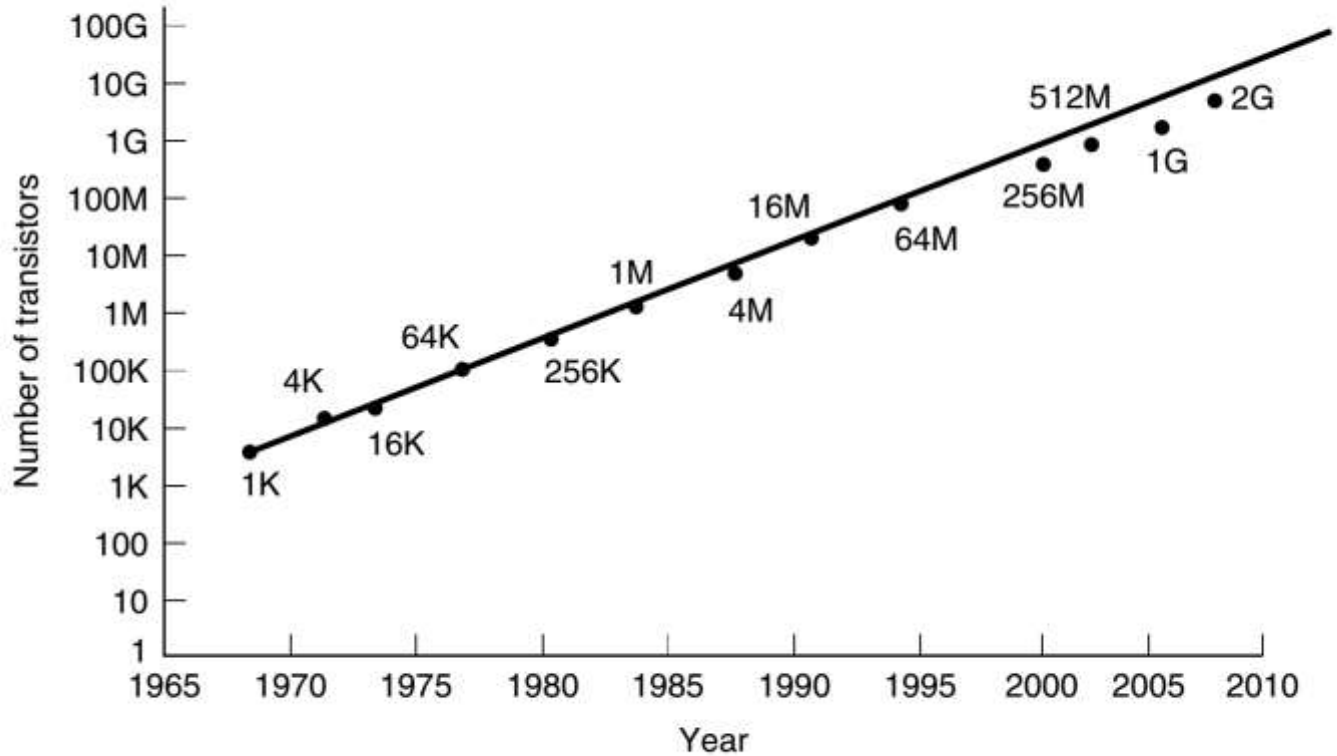
Technology (1)

- Chip manufacturers pack more and more transistors per chip every year.
- Moore’s Law (1965)
    - Integrated circuit density-doubling would occur every 24 months, integrated circuits would double in performance every 18 months.
- Moore’s law has long been associated with the number of bits in a memory, it applies equally well to CPU chips.

Technology (2)

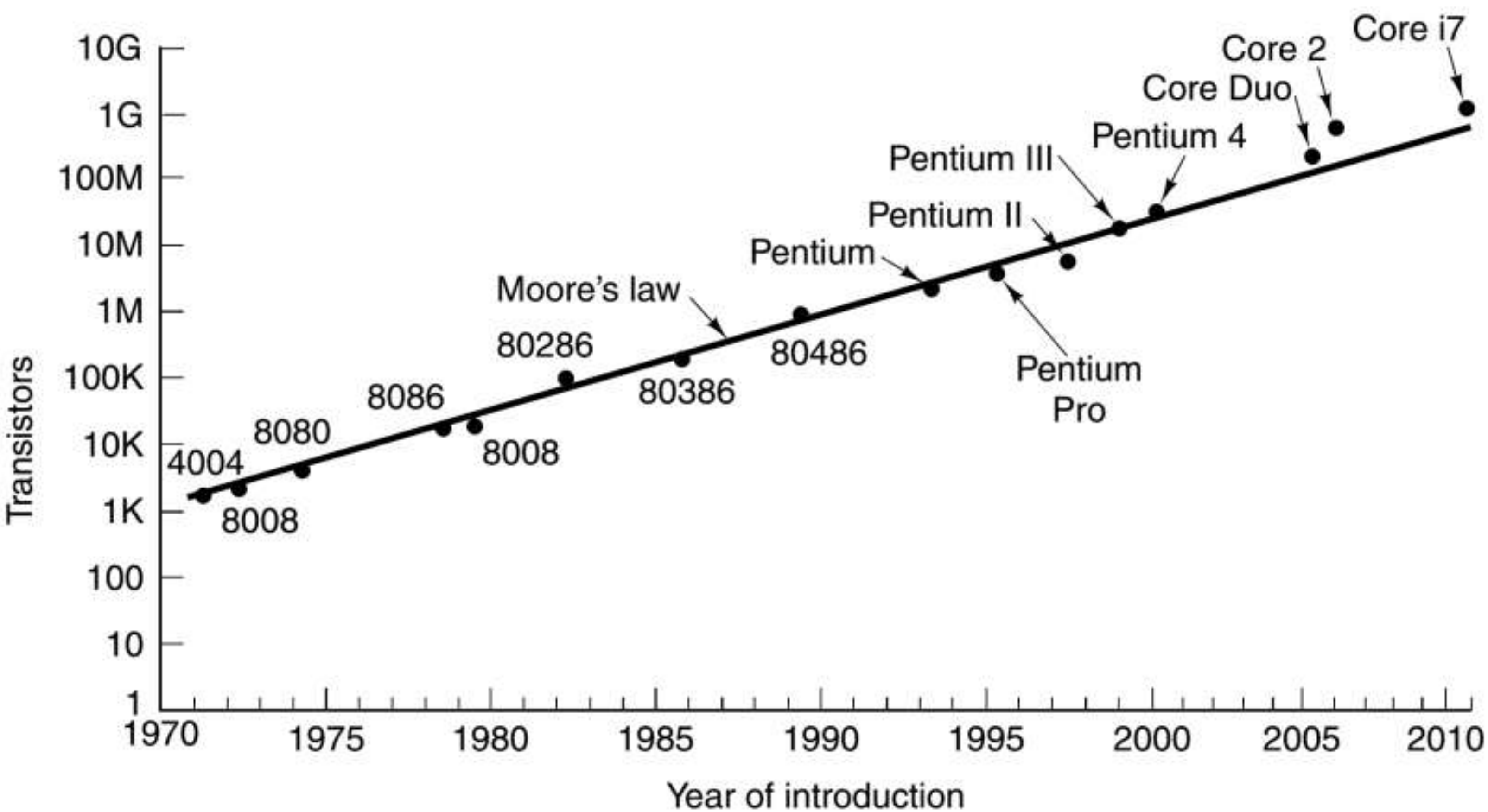
- Moore's Law for Memory Chips

A 60 percent annual increase in the number of transistors that can be put on a chip.



Technology (3)

- Moore's Law for CPU(Intel) Chips

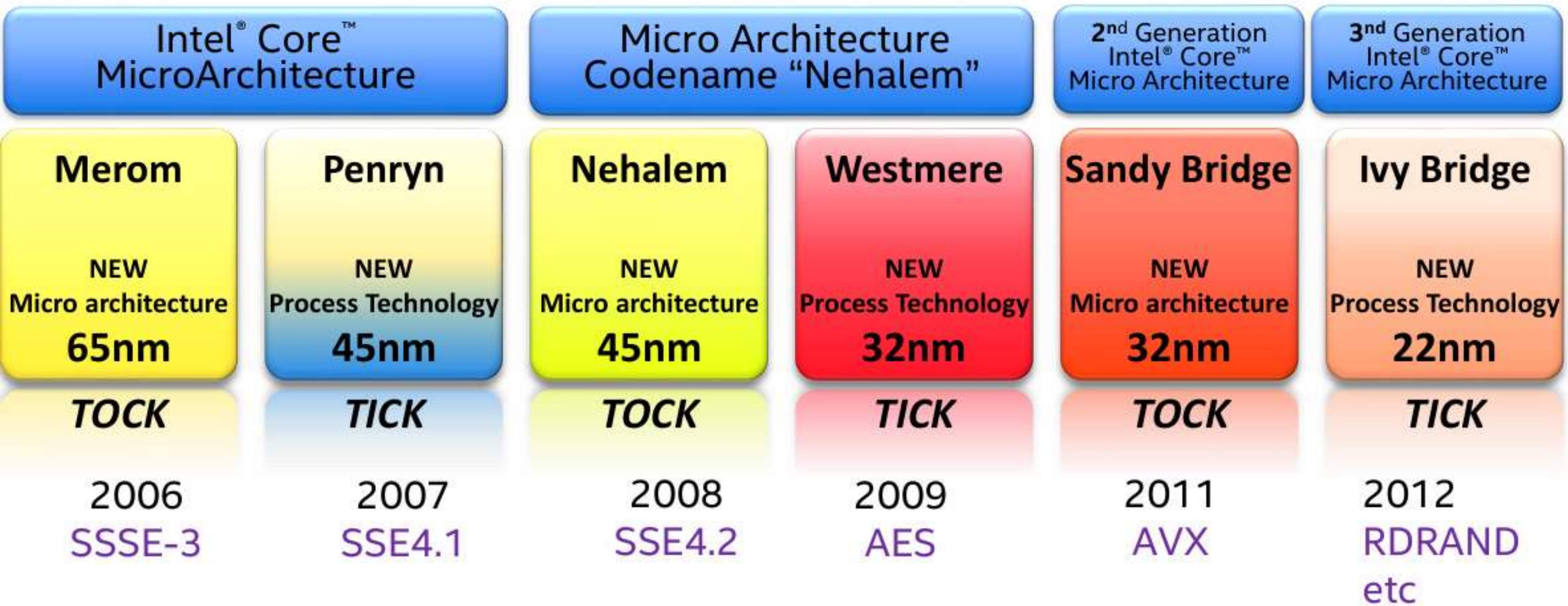


Technology (4)

- Tick-Tock Model (2007)
  - An aggressive development model introduced by Intel for their mainstream microprocessors.
  - Tick 芯片工艺提升，晶体管变小
    - With each tick, Intel advances their manufacturing process technology in line with Moore's Law.
  - Tock 工艺不变，芯片核心架构的升级
    - With each tock, Intel uses the their latest manufacturing process technology from their "tick" to manufacture a newly designed microarchitecture.

Technology (5)

- Tick-Tock Roadmap



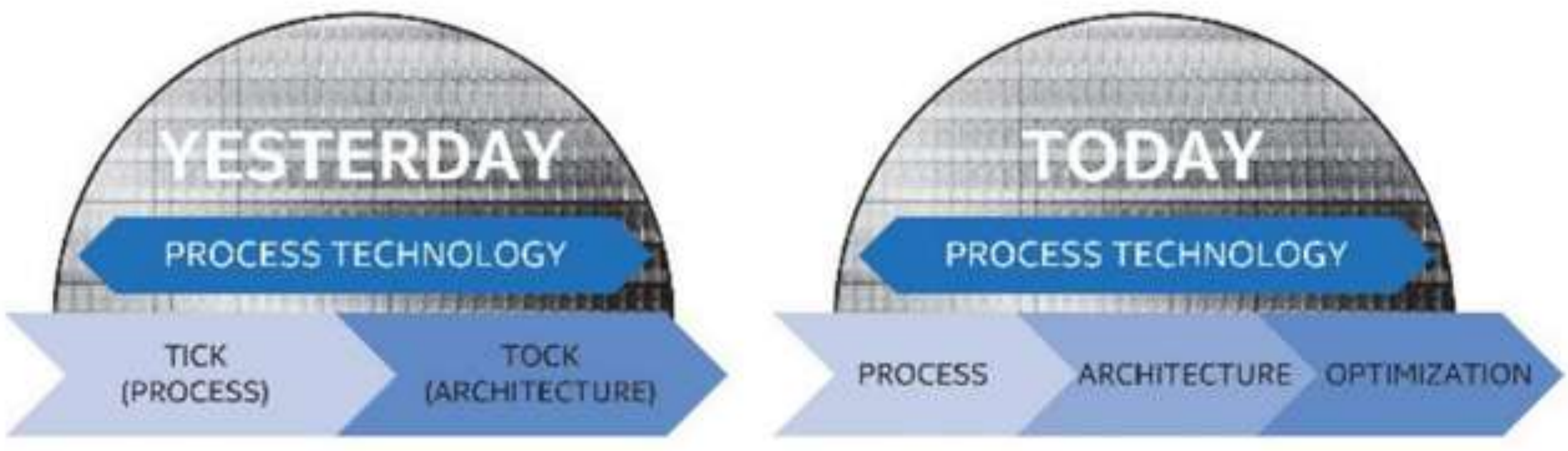
Technology (6)

- Tick-Tock Roadmap (ctd.)

4nth Generation Intel® Core™ Micro Architecture	TBD	TBD	TBD	TBD	TBD
Haswell	**Broadwell**	**Skylake**	**TBD**	**TBD**	**TBD**
NEW Micro architecture	**NEW** Process Technology	**NEW** Micro architecture	**NEW** Process Technology	**NEW** Micro architecture	**NEW** Process Technology
22nm	**14nm**	**14nm**	**10nm**	**10nm**	**7nm**
TICK	***TOCK***	***TICK***	***TOCK***	***TICK***	***TOCK***
2013	2015 !	2017	???	???	???
	5 new Inst.				
AVX-2					

Technology (6)

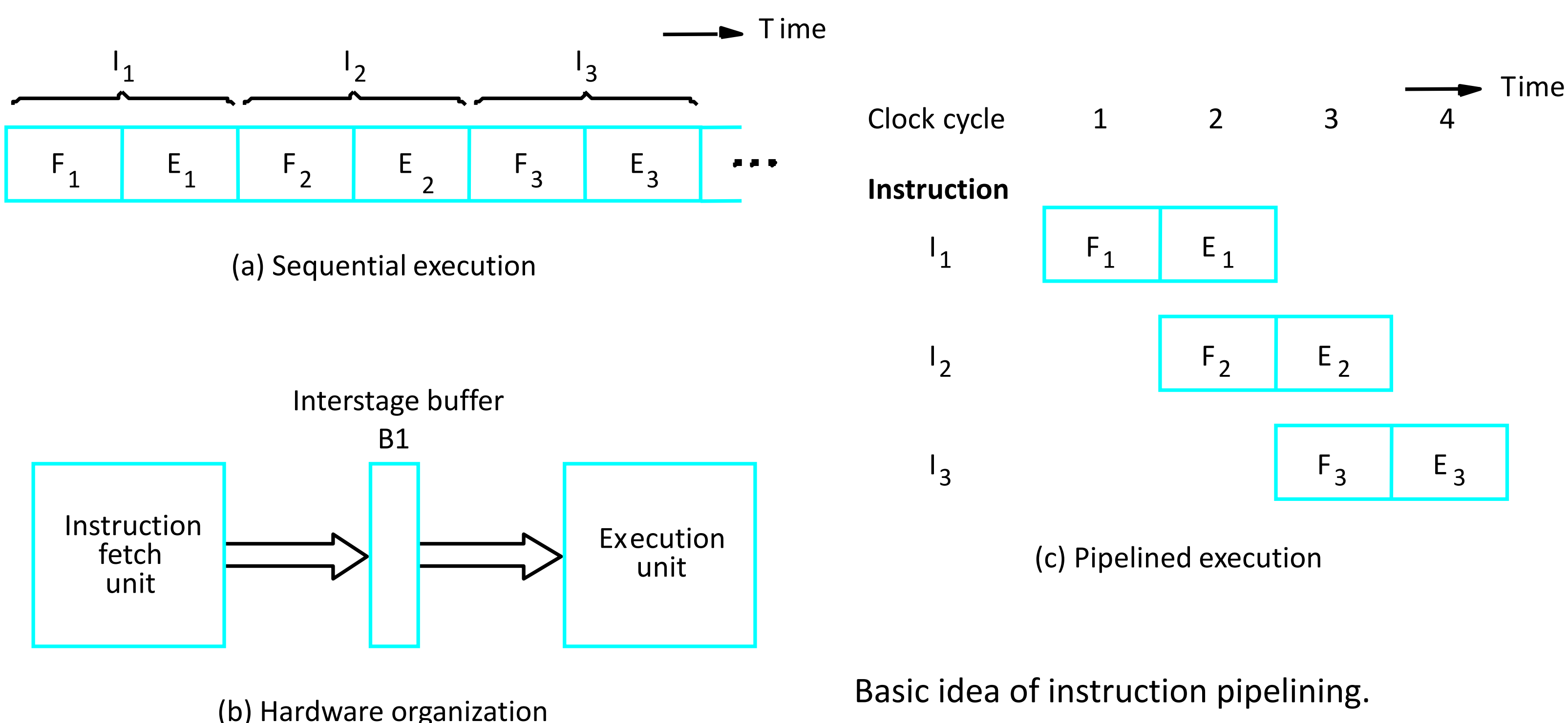
- Process-Architecture-Optimization
  - 10nm工艺被认为硅基半导体的转折点，再往后是7nm工艺，已经进入量子学范畴，半导体特性都有变化了，所以研发难度更大，需要更多黑科技材料加入，也需要EUV等新一代光刻设备。



Parallelism (1)

- Instruction-level Parallelism
  - Pipelining: beginning other waiting instructions before the first finishes

Fetch + Execution



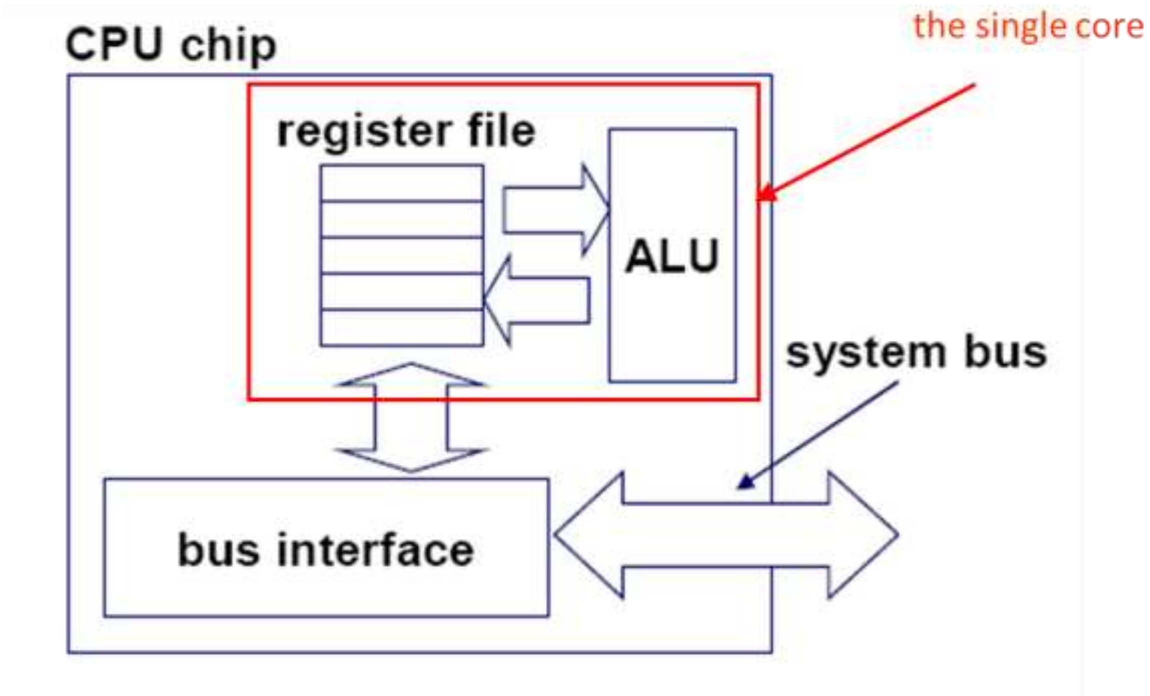
Basic idea of instruction pipelining.

Parallelism (2)

- Multicore Processors
  - Problems in Single Core
    - To execute the tasks faster you must increase the clock time.
    - Increasing clock times too high drastically increases power consumption(higher voltage) and heat dissipation to extremely high levels, making the processor inefficient.
    - Power consumed and heat dissipated is proportional to the square of the voltage.
    - E.g. At 3.6 GHz, the Pentium 4 consumes 115 watts of power. That means it gets about as hot as a 100-watt light bulb.

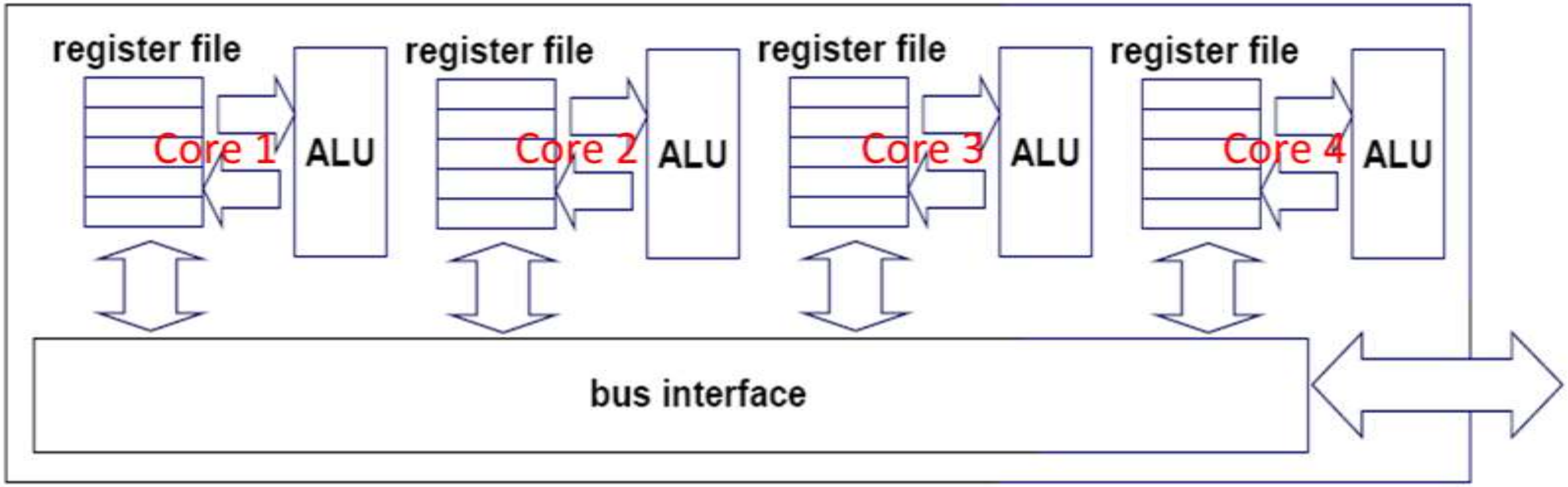
Parallelism (3)

- Multicore Processors (ctd.)
  - Single-core CPU Chip



Parallelism (4)

- Multicore Processors (ctd.)
  - Multi-core CPU Chip

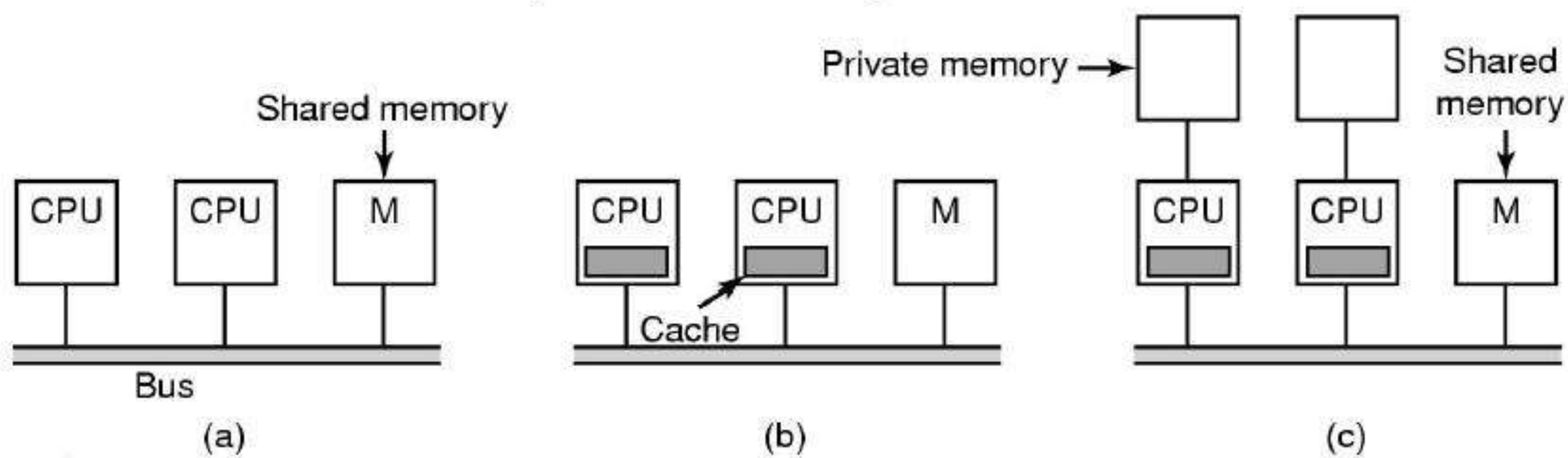


Parallelism (5)

- Multiprocessors
  - A shared-memory multiprocessor (or just multiprocessor henceforth) is a computer system in which two or more CPUs share full access to a common RAM.
  - Why Multiprocessors?
    - For improved latency: e.g. Meteorologist running weather analysis program
    - For improved throughput: e.g. Transaction processing systems at banks, Google search engine
    - For improved reliability: e.g. Computer system in a spacecraft; it should continue working correctly even when some CPUs go wrong

Parallelism (6)

- Shared-Memory Multiprocessor
  - (a) Simplest MP: More than one processor on a single bus connect to memory, bus bandwidth becomes a bottleneck.
  - (b) Each processor has a cache to reduce the need to access to memory.
  - (c) To further scale the number of processors, each processor is given private local memory.



Content of this lecture

- 1.1 Computer Types
- 1.2 Functional Units
- 1.3 Basic Operational Concepts
- 1.6 Performance
- 1.7 Historical Perspective
- Summary

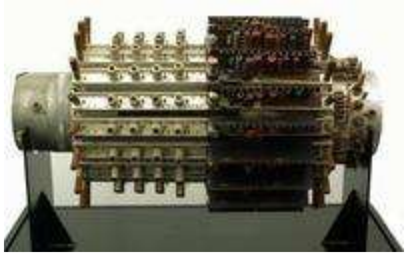
Computer Generations (1)

- The First Generation : Vacuum Tubes (1945-1955)
  - ENIAC(Electronic Numerical Integrator And Computer)
    - The world's first general-purpose electronic digital computer
    - The Moore school of the University of Pennsylvania

Consisted of 18,000 vacuum tubes and 1500 relays

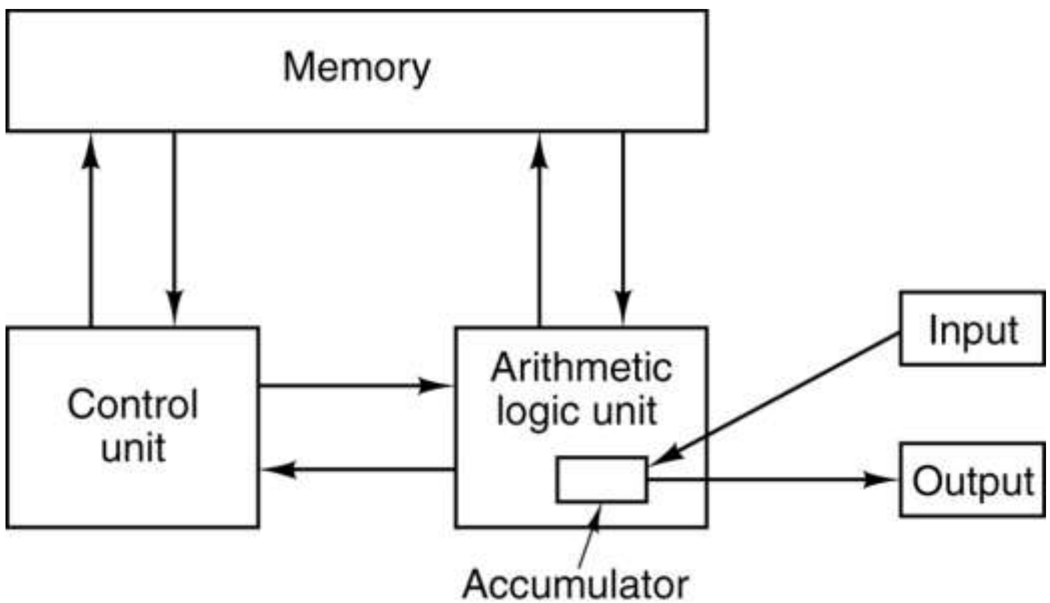
Computer Generations (2)

- The First Generation : Vacuum Tubes (1945-1955) (ctd.)
  - Stored-program concept (John von Neumann)
    - Data and instructions are stored in a single read-write memory.
    - The contents of this memory are addressable by location , without regard to the type of data contained there.
    - Execution occurs in a sequential fashion (unless explicitly modified) from one instruction to the next.
  - Assembly Language
  - Mercury delay-line memory



Computer Generations (3)

- The First Generation : Vacuum Tubes (1945-1955) (ctd.)
  - The Original Von Neumann Machine



Computer Generations (4)

- The Second Generation : Transistors (1955-1965)
  - Magnetic core memories and magnetic drum storage devices
  - High-level languages, e.g. Fortran
  - Compiler
  - Separate I/O processors
  - PDP-8, IBM7094, CDC6600…

Computer Generations (5)

- The Third Generation : Integrated Circuits (1965-1975)
  - An IC comprises a number of circuit components. They are interconnected in a single small package to perform the desired electronic function.
  - Small Scale Integration: transistors <100
  - Medium Scale Integration: transistors101~1k
  - Integrated-circuit memories
  - Microprogramming
  - Parallelism
  - Pipelining
  - Cache
  - Virtual memories
  - IBM360 Computer family, PDP-11, …

Computer Generations (6)

- The Fourth Generation : LSI & VLSI (1975-1990)
  - Complete processors and large sections of the main memory of small computers could be implemented on single chips.
  - LSI: Large Scale Integration
    - Transistors 1,001~10k
  - VLSI: Very Large Scale Integration
    - Transistors 10,001~100k
  - RISC takes place of CISC
  - FPGA
  - Multicore and Parallel Programming

Computer Generations (7)

- Beyond the Generation (1991-   )
  - Some computer system that have a dominant organizational or application-driven feature.
  - ULSI: Ultra Large Scale Integration
    - Transistors 100,001~10M
  - GLSI: Giga Scale Integration
    - Transistors>10,000,001

Computer Generations (8)

- Beyond the Generation (1991-   ) (ctd.)
  - 三大主流厂商晶体管密度比较
    - 单位：百万个晶体管每平方毫米

排名	工艺技术	近似晶体管密度（MTr/mm²）
1	英特尔 10nm	100.8
2	台积电 7nm FF/FF+	96.5
3	三星 7nm	95.3
4	台积电 7nm HPC	66.7
5	三星 8nm	61.2
6	台积电 10nm	60.3
7	三星 10nm	51.8
8	英特尔 14nm	43.5
9	GlobalFoundries 12nm	36.7
10	台积电 12nm	33.8
11	三星/GlobalFoundries 14nm	32.5
12	台积电 16nm	28.2

Summary (1)

- Computer Types
- Five basic functional units of computers
  - Arithmetic and logic unit
  - Control unit
  - Memory
  - Input unit
  - Output unit

Summary (2)

- Computer Generations
  - The first generation: vacuum tubes
  - The second generation: transistors
  - The third generation: integrated circuits
  - The fourth generation: LSI&VLSI