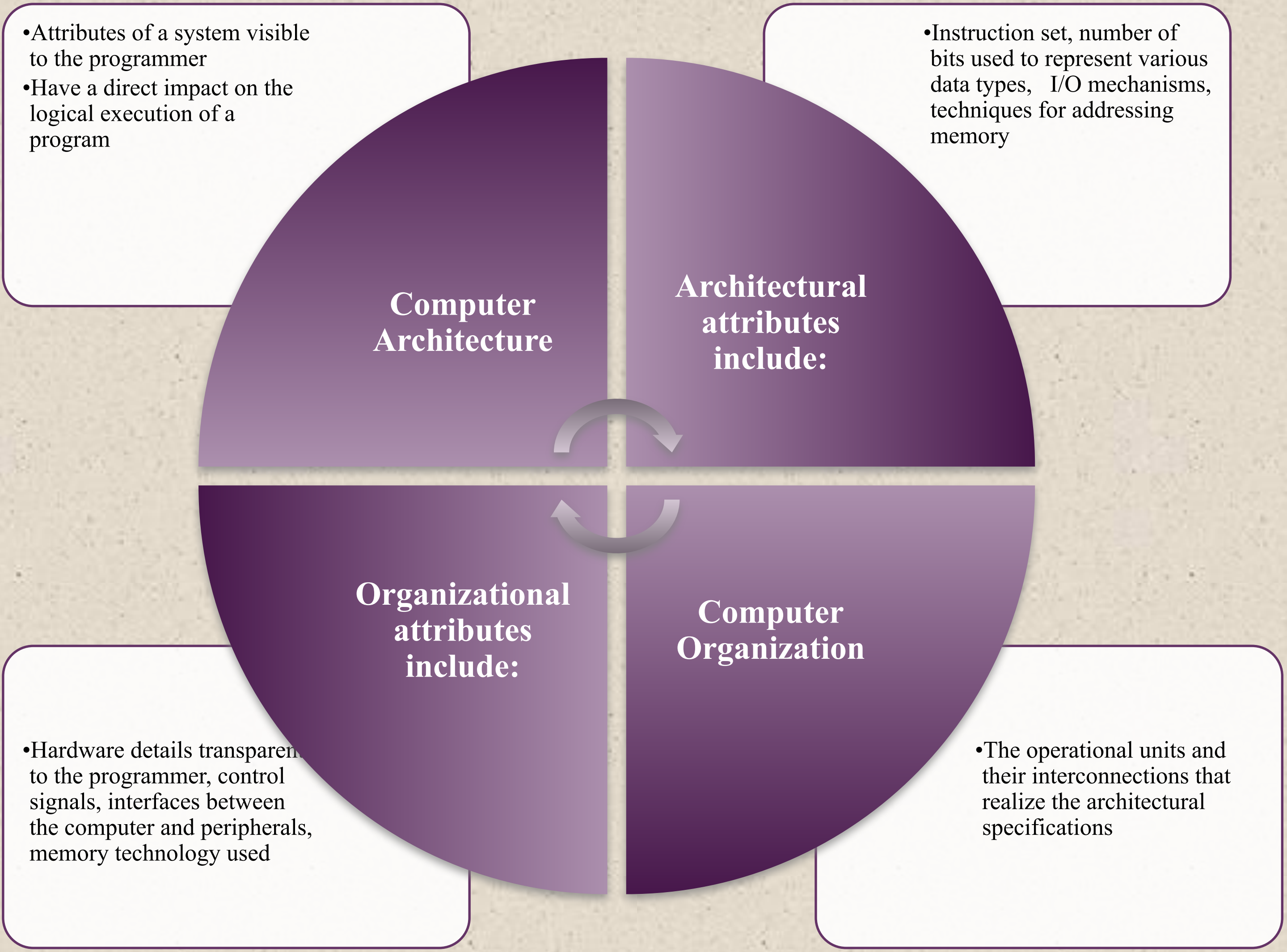
Computer Architecture

Computer Organization



Computer Architecture

- Attributes of a system visible to the programmer
- Have a direct impact on the logical execution of a program

Architectural attributes include:

- Instruction set, number of bits used to represent various data types, I/O mechanisms, techniques for addressing memory

Computer Organization

- The operational units and their interconnections that realize the architectural specifications

Organizational attributes include:

- Hardware details transparent to the programmer, control signals, interfaces between the computer and peripherals, memory technology used

Structure and Function

- Hierarchical system
  - Set of interrelated subsystems
- Hierarchical nature of complex systems is essential to both their design and their description
- Designer need only deal with a particular level of the system at a time
  - Concerned with structure and function at each level

- Structure
  - The way in which components relate to each other
- Function
  - The operation of individual components as part of the structure

+

Function

A computer can perform four basic functions:

- Data processing
- Data storage
- Data movement
- Control

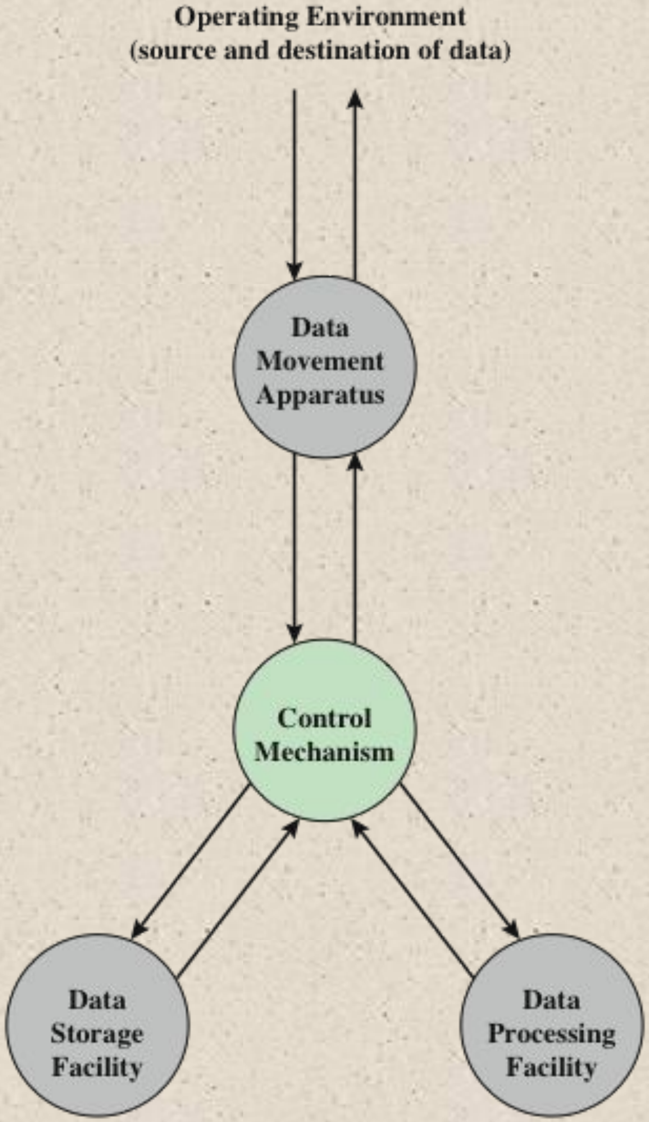


Figure 1.1 A Functional View of the Computer

+

Operations

(a)
Data movement

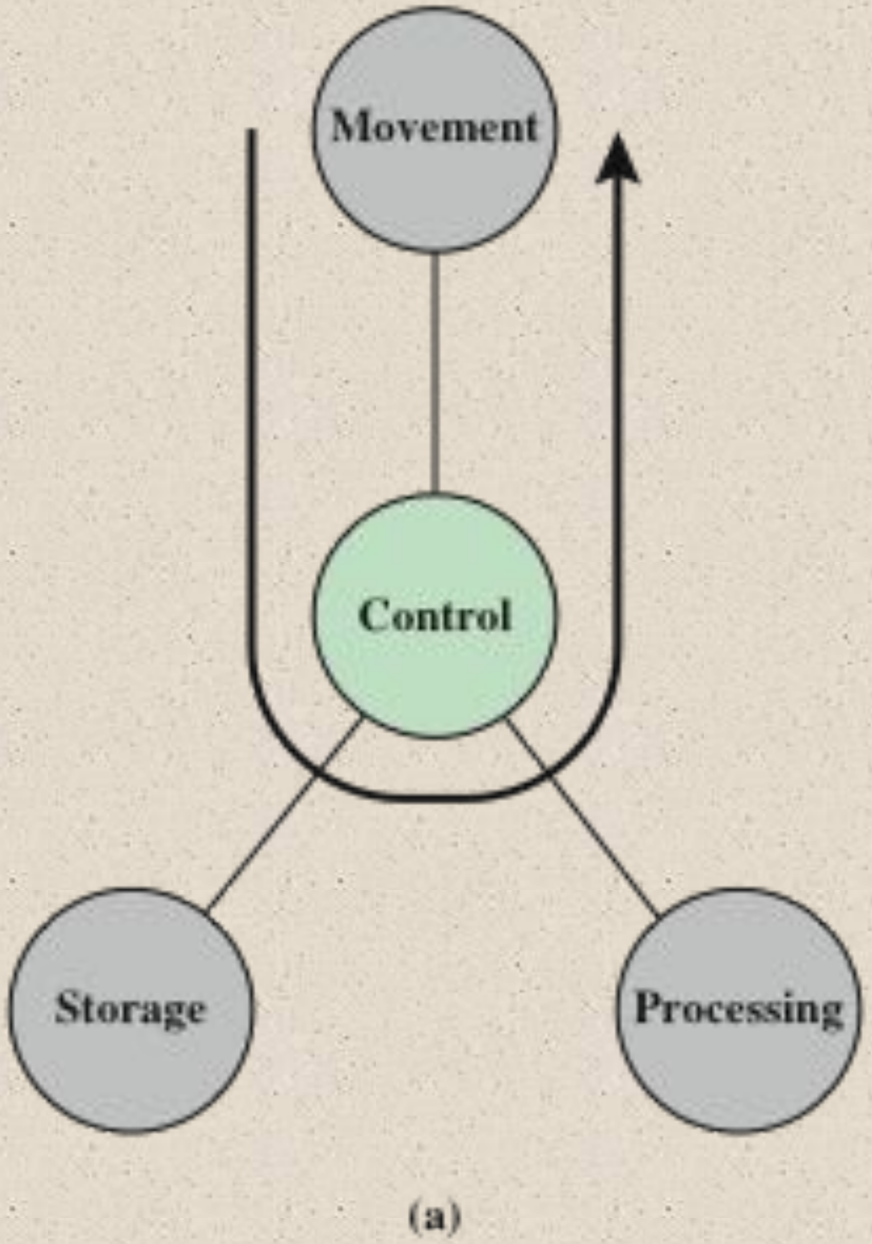
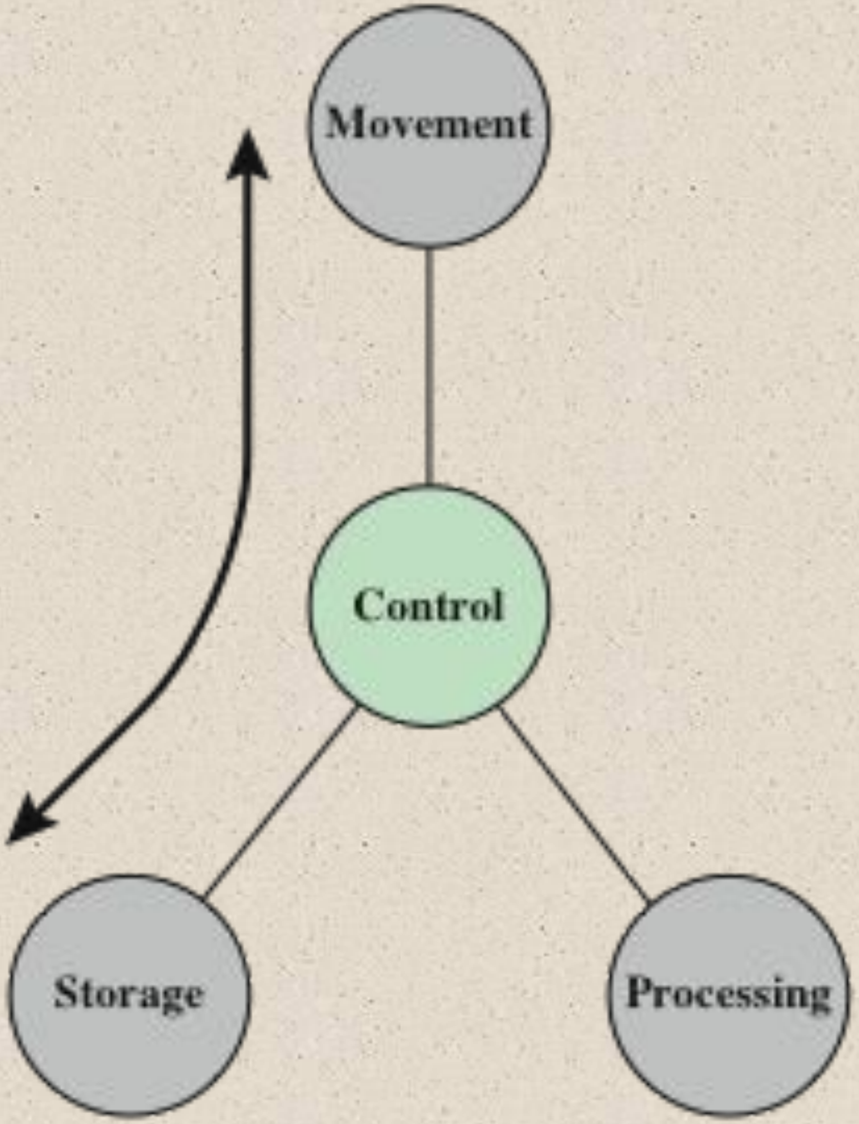


Figure 1.2 Possible Computer Operations

Operations

(b)
Data storage



(b)

Figure 1.2 Possible Computer Operations

+

Operations

(c)
Data processing

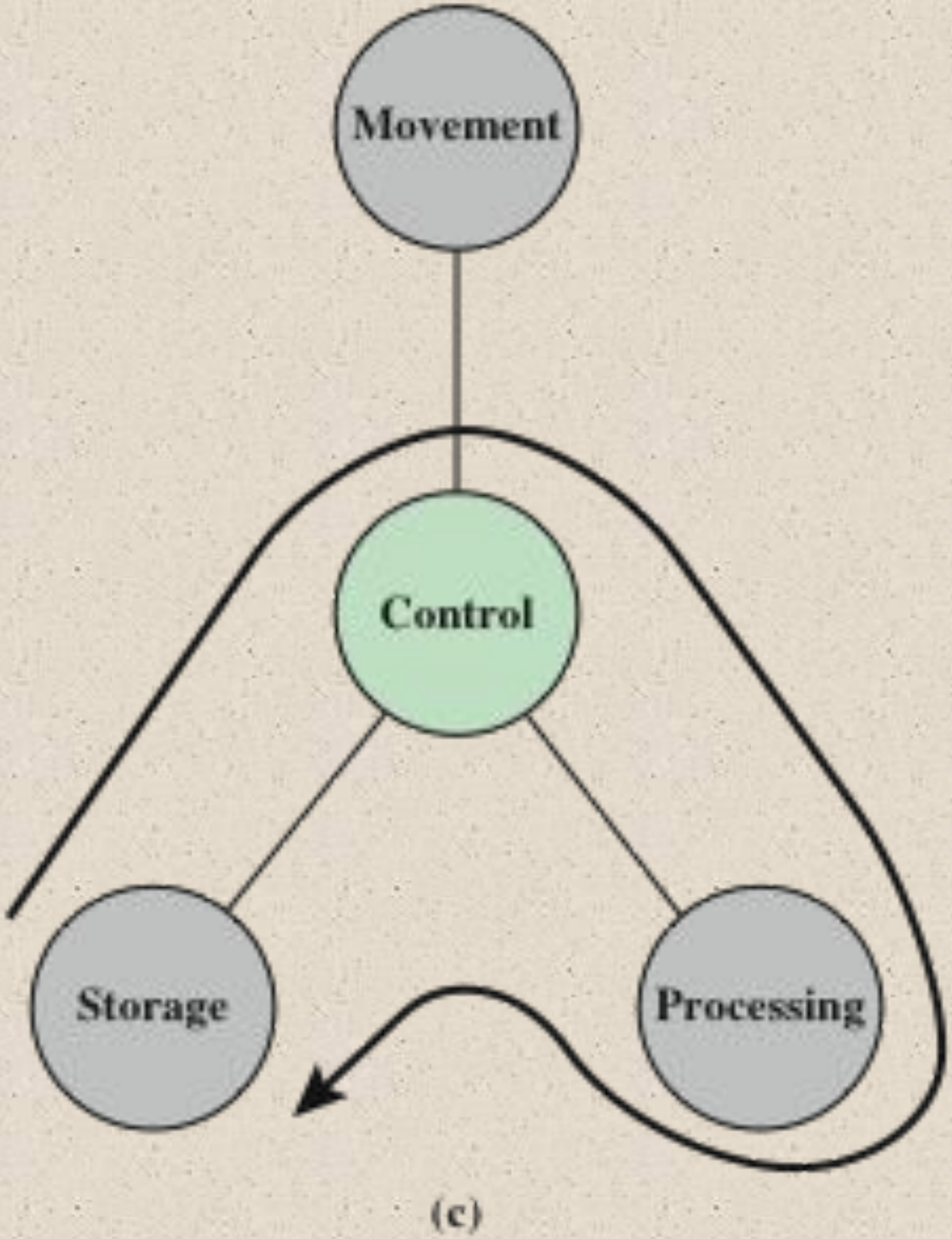


Figure 1.2 Possible Computer Operations

+

Operations

(d)
Control

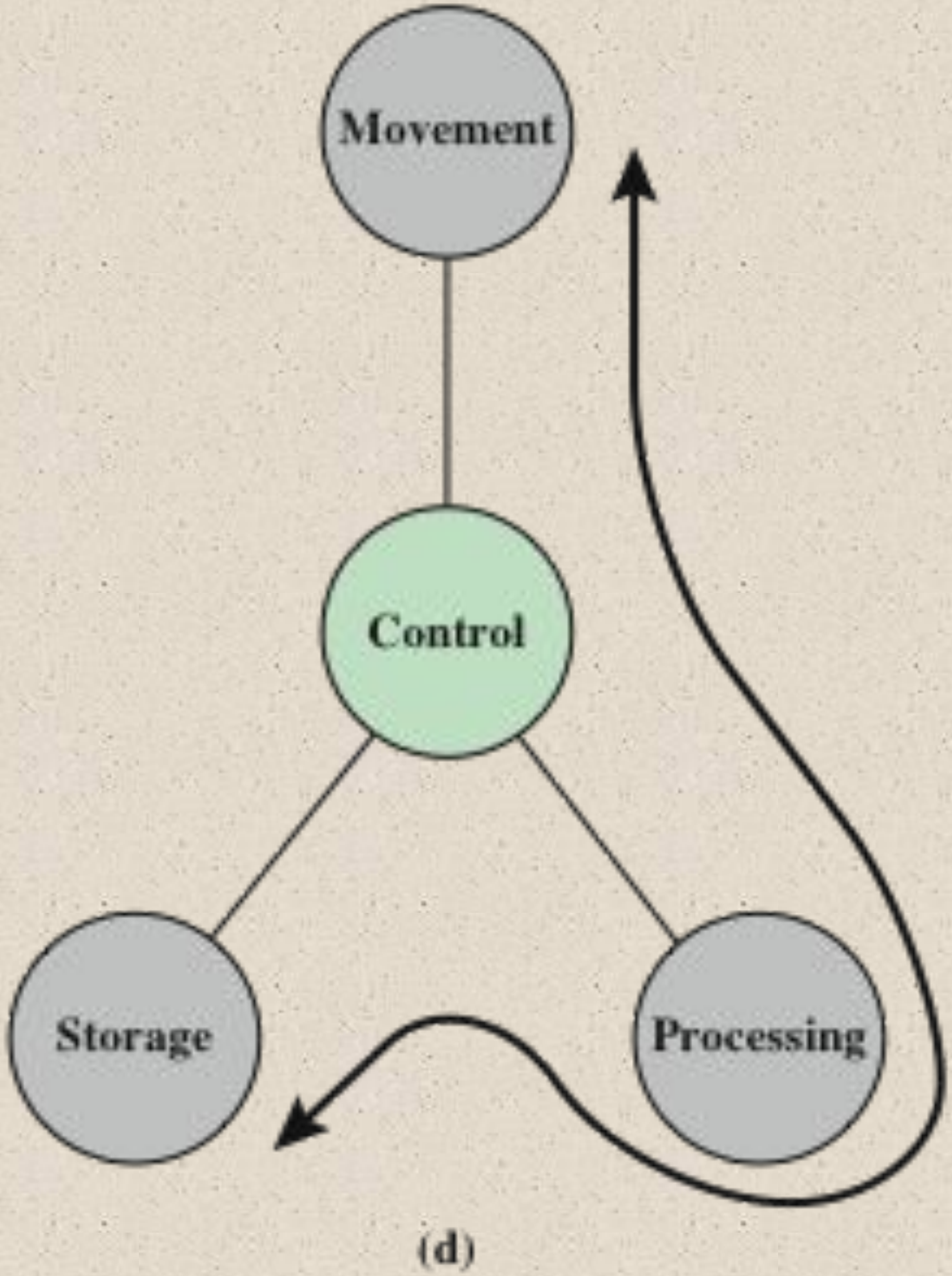


Figure 1.2 Possible Computer Operations

The Computer

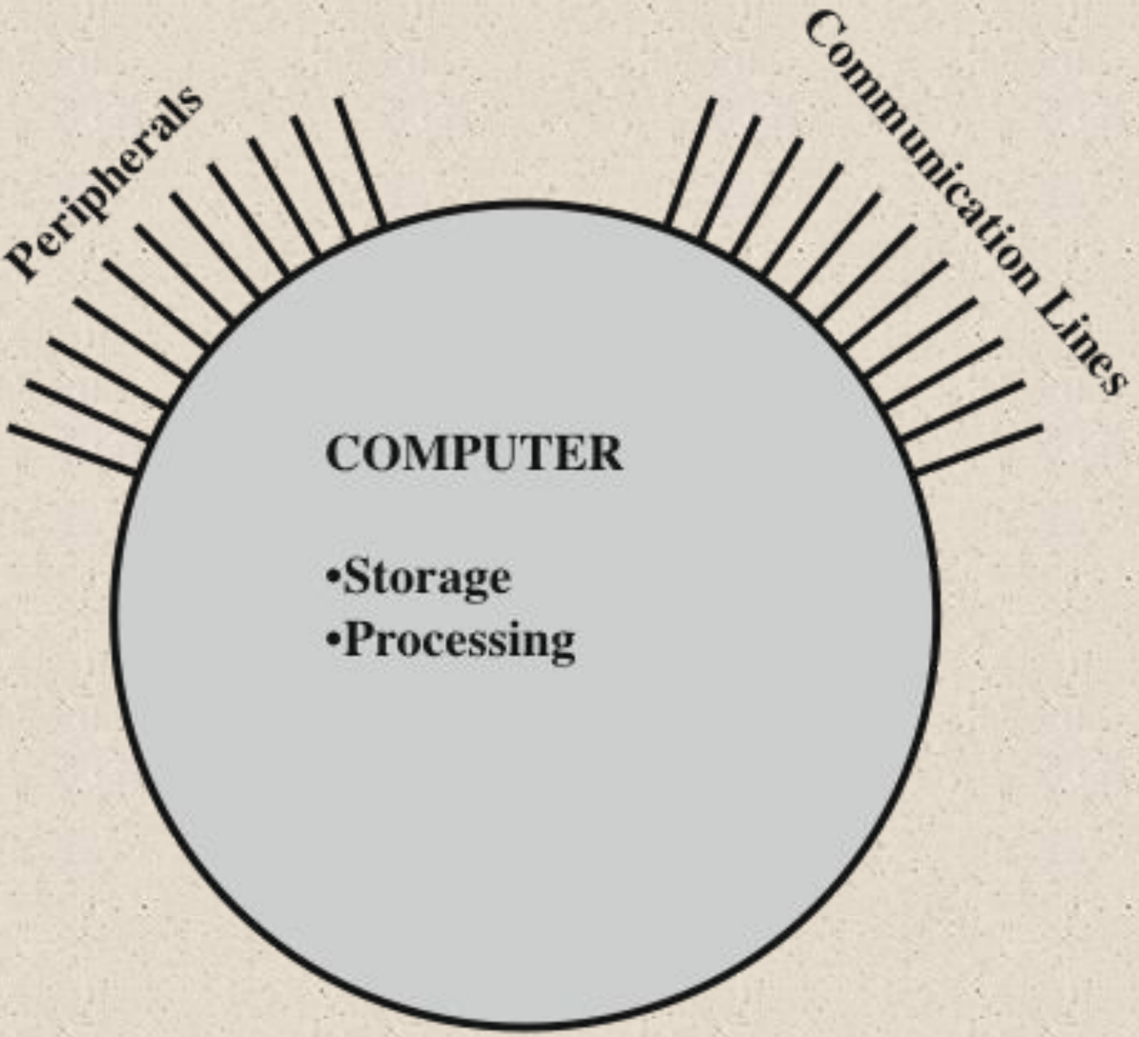
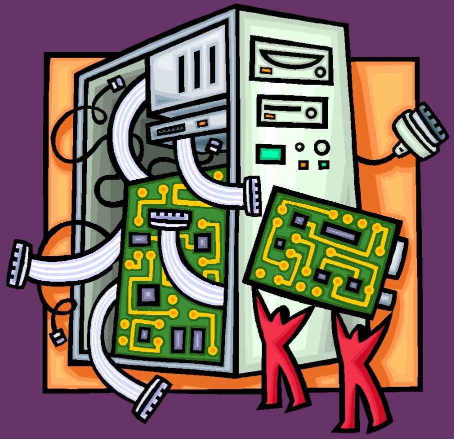


Figure 1.3 The Computer

+

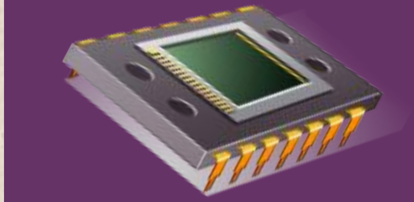
There are four main structural components of the computer:

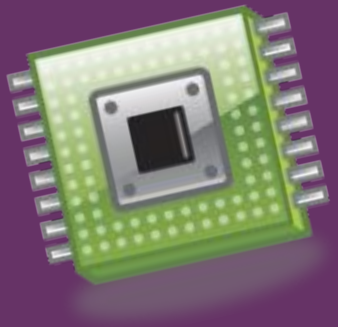


- CPU – controls the operation of the computer and performs its data processing functions
- Main Memory – stores data
- I/O – moves data between the computer and its external environment
- System Interconnection – some mechanism that provides for communication among CPU, main memory, and I/O

CPU

Major structural components:

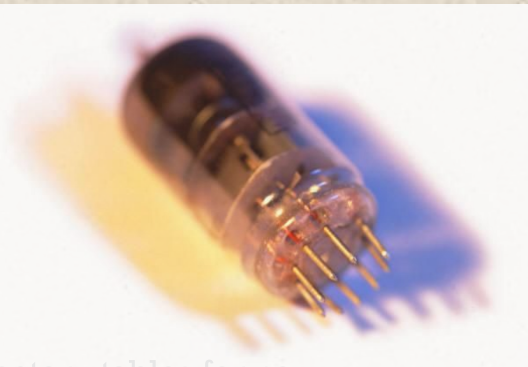




- Control Unit
  - Controls the operation of the CPU and hence the computer
- Arithmetic and Logic Unit (ALU)
  - Performs the computer’s data processing function
- Registers
  - Provide storage internal to the CPU
- CPU Interconnection
  - Some mechanism that provides for communication among the control unit, ALU, and registers

History of Computers

First Generation: Vacuum Tubes



- ENIAC
  - Electronic Numerical Integrator And Computer
- Designed and constructed at the University of Pennsylvania
  - Started in 1943 – completed in 1946
  - By John Mauchly and John Eckert
- World's first general purpose electronic digital computer
  - Army's Ballistics Research Laboratory (BRL) needed a way to supply trajectory tables for new weapons accurately and within a reasonable time frame
  - Was not finished in time to be used in the war effort
- Its first task was to perform a series of calculations that were used to help determine the feasibility of the hydrogen bomb
- Continued to operate under BRL management until 1955 when it was disassembled

ENIAC

Weighed 30 tons

Occupied 1500 square feet of floor space

Contained more than 18,000 vacuum tubes

140 kW Power consumption

Capable of 5000 additions per second

Decimal rather than binary machine

Memory consisted of 20 accumulators, each capable of holding a 10 digit number

Major drawback was the need for manual programming by setting switches and plugging/ unplugging cables

+

John von Neumann

EDVAC (Electronic Discrete Variable Computer)

- First publication of the idea was in 1945

- Stored program concept
  - Attributed to ENIAC designers, most notably the mathematician John von Neumann
  - Program represented in a form suitable for storing in memory alongside the data

- IAS computer
  - Princeton Institute for Advanced Studies
  - Prototype of all subsequent general-purpose computers
  - Completed in 1952

Structure of von Neumann Machine

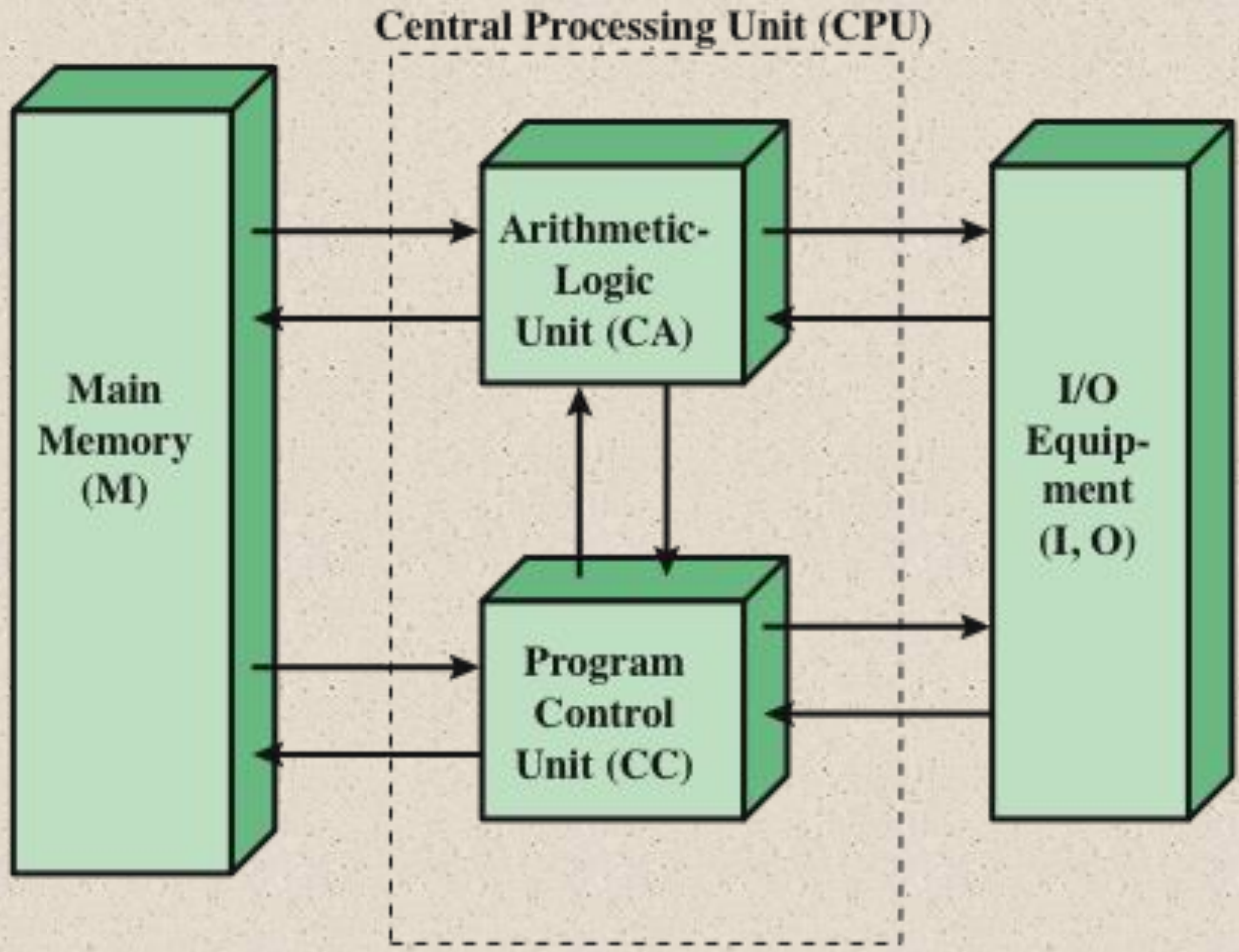

Central Processing Unit (CPU)
Main Memory (M)
Arithmetic-Logic Unit (CA)
Program Control Unit (CC)
I/O Equip-ment (I, O)

Figure 2.1 Structure of the IAS Computer

IAS Memory Formats

- The memory of the IAS consists of 1000 storage locations (called ***words***) of 40 bits each

- Both data and instructions are stored there

- Numbers are represented in binary form and each instruction is a binary code

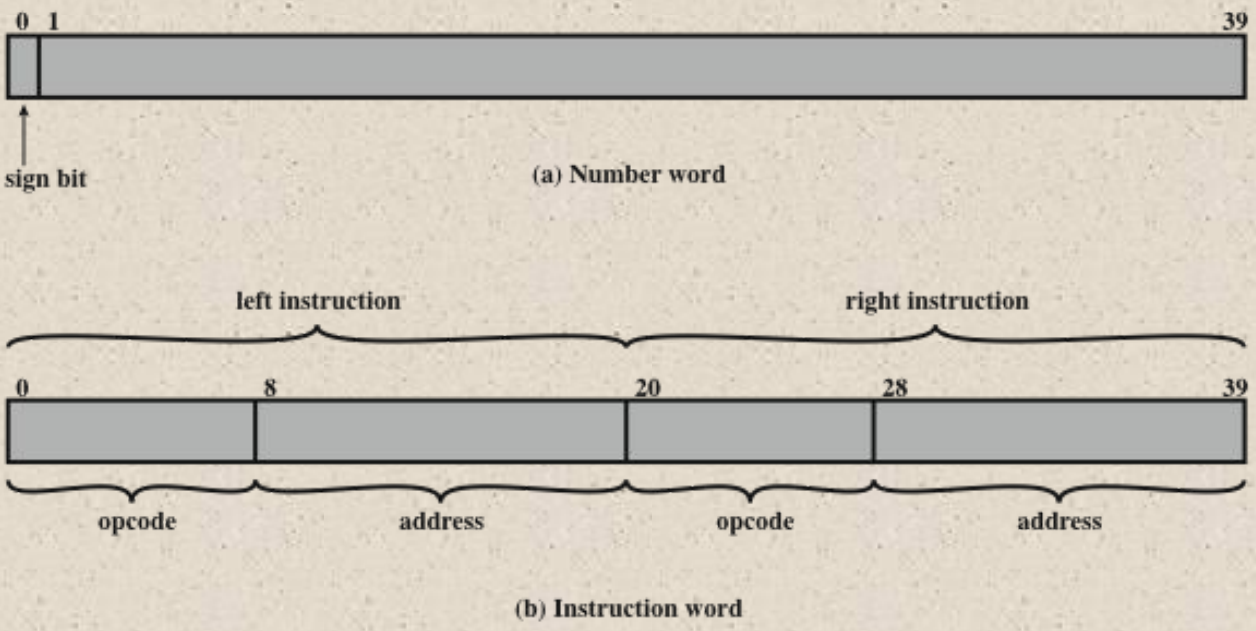


Figure 2.2 IAS Memory Formats

+

Structure of IAS Computer

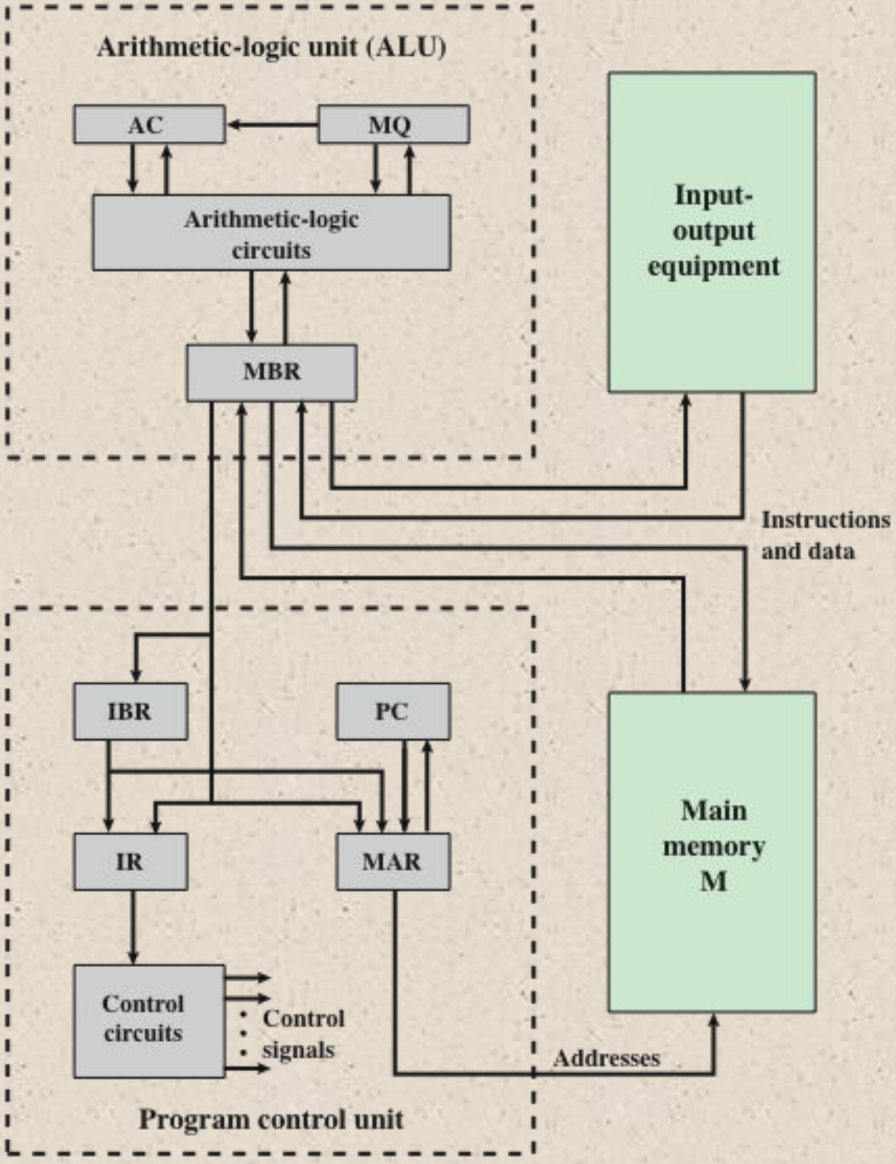


Figure 2.3 Expanded Structure of IAS Computer

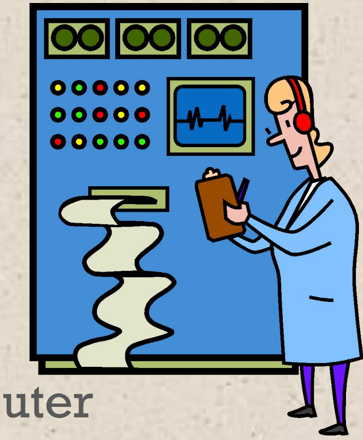
Registers

Memory buffer register (MBR)	• Contains a word to be stored in memory or sent to the I/O unit • Or is used to receive a word from memory or from the I/O unit
Memory address register (MAR)	• Specifies the address in memory of the word to be written from or read into the MBR
Instruction register (IR)	• Contains the 8-bit opcode instruction being executed
Instruction buffer register (IBR)	• Employed to temporarily hold the right-hand instruction from a word in memory
Program counter (PC)	• Contains the address of the next instruction pair to be fetched from memory
Accumulator (AC) and multiplier quotient (MQ)	• Employed to temporarily hold operands and results of ALU operations

+

Commercial Computers

UNIVAC



- 1947 – Eckert and Mauchly formed the Eckert-Mauchly Computer Corporation to manufacture computers commercially
- UNIVAC I (Universal Automatic Computer)
  - First successful commercial computer
  - Was intended for both scientific and commercial applications
  - Commissioned by the US Bureau of Census for 1950 calculations
- The Eckert-Mauchly Computer Corporation became part of the UNIVAC division of the Sperry-Rand Corporation
- UNIVAC II – delivered in the late 1950's
  - Had greater memory capacity and higher performance
- Backward compatible

History of Computers
Second Generation: Transistors

- Smaller
- Cheaper
- Dissipates less heat than a vacuum tube
- Is a *solid state device* made from silicon
- Was invented at Bell Labs in 1947
- It was not until the late 1950's that fully transistorized computers were commercially available

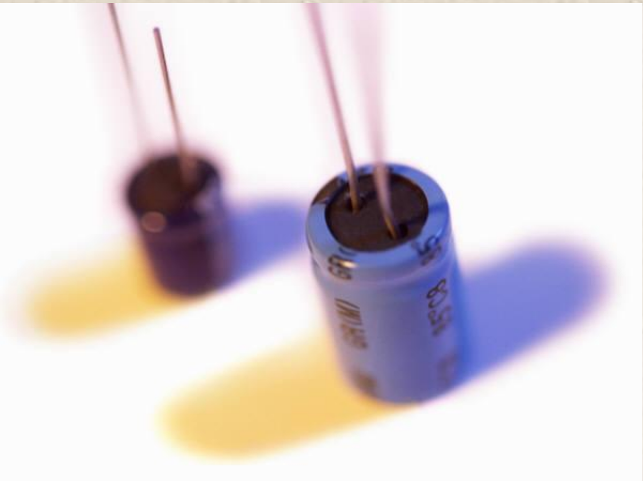
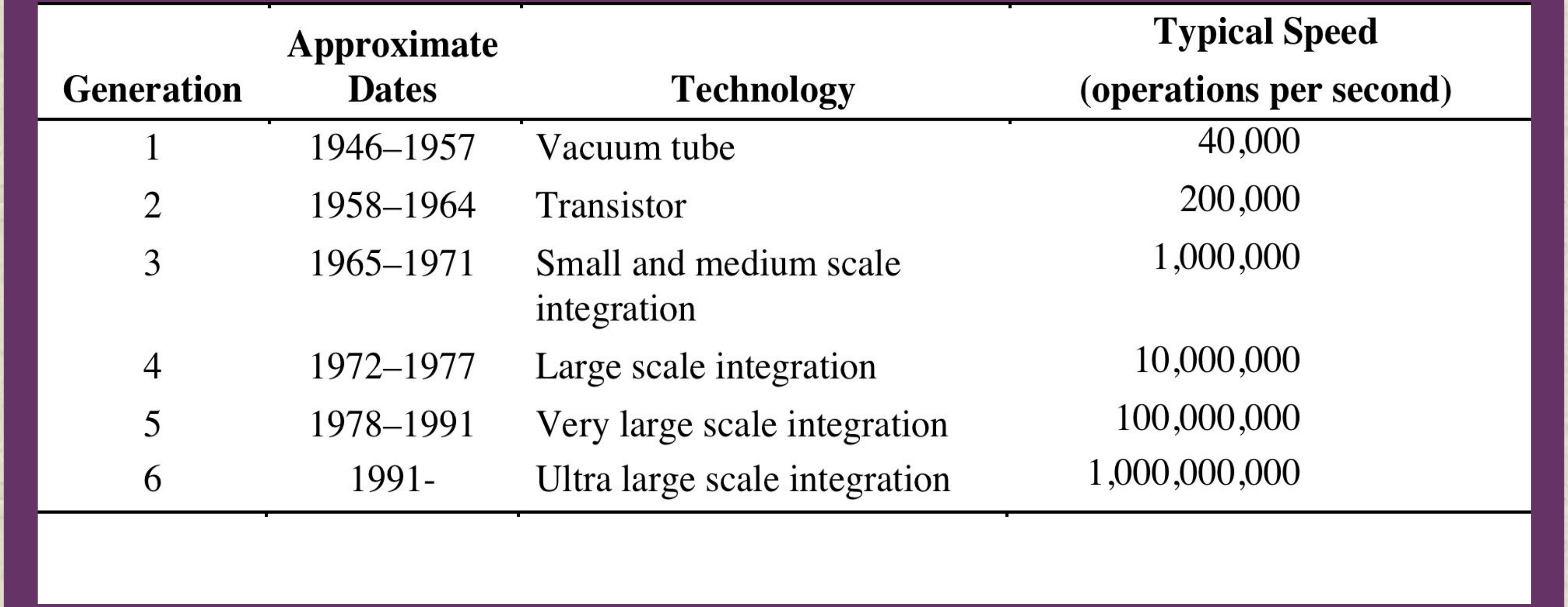


Table 2.2
Computer Generations



Generation	Approximate Dates	Technology	Typical Speed (operations per second)
1	1946–1957	Vacuum tube	40,000
2	1958–1964	Transistor	200,000
3	1965–1971	Small and medium scale integration	1,000,000
4	1972–1977	Large scale integration	10,000,000
5	1978–1991	Very large scale integration	100,000,000
6	1991-	Ultra large scale integration	1,000,000,000

Second Generation Computers

- Introduced:
  - More complex arithmetic and logic units and control units
  - The use of high-level programming languages
  - Provision of *system software* which provided the ability to:
    - load programs
    - move data to peripherals and libraries
    - perform common computations

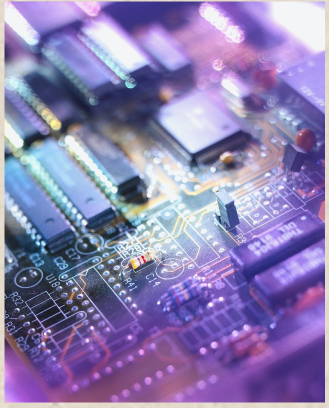
- Appearance of the Digital Equipment Corporation (DEC) in 1957

- PDP-1 was DEC's first computer

- This began the mini-computer phenomenon that would become so prominent in the third generation

History of Computers

Third Generation: Integrated Circuits



- 1958 – the invention of the integrated circuit
- *Discrete component*
  - Single, self-contained transistor
  - Manufactured separately, packaged in their own containers, and soldered or wired together onto masonite-like circuit boards
  - Manufacturing process was expensive and cumbersome
- The two most important members of the third generation were the IBM System/360 and the DEC PDP-8

+

Microelectronics

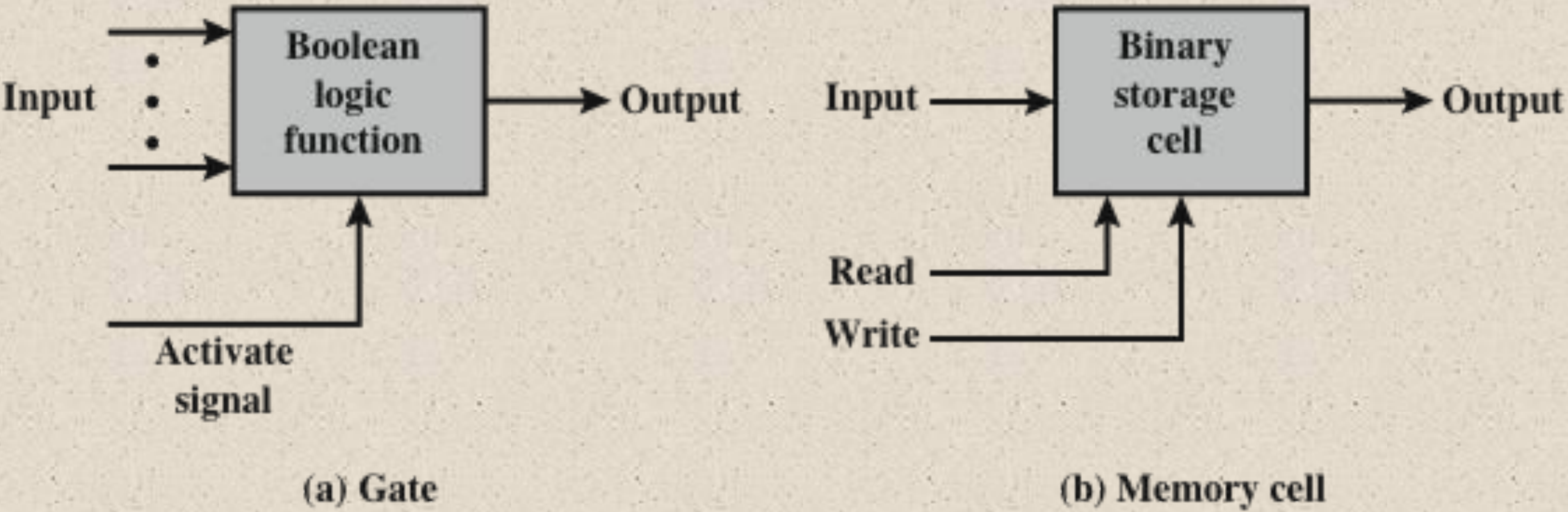


Figure 2.6 Fundamental Computer Elements

Integrated Circuits

- Data storage – provided by memory cells
- Data processing – provided by gates
- Data movement – the paths among components are used to move data from memory to memory and from memory through gates to memory
- Control – the paths among components can carry control signals

- A computer consists of gates, memory cells, and interconnections among these elements
- The gates and memory cells are constructed of simple digital electronic components
- Exploits the fact that such components as transistors, resistors, and conductors can be fabricated from a semiconductor such as silicon
- Many transistors can be produced at the same time on a single wafer of silicon
- Transistors can be connected with a processor metallization to form circuits

+

Wafer, Chip, and Gate Relationship

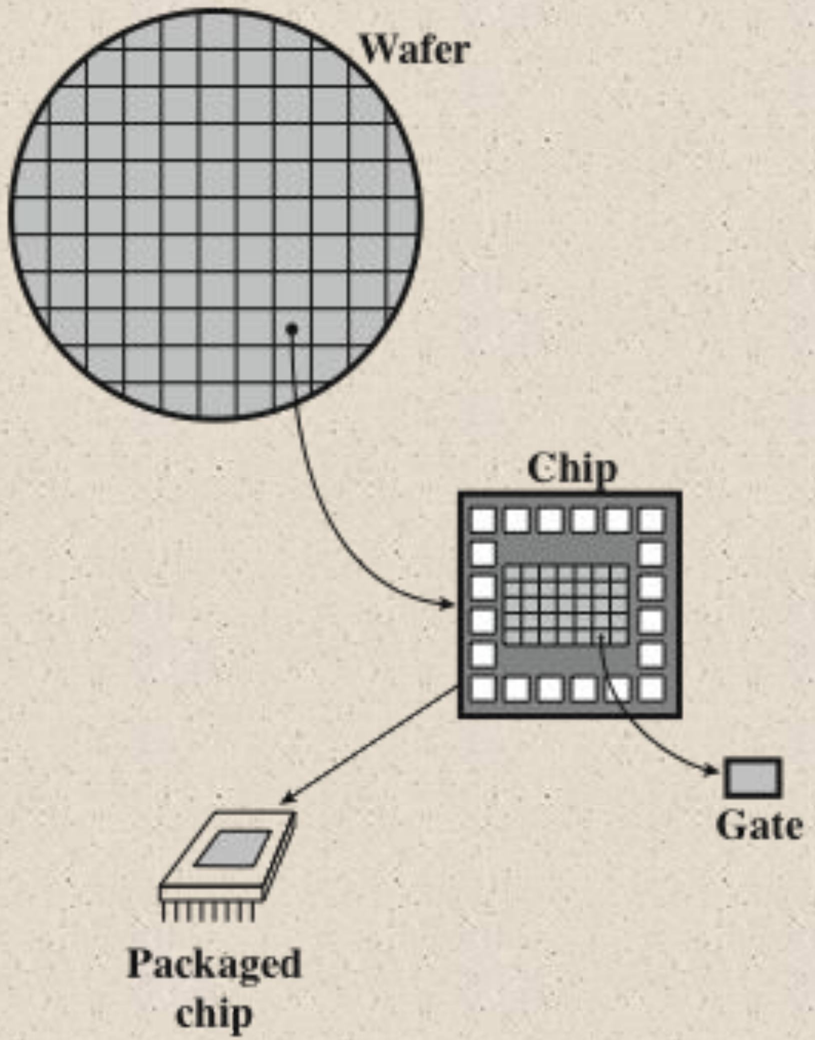


Figure 2.7 Relationship Among Wafer, Chip, and Gate

+

Chip Growth

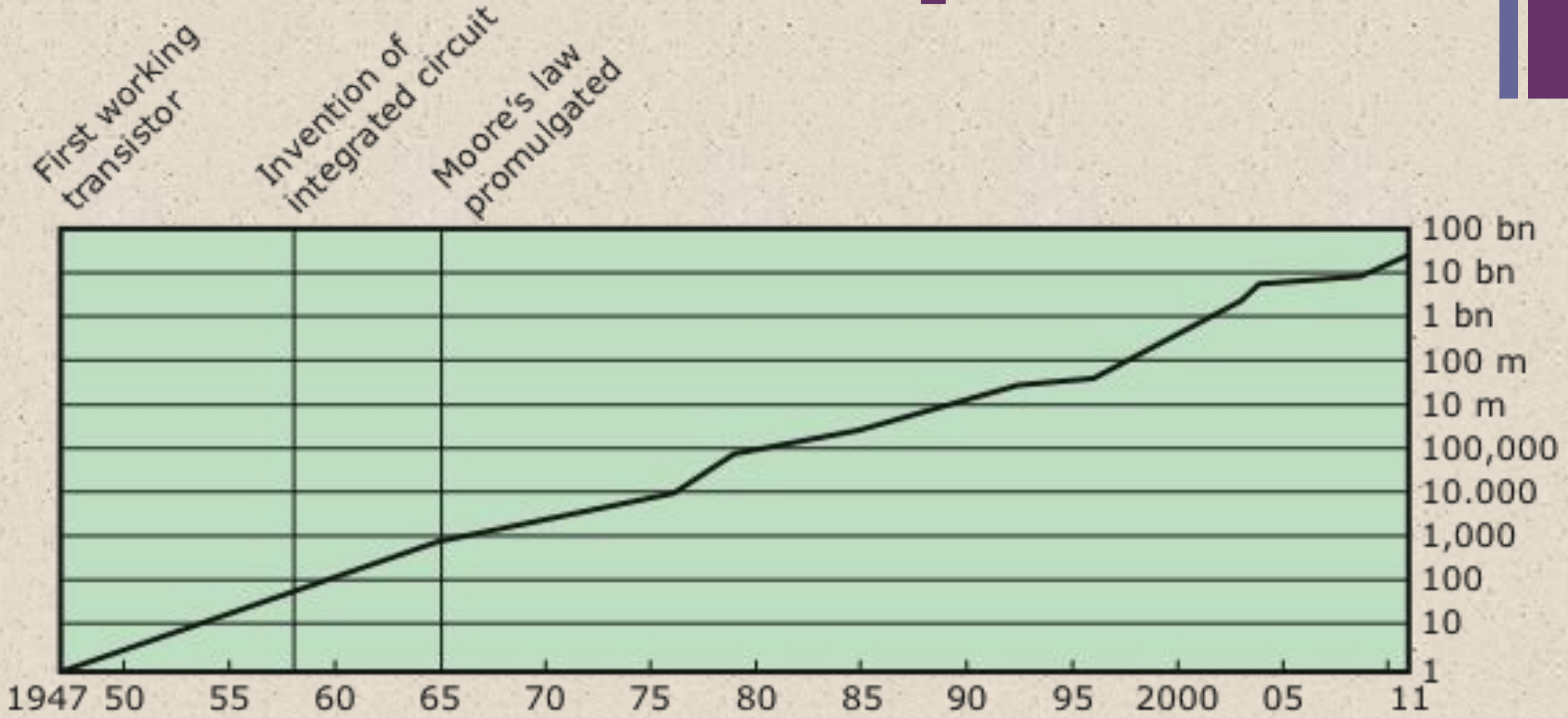


Figure 2.8 Growth in Transistor Count on Integrated Circuits (DRAM memory)

Moore's Law

1965; Gordon Moore – co-founder of Intel

Observed number of transistors that could be put on a single chip was doubling every year

The pace slowed to a doubling every 18 months in the 1970's but has sustained that rate ever since

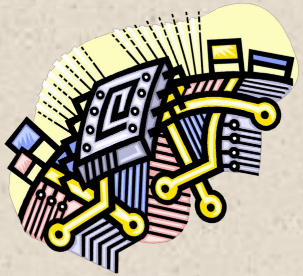
Consequences of Moore's law:

- The cost of computer logic and memory circuitry has fallen at a dramatic rate
- The electrical path length is shortened, increasing operating speed
- Computer becomes smaller and is more convenient to use in a variety of environments
- Reduction in power and cooling requirements
- Fewer interchip connections

Later Generations

LSI
Large
Scale
Integration

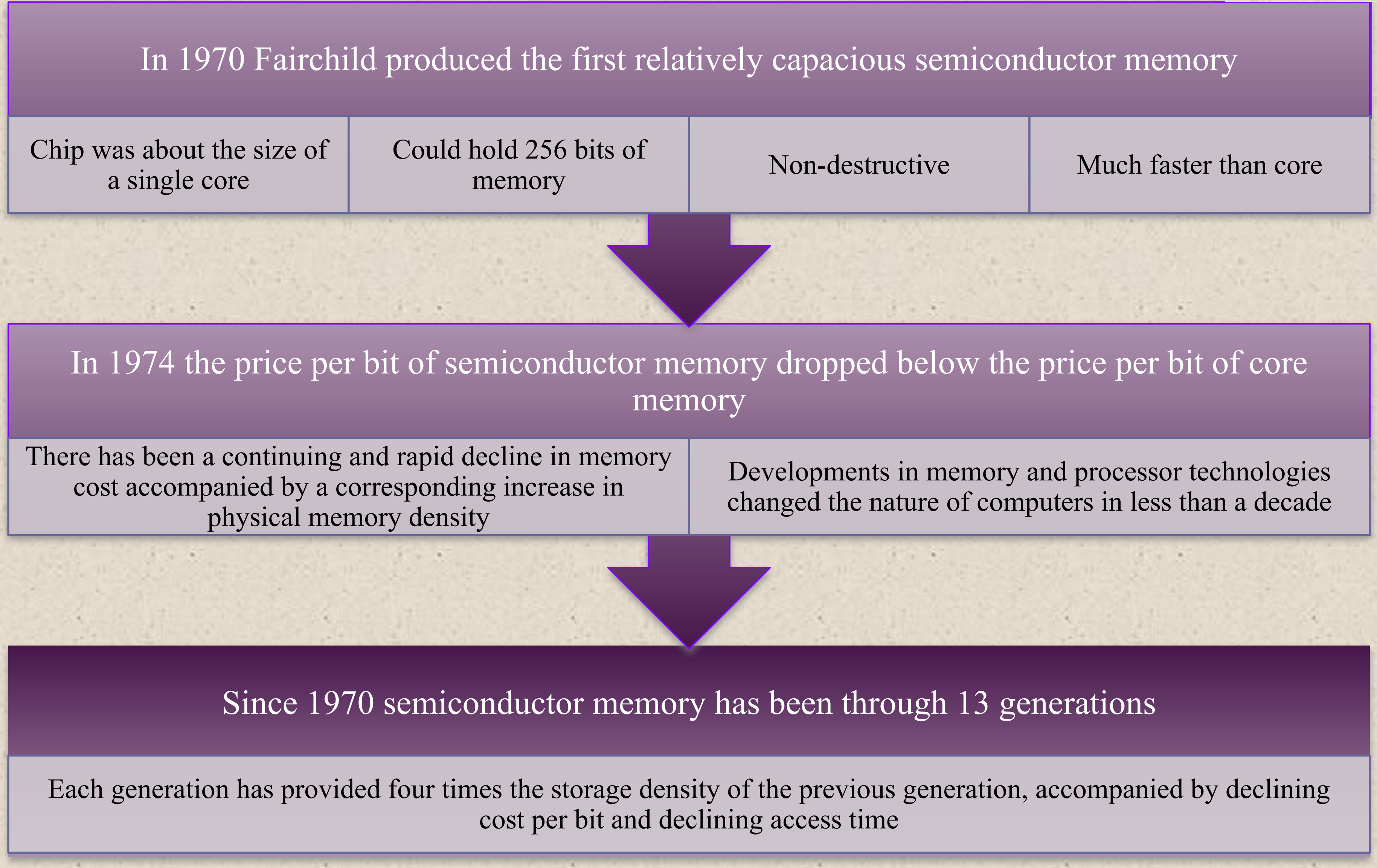
VLSI
Very Large
Scale
Integration



Semiconductor Memory
Microprocessors

ULSI
Ultra Large
Scale
Integration

Semiconductor Memory



In 1970 Fairchild produced the first relatively capacious semiconductor memory

- Chip was about the size of a single core
- Could hold 256 bits of memory
- Non-destructive
- Much faster than core

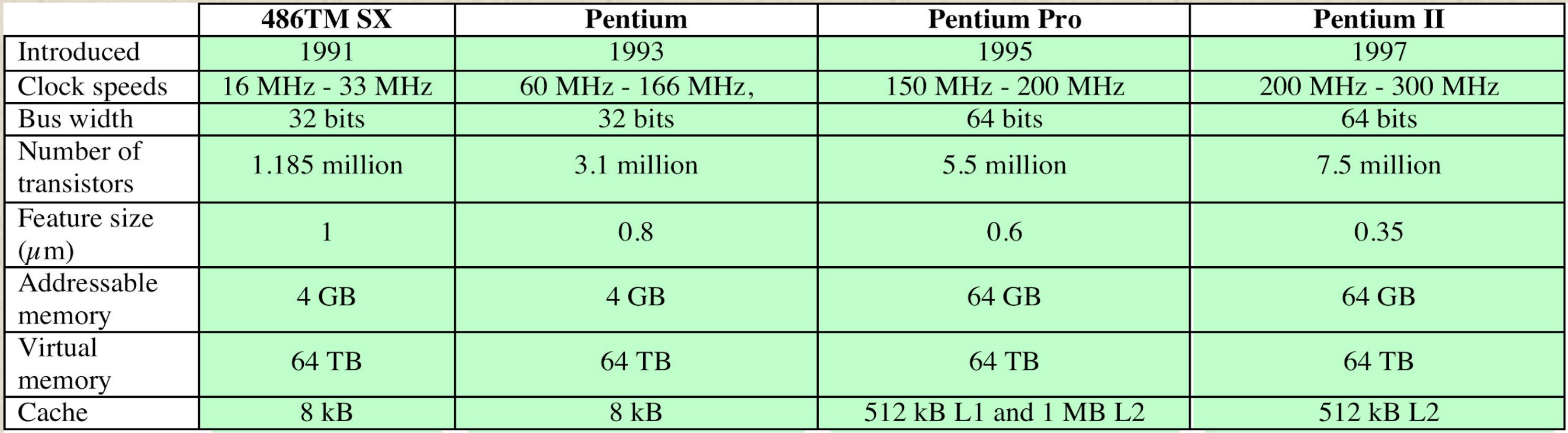
In 1974 the price per bit of semiconductor memory dropped below the price per bit of core memory

- There has been a continuing and rapid decline in memory cost accompanied by a corresponding increase in physical memory density
- Developments in memory and processor technologies changed the nature of computers in less than a decade

Since 1970 semiconductor memory has been through 13 generations

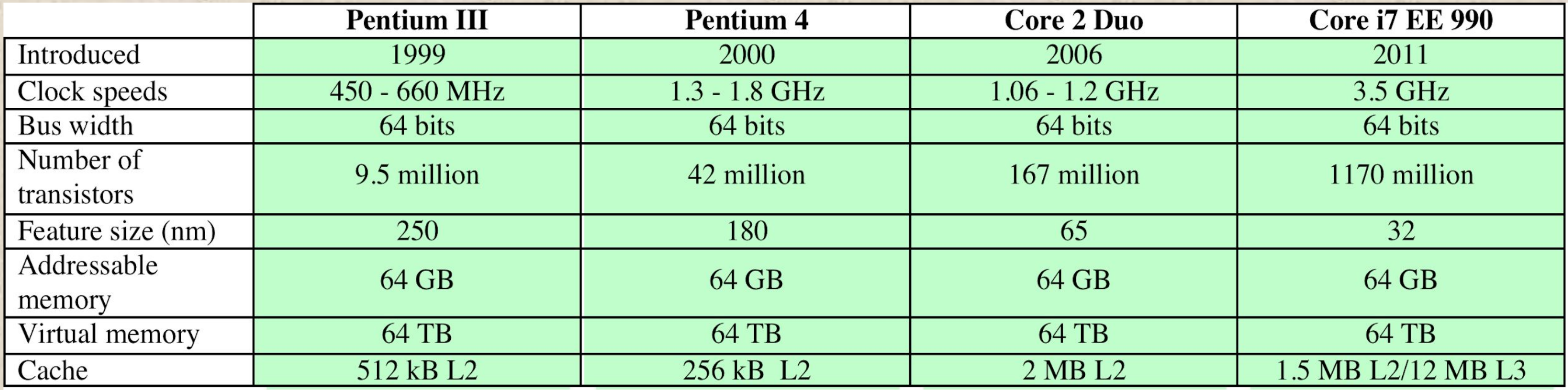
- Each generation has provided four times the storage density of the previous generation, accompanied by declining cost per bit and declining access time

Evolution of Intel Microprocessors



	486TM SX	Pentium	Pentium Pro	Pentium II
Introduced	1991	1993	1995	1997
Clock speeds	16 MHz - 33 MHz	60 MHz - 166 MHz,	150 MHz - 200 MHz	200 MHz - 300 MHz
Bus width	32 bits	32 bits	64 bits	64 bits
Number of transistors	1.185 million	3.1 million	5.5 million	7.5 million
Feature size (μm)	1	0.8	0.6	0.35
Addressable memory	4 GB	4 GB	64 GB	64 GB
Virtual memory	64 TB	64 TB	64 TB	64 TB
Cache	8 kB	8 kB	512 kB L1 and 1 MB L2	512 kB L2

c. 1990s Processors



	Pentium III	Pentium 4	Core 2 Duo	Core i7 EE 990
Introduced	1999	2000	2006	2011
Clock speeds	450 - 660 MHz	1.3 - 1.8 GHz	1.06 - 1.2 GHz	3.5 GHz
Bus width	64 bits	64 bits	64 bits	64 bits
Number of transistors	9.5 million	42 million	167 million	1170 million
Feature size (nm)	250	180	65	32
Addressable memory	64 GB	64 GB	64 GB	64 GB
Virtual memory	64 TB	64 TB	64 TB	64 TB
Cache	512 kB L2	256 kB L2	2 MB L2	1.5 MB L2/12 MB L3

d. Recent Processors

+ Performance Balance

- Adjust the organization and architecture to compensate for the mismatch among the capabilities of the various components
- Architectural examples include:

Increase the number of bits that are retrieved at one time by making DRAMs “wider” rather than “deeper” and by using wide bus data paths

Reduce the frequency of memory access by incorporating increasingly complex and efficient cache structures between the processor and main memory

Change the DRAM interface to make it more efficient by including a cache or other buffering scheme on the DRAM chip

Increase the interconnect bandwidth between processors and memory by using higher speed buses and a hierarchy of buses to buffer and structure data flow

Typical I/O Device Data Rates

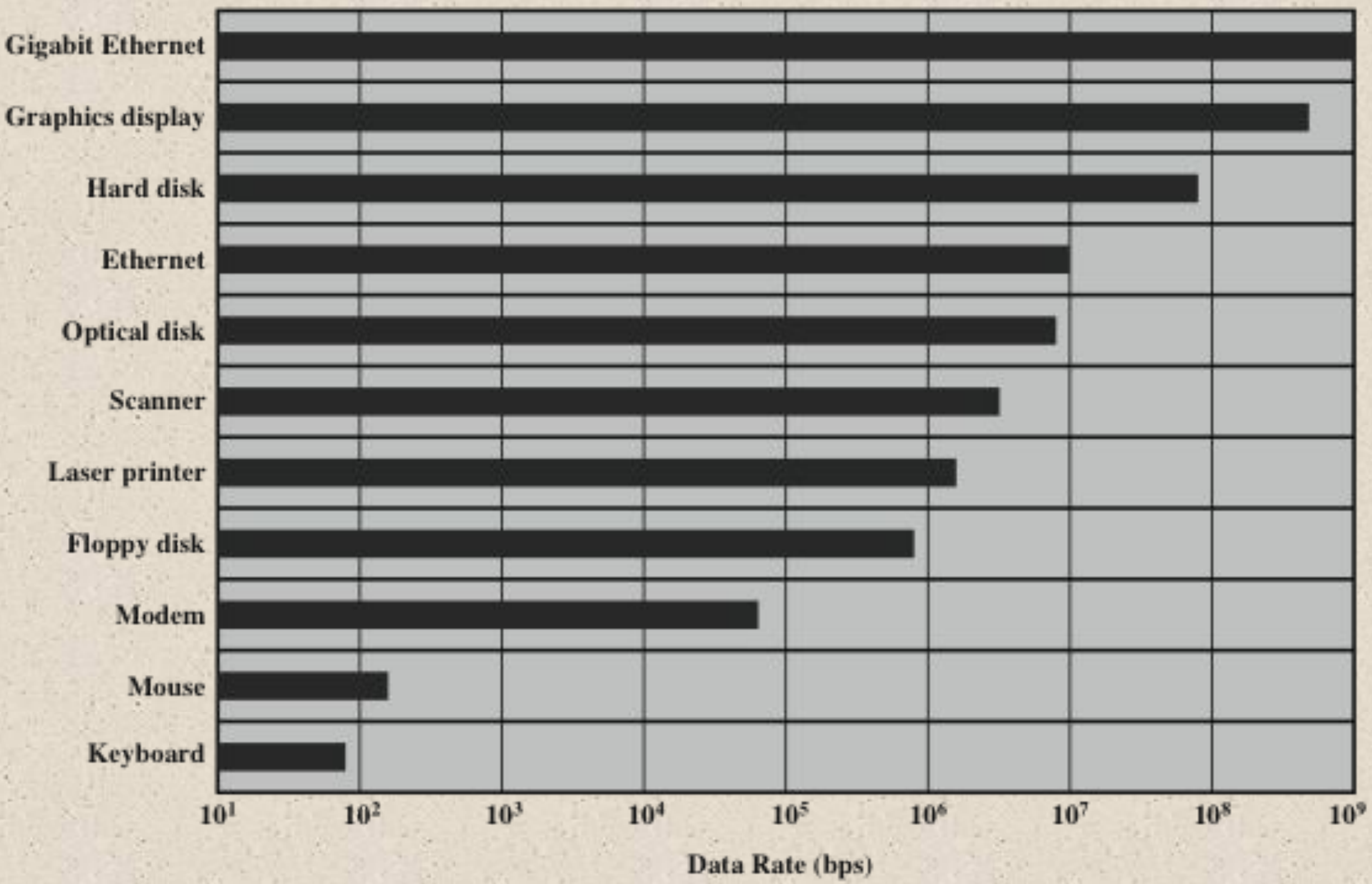


Figure 2.10 Typical I/O Device Data Rates

+

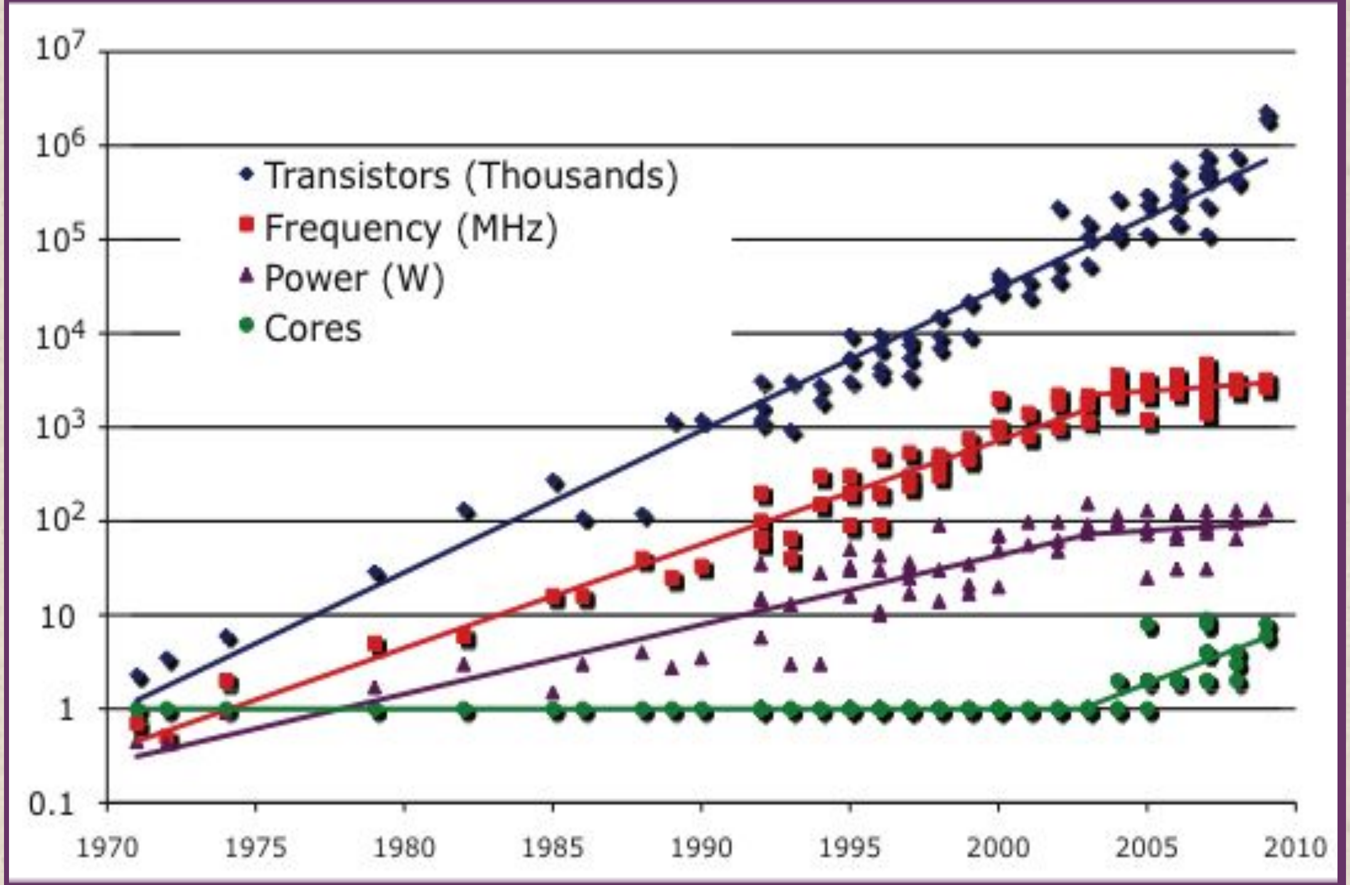
Improvements in Chip Organization and Architecture

- Increase hardware speed of processor
  - Fundamentally due to shrinking logic gate size
    - More gates, packed more tightly, increasing clock rate
    - Propagation time for signals reduced
- Increase size and speed of caches
  - Dedicating part of processor chip
    - Cache access times drop significantly
- Change processor organization and architecture
  - Increase effective speed of instruction execution
  - Parallelism

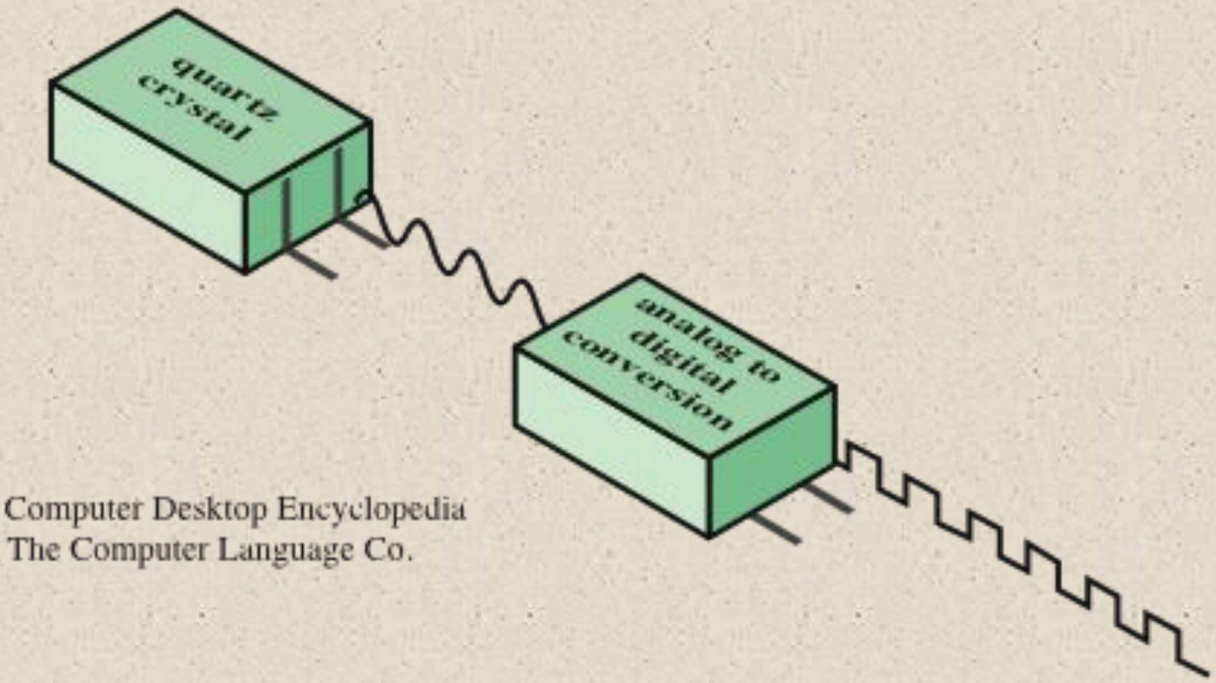
Problems with Clock Speed and Login Density

- Power
  - Power density increases with density of logic and clock speed
  - Dissipating heat
- RC delay
  - Speed at which electrons flow limited by resistance and capacitance of metal wires connecting them
  - Delay increases as RC product increases
  - Wire interconnects thinner, increasing resistance
  - Wires closer together, increasing capacitance
- Memory latency
  - Memory speeds lag processor speeds

Processor Trends



System Clock

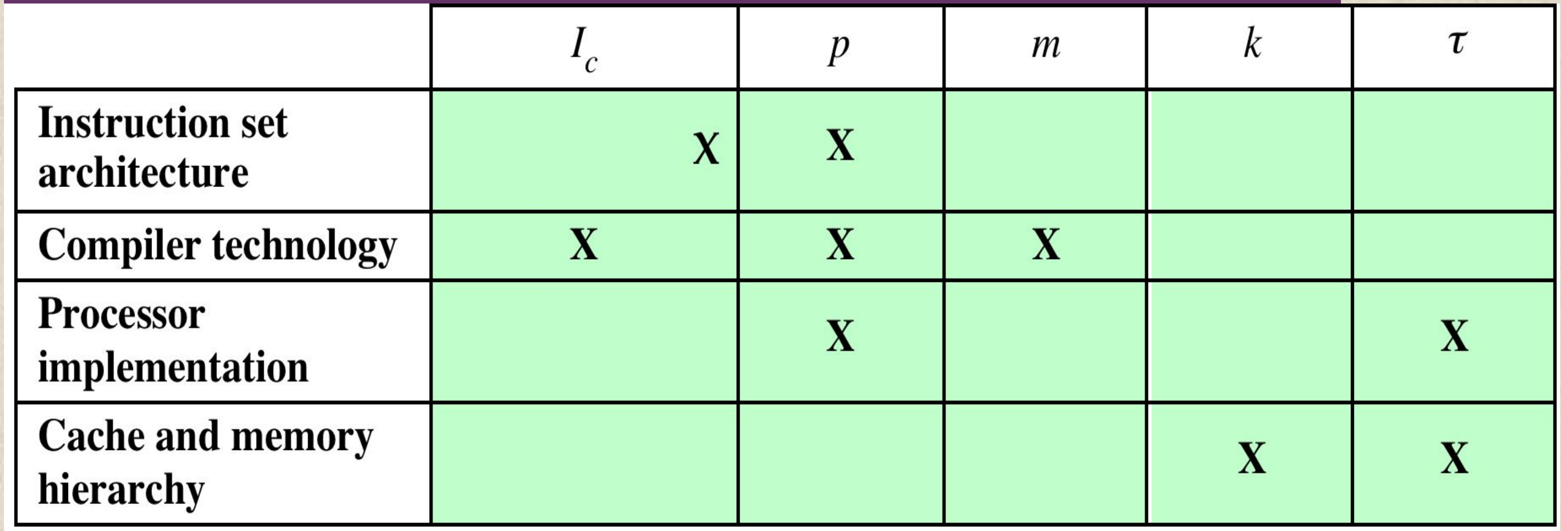


From Computer Desktop Encyclopedia
1998, The Computer Language Co.

Figure 2.13 System Clock

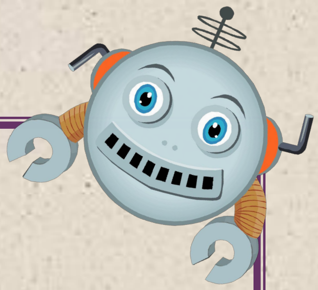
Performance Factors and System Attributes

Table 2.9



	I_c	p	m	k	τ
Instruction set architecture	**X**	**X**			
Compiler technology	**X**	**X**	**X**		
Processor implementation		**X**			**X**
Cache and memory hierarchy				**X**	**X**

Benchmarks



For example, consider this high-level language statement:

A = B + C /* assume all quantities in main memory */

With a traditional instruction set architecture, referred to as a complex instruction set computer (CISC), this instruction can be compiled into one processor instruction:

add mem(B), mem(C), mem (A)

On a typical RISC machine, the compilation would look something like this:

```
load mem(B), reg(1);
load mem(C), reg(2);
add reg(1), reg(2), reg(3);
store reg(3), mem (A)
```

Desirable Benchmark Characteristics

- Written in a high-level language, making it portable across different machines
- Representative of a particular kind of programming style, such as system programming, numerical programming, or commercial programming
- Can be measured easily
- Has wide distribution

+

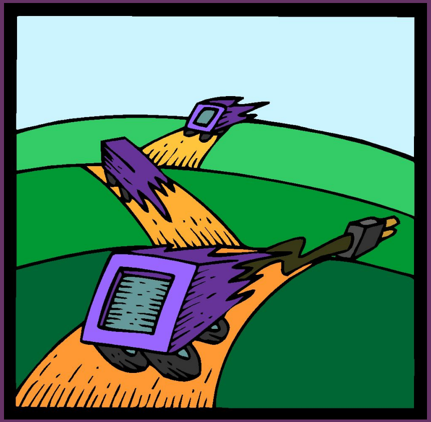
System Performance Evaluation Corporation (SPEC)

- Benchmark suite
  - A collection of programs, defined in a high-level language
  - Attempts to provide a representative test of a computer in a particular application or system programming area
- SPEC
  - An industry consortium
  - Defines and maintains the best known collection of benchmark suites
  - Performance measurements are widely used for comparison and research purposes

SPEC CPU2006

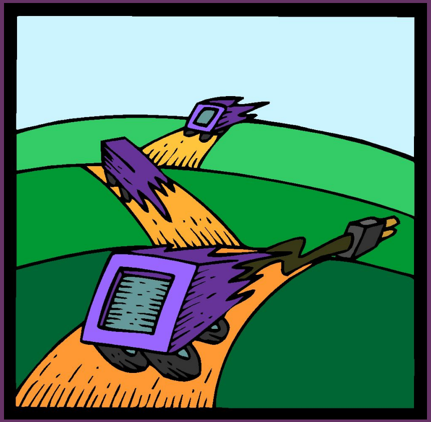
- Best known SPEC benchmark suite
- Industry standard suite for processor intensive applications
- Appropriate for measuring performance for applications that spend most of their time doing computation rather than I/O
- Consists of 17 floating point programs written in C, C++, and Fortran and 12 integer programs written in C and C++
- Suite contains over 3 million lines of code
- Fifth generation of processor intensive suites from SPEC

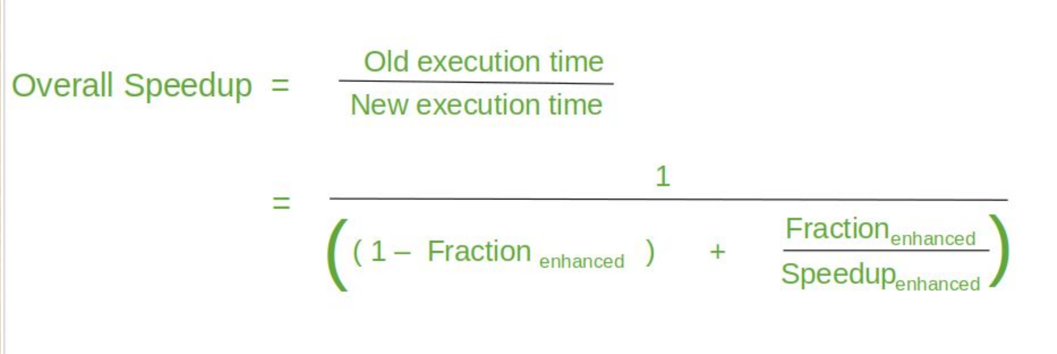
Amdahl's Law



- Gene Amdahl [AMDA67]
- Deals with the potential speedup of a program using multiple processors compared to a single processor
- Illustrates the problems facing industry in the development of multi-core machines
    - Software must be adapted to a highly parallel execution environment to exploit the power of parallel processing
- Can be generalized to evaluate and design technical improvement in a computer system

Amdahl's Law





$$\text{Overall Speedup} = \frac{\text{Old execution time}}{\text{New execution time}}$$

$$= \frac{1}{\left((1 - \text{Fraction}_{\text{enhanced}}) + \frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}} \right)}$$

+

Amdahl's Law

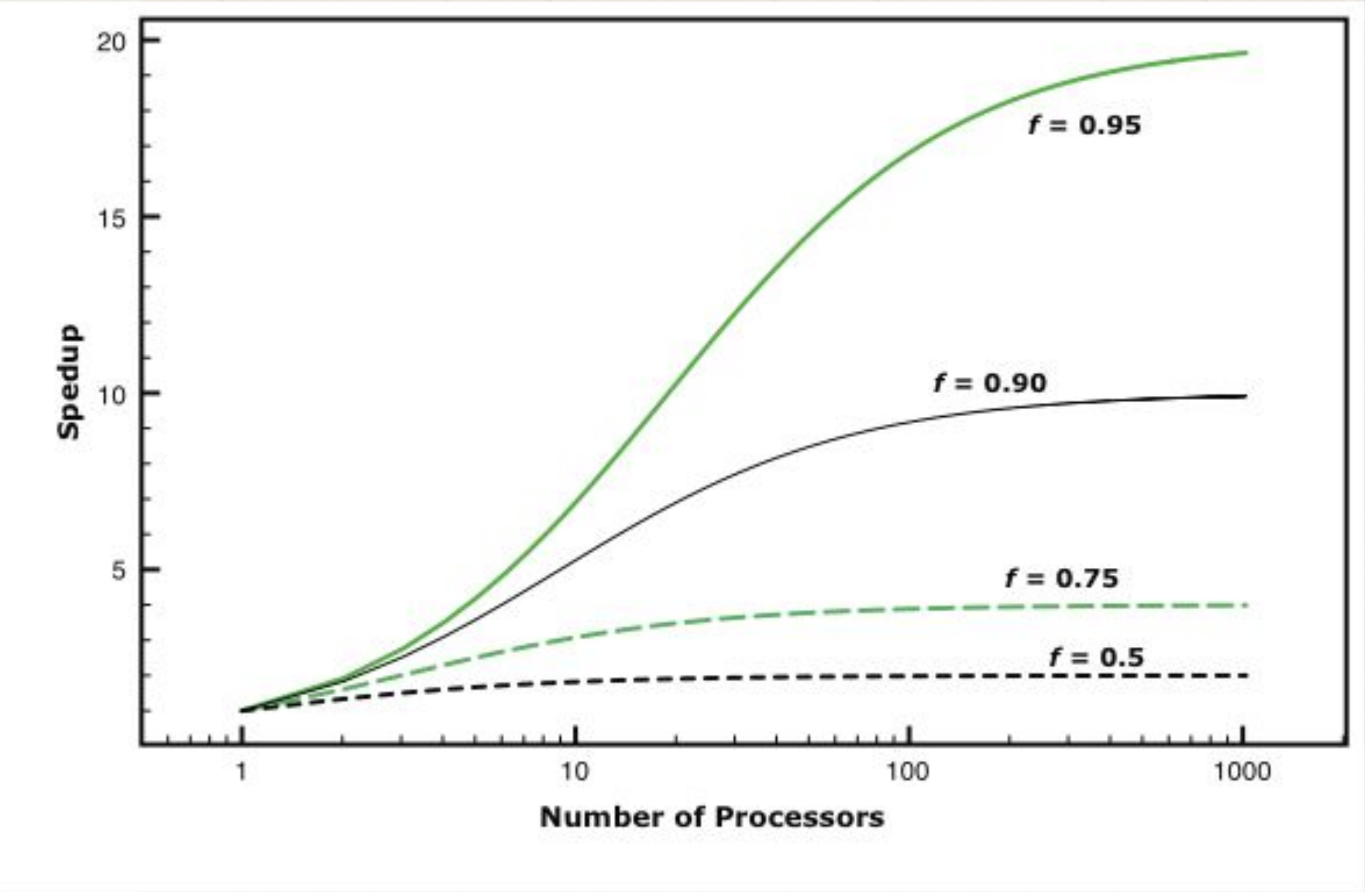


Figure 2.14 Amdahl's Law for Multiprocessors