Tishk International University
Mechatronics Engineering Department
Computer Architecture ME228/A
Lecture 2: 22/02/2022

Computer Architecture :Chapter-2
Computer Evolution and Performance

Outline

- A Brief History of Computers
- Designing for Performance
- Multicore, MICs and GPGPUs
- The Evolution of the Intel x86 Architecture
- Embedded Systems and the ARM
- Performance Assessment

Brief History of Computers

- First Generation: Vacuum Tubes
- Second Generation: Transistor
- Third Generation: Integrated Circuit

First Generation: Vacuum Tubes

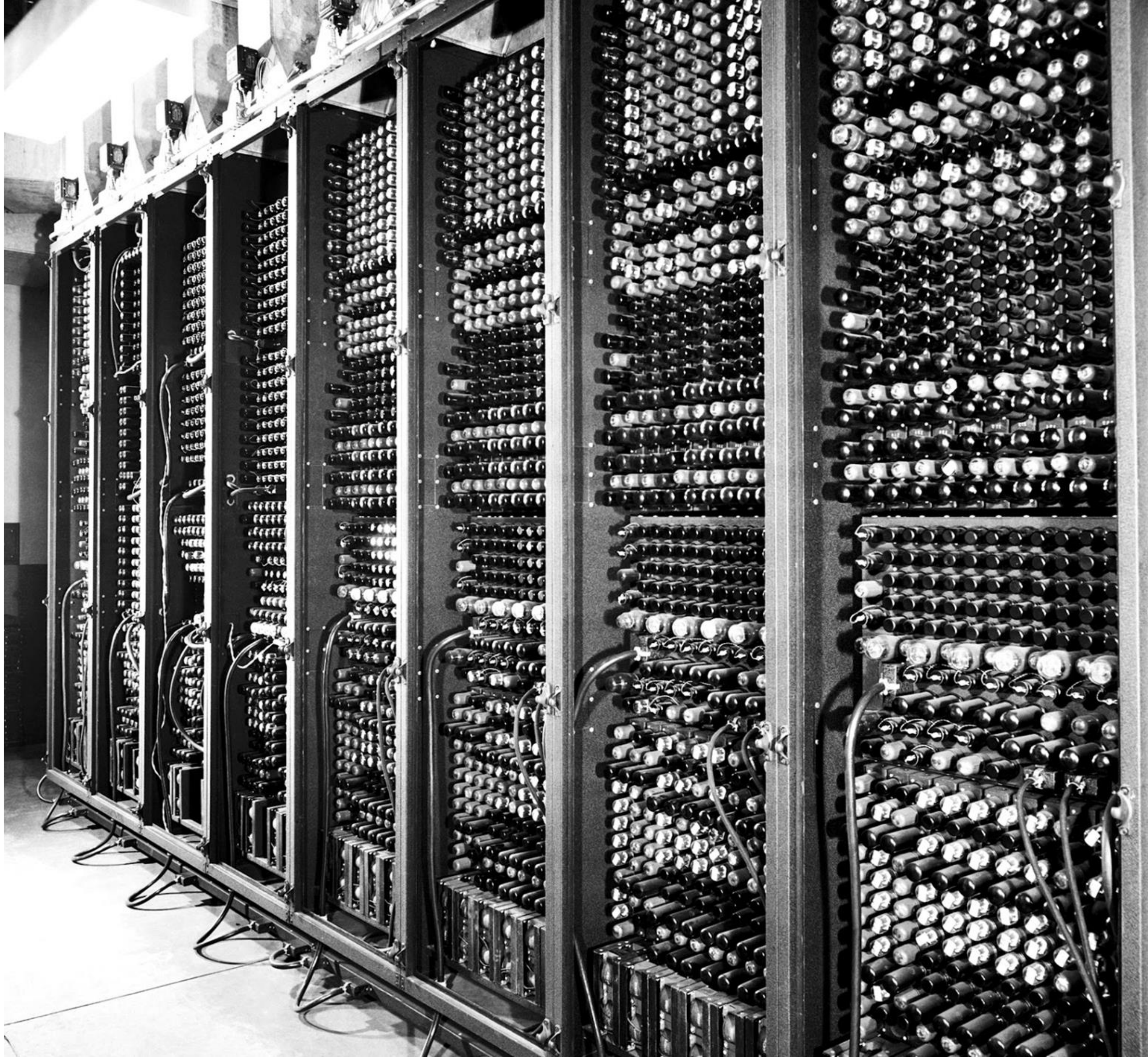
ENIAC: the ENIAC (Electronic Numerical Integrator And Computer)

- Designed and constructed at the University of Pennsylvania.
- Was the world's first general purpose electronic digital computer.
- The project was a response to U.S. needs during World War II.
- The Army's Ballistics Research Laboratory (BRL), which was responsible for developing range and trajectory tables for new weapons.
- The BRL employed more than 200 people who, using desktop calculators, solved the necessary equations.
- Preparation of the tables for a single weapon would take one person many hours, even days

First Generation: Vacuum Tubes

- John Mauchly, a professor of electrical engineering at the University of Pennsylvania, and John Eckert, one of his graduate students, proposed to build a general-purpose computer using vacuum tubes for the BRL's application.
- In 1943, the Army accepted this proposal, and work began on the ENIAC. The resulting machine was enormous:
  - Weighting 30 tons
  - Occupying 1500 square feet of floor space
  - Containing more than 18,000 vacuum tubes.
  - It consumed 140 kilowatts of power.
  - It was capable of 5000 additions per second
  - Decimal rather than binary

Here you see the nearly 18,000 vacuum tubes and 6,000 switches of the ENIAC, the first electronic computer.

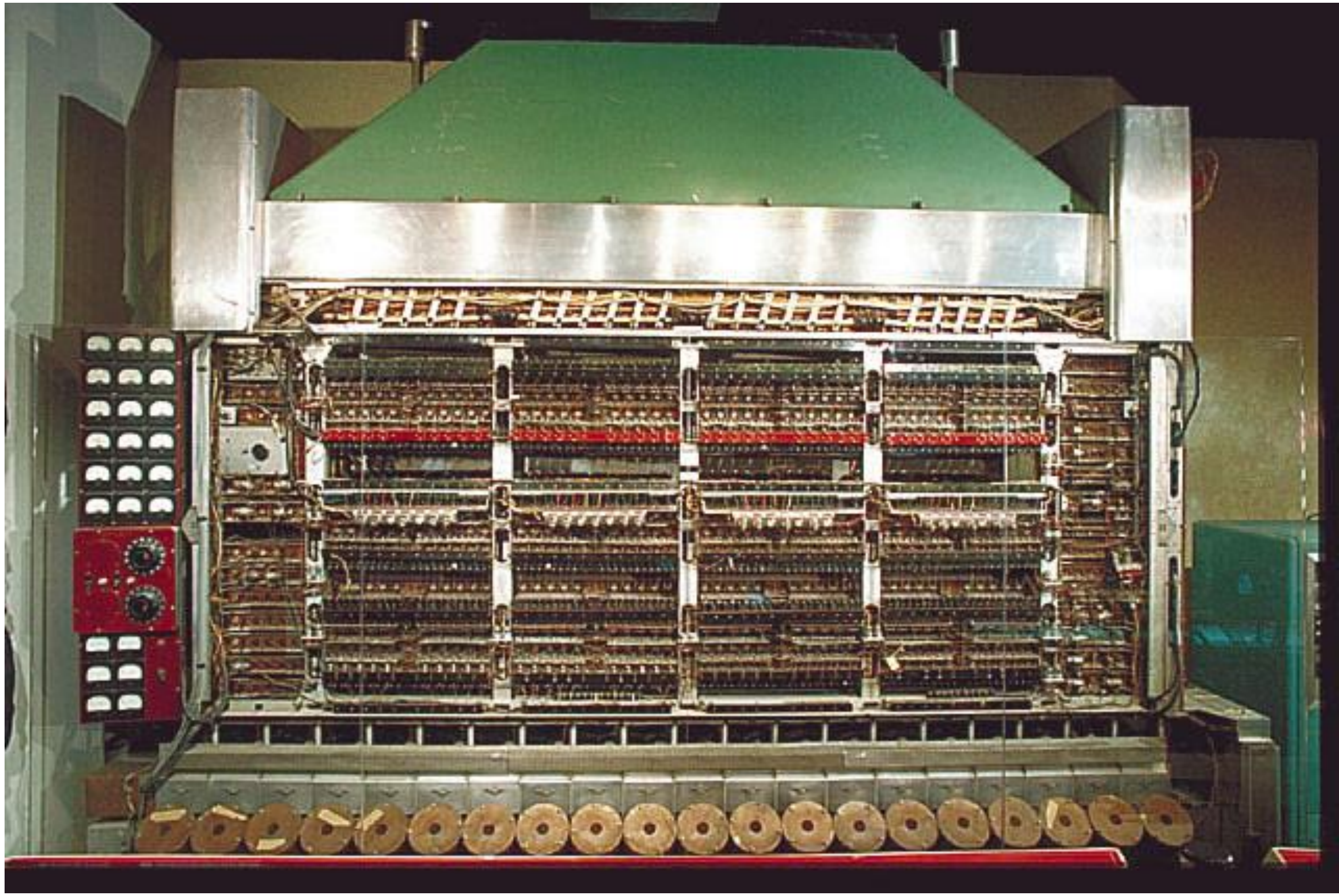


First Generation: Vacuum Tubes

- Its memory consisted of 20 accumulators, each capable of holding a 10-digit decimal number.
- A ring of 10 vacuum tubes represented each digit.
- One vacuum tube was in the ON state, representing one of the 10 digits.
- The major drawback of the ENIAC was that it had to be programmed manually by setting switches and plugging and unplugging cables.
- The ENIAC was completed in 1946.
- The ENIAC continued to operate under BRL management until 1955, when it was disassembled.

First Generation: Vacuum Tubes

- In 1946, von Neumann and his colleagues began the design of a new stored program computer, referred to as the **IAS Computer**, at the New-jersey Institute for Advanced Studies.
- The IAS computer completed in 1952



IAS Machine

First Generation: Vacuum Tubes

Figure 1 shows the general structure of the IAS computer. It consists of:

- A main memory , which stores both data and instructions.
- An arithmetic and logic unit (ALU) capable of operating on binary data.
- A control unit, which interprets the instructions in memory and causes them to be executed.
- Input/output (I/O) equipment operated by the control unit.

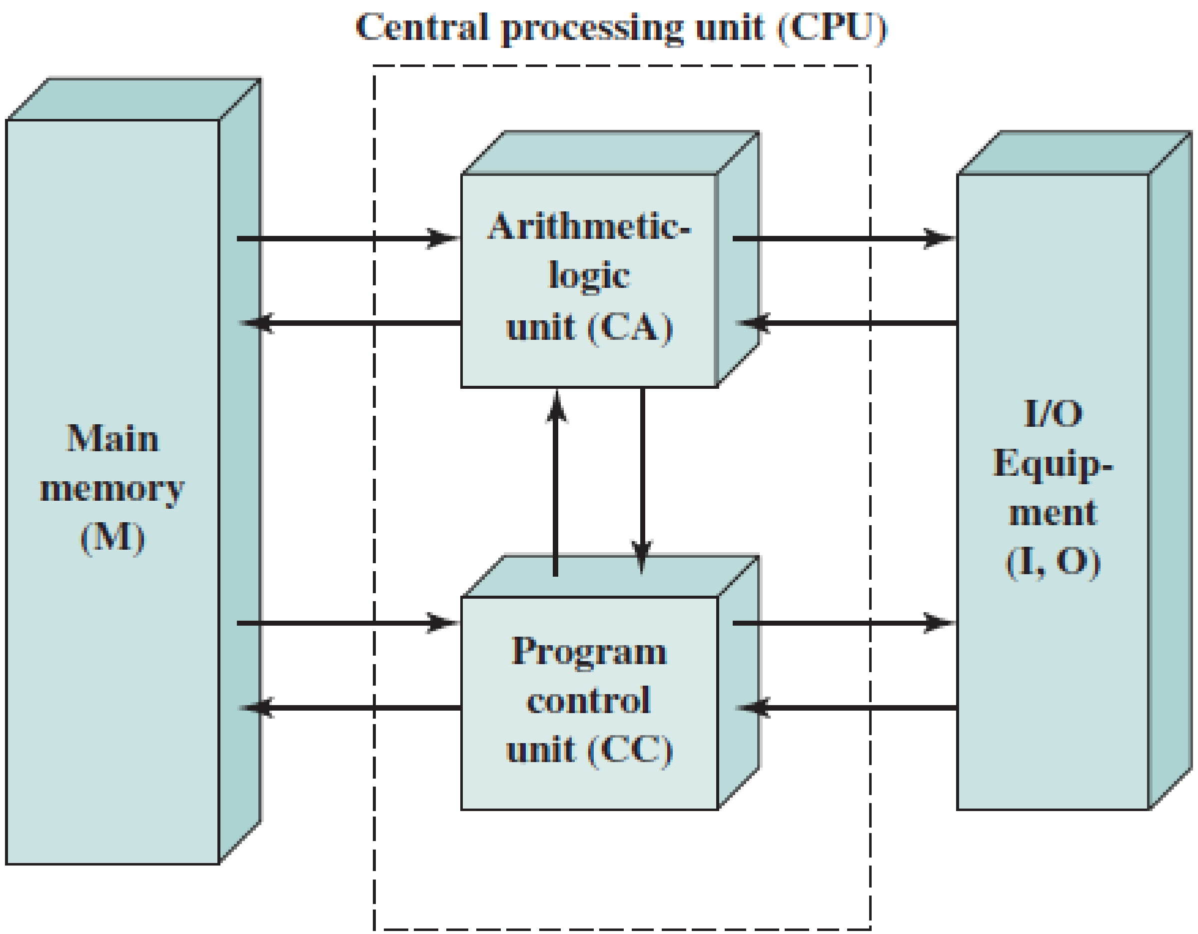


Fig. 1: Structure of the IAS Computer

IAS - details

- Consists 1000 storage locations called word, each has 40 bits.
- Used to store data and instruction
- Numbers represented in binary form
  - 1-bit for sign
  - 39-bits for value
- Instruction is a binary code
  - two 20-bits instructions
  - each consists of 8-bit operation code (opcode)
  - 12-bit address, to specify the word in the memory (0-999)

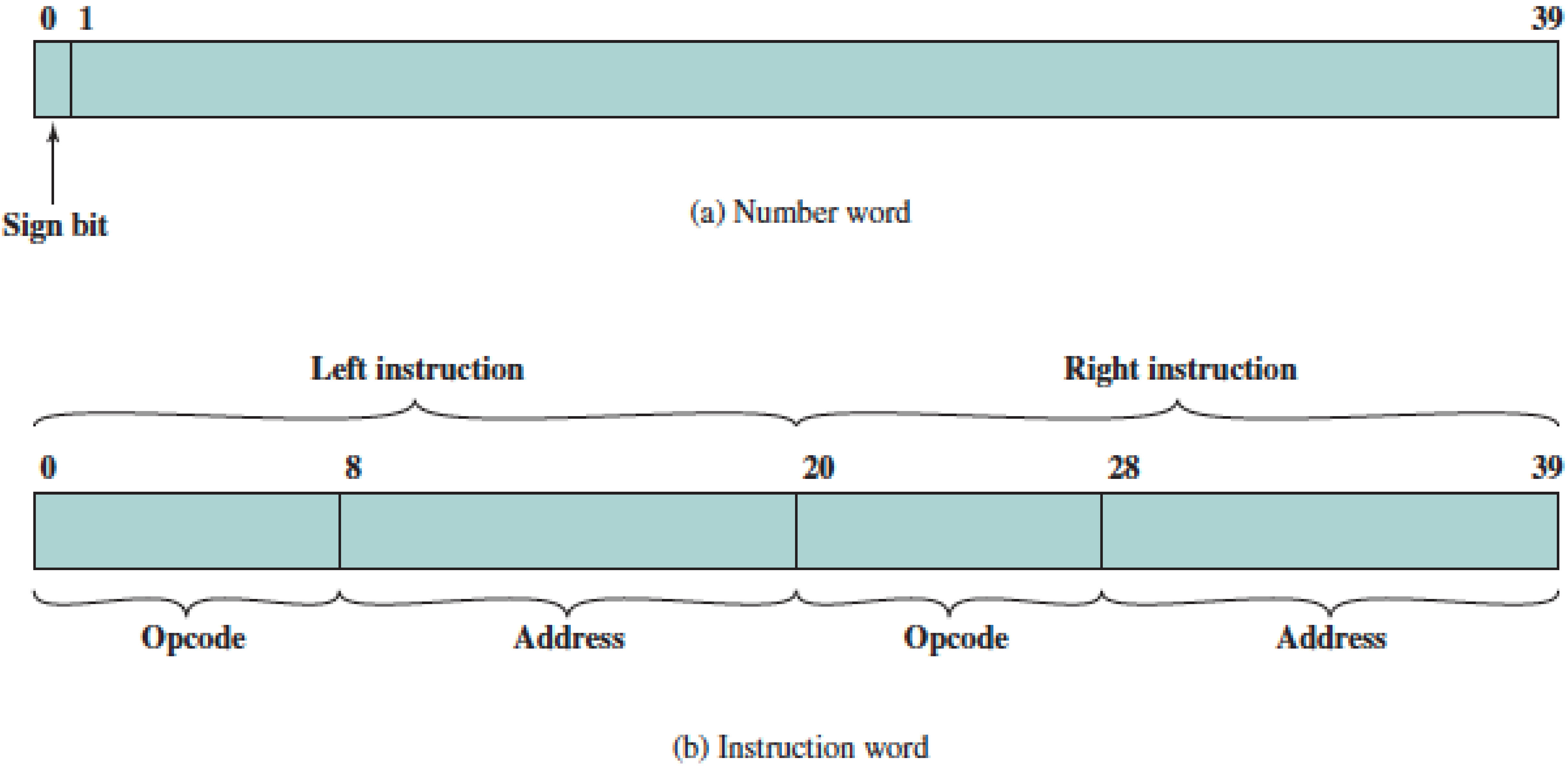


Fig. 2: IAS Memory Formats

Structure of IAS details*

Figure 3 reveals that both the control unit and the ALU contain storage locations, called registers , defined as follows:

- **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.

Structure of IAS details

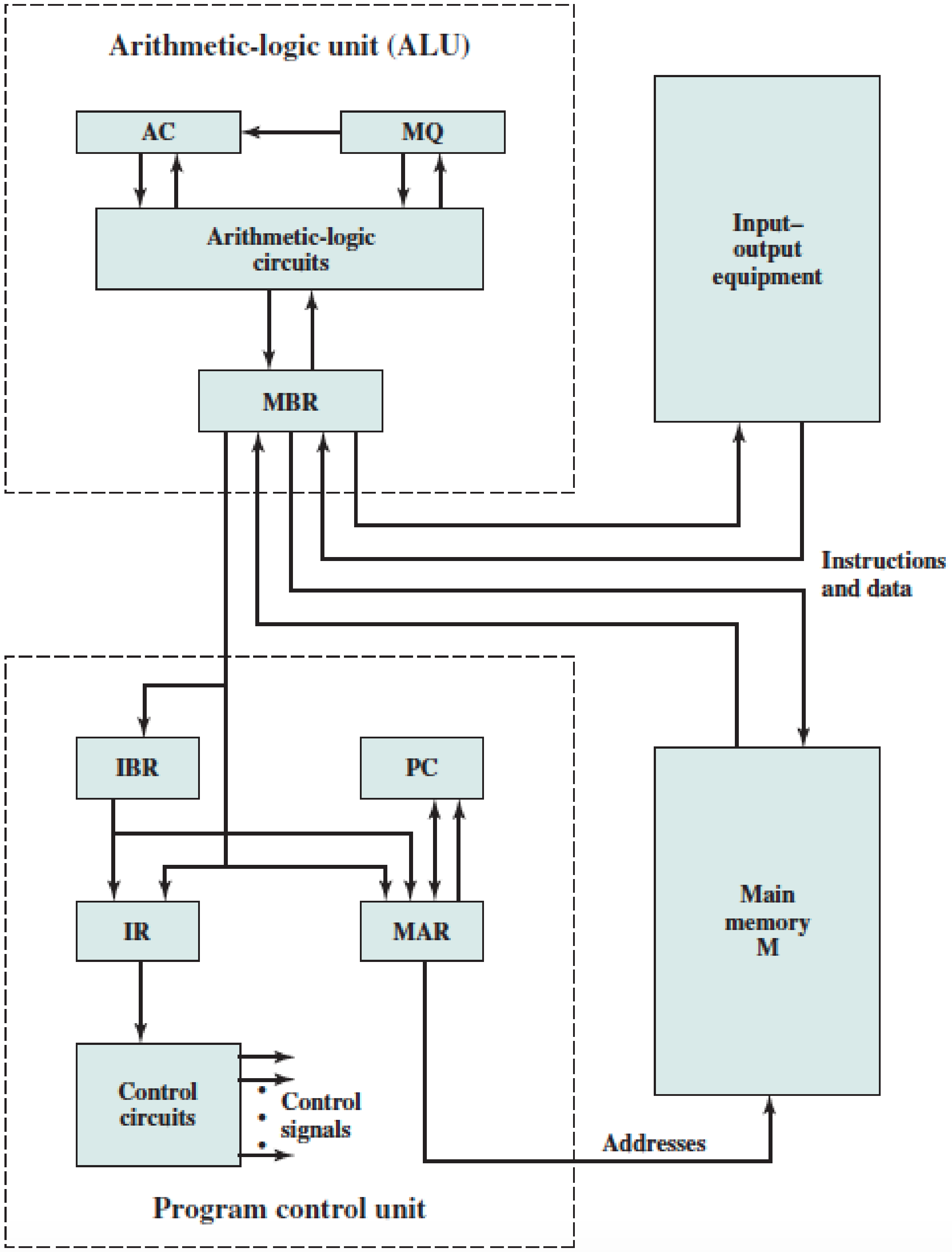


Fig. 3: Expanded Structure of IAS Computer

Structure of IAS details

- **Instruction buffer register (IBR)**: Employed to hold temporarily 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 hold temporarily operands and results of ALU operations. For example, the result of multiplying two 40-bit numbers is an 80-bit number; the most significant 40 bits are stored in the AC and the least significant in the MQ.

Structure of IAS details

The IAS computer had a total of 21 instructions, these can be grouped as follows:

- **Data transfer:** Move data between memory and ALU registers or between two ALU registers.
- **Unconditional branch:** Normally, the control unit executes instructions in sequence from memory. This sequence can be changed by a branch instruction, which facilitates repetitive operations.
- **Conditional branch:** The branch can be made dependent on a condition, thus allowing decision points.
- **Arithmetic:** Operations performed by the ALU.
- **Address modify:** Permits addresses to be computed in the ALU and then inserted into instructions stored in memory. This allows a program considerable addressing flexibility.

The Second Generation: Transistor

- Transistor is a replacement of vacuum tube.
- Transistor is:
  - smaller
  - cheaper
  - less heat
  - Made from silicon
- The transistor was invented at Bell Labs in 1947
- By the 1950s had launched an electronic revolution.
- At late 1950s, fully transistorized computers were commercially available.

Transistor based Computers

- The use of the transistor defines the second generation of computers.
- Classify computers into generations based on the fundamental hardware technology employed (Table 1).
- Each new generation is characterized by greater processing performance, larger memory capacity, and smaller size than the previous one.

Table 1: 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

The Third Generation: Integrated Circuits
1. Microelectronics

- Small electronics.
- The basic elements of a digital computer are: **storage, movement, processing, and control functions.**
- Two fundamental types of components are required: **gates and memory cells.**
- A gate is a device that implements a simple Boolean or logical function, such as:

IF A AND B ARE TRUE THEN C IS TRUE (AND gate)

- Such devices are called gates because they control data flow in much the same way that canal gates control the flow of water.
- The memory cell is a device that can store one bit of data; that is, the device can be in one of two stable states at any time.

Microelectronics

- The computer consists of: gates, memory cells, and interconnections.
- The gates and memory cells are constructed of simple digital electronic components.
- The integrated circuit consists components such 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.

Microelectronics

- Figure 4 reflects the famous Moore’s law, which was propounded by Gordon Moore, cofounder of Intel, in 1965.
- Moore observed that the number of transistors that could be put on a single chip was doubling every year and correctly predicted that this pace would continue into the near future.
- To the surprise of many, including Moore, the pace continued year after year and decade after decade. The pace slowed to a doubling every 18 months in the 1970s but has sustained that rate ever since

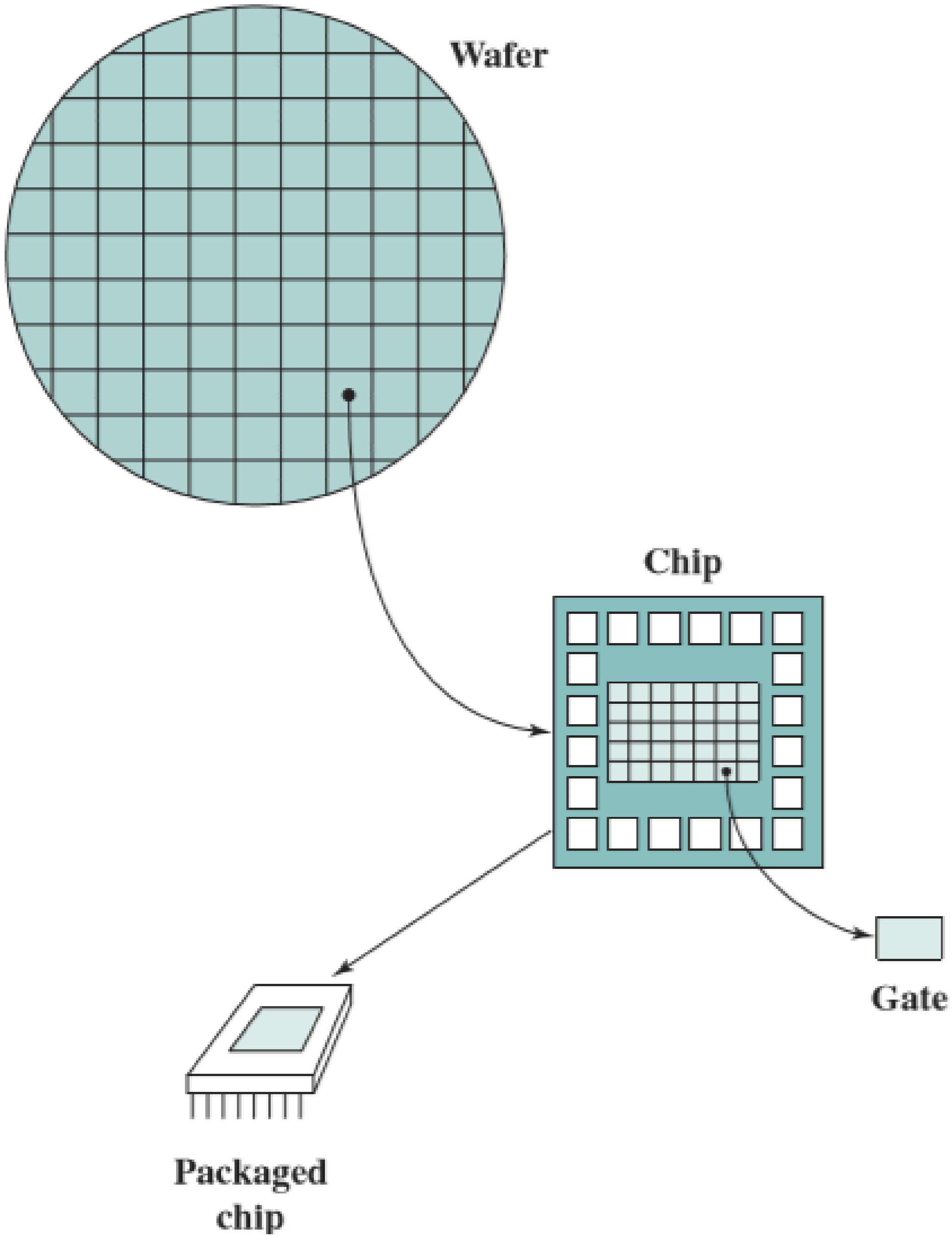


Figure 4: Relation between wafer, chip and gate

Moore's Law

1. The cost of a chip has remained virtually unchanged during this period of rapid growth in density. This means that the cost of computer logic and memory circuitry has fallen at a dramatic rate.
2. Because logic and memory elements are placed closer together on more densely packed chips, the electrical path length is shortened, increasing operating speed.
3. The computer becomes smaller, making it more convenient to place in a variety of environments.
4. There is a reduction in power and cooling requirements.
5. The interconnections on the integrated circuit are much more reliable than solder connections. With more circuitry on each chip, there are fewer inter-chip connections.

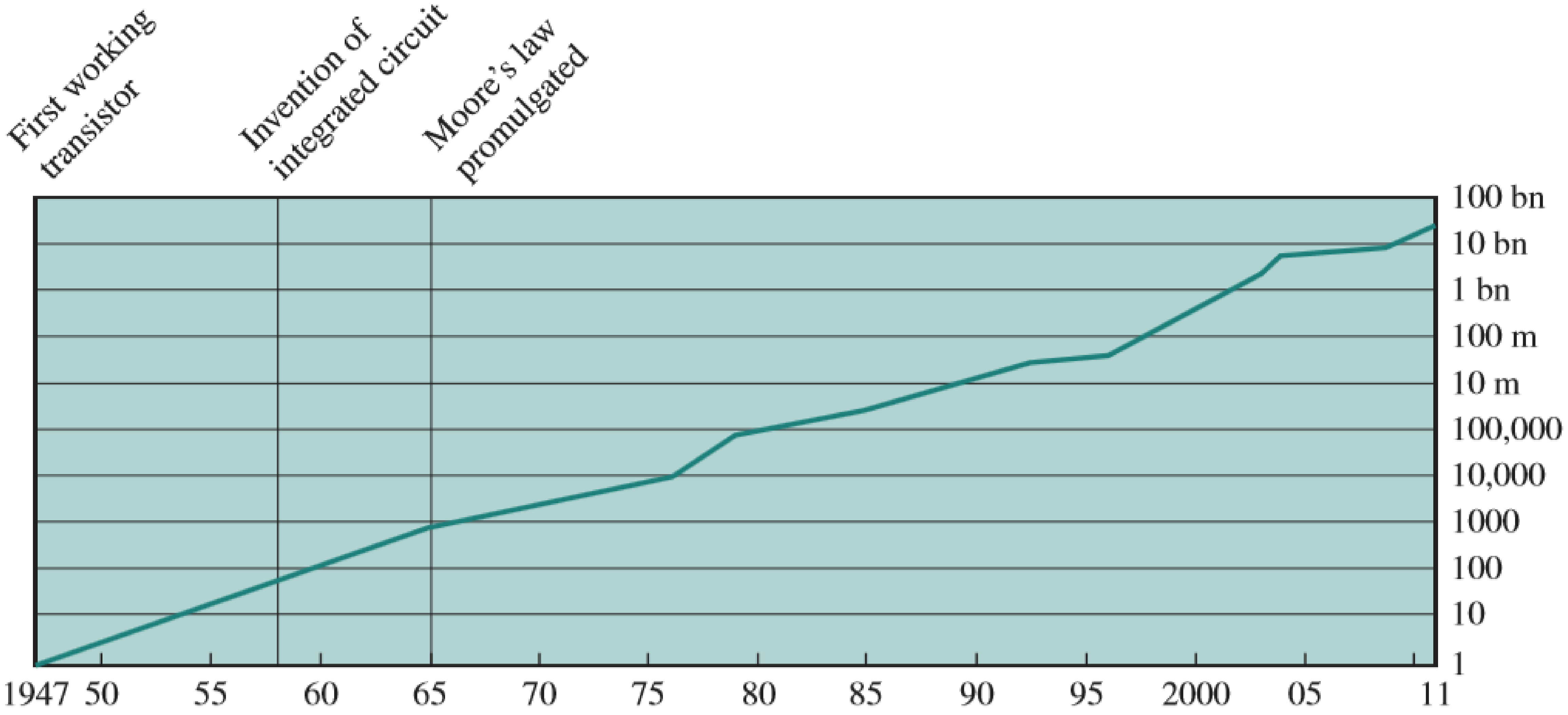


Figure 5: Growth in Transistor Count on Integrated Circuits

IBM System / 360

- In 1964:
  - IBM announced the System/360, a new family of computer products
  - the 360 product line was incompatible with older IBM machines
  - The System/360 was the industry’s first planned family of computers.
  - The family covered a wide range of performance and cost.
  - Table 1 indicates some of the key characteristics of the various models in 1965 (each member of the family is distinguished by a model number).
  - The models were compatible in the sense that a program written for one model should be capable of being executed by another model in the series, with only a difference in the time it takes to execute.

IBM System / 360*

Characteristic	Model 30	Model 40	Model 50	Model 65	Model 75
Maximum memory size (bytes)	64K	256K	256K	512K	512K
Data rate from memory (Mbytes/s)	0.5	0.8	2.0	8.0	16.0
Processor cycle time (μs)	1.0	0.625	0.5	0.25	0.2
Relative speed	1	3.5	10	21	50
Maximum number of data channels	3	3	4	6	6
Maximum data rate on one channel (Kbytes/s)	250	400	800	1250	1250

The characteristics of System/360 family are as follows:

- **Similar or identical instruction set:** In many cases, the exact same set of machine instructions is supported on all members of the family. Thus, a program that executes on one machine will also execute on any other. In some cases, the lower end of the family has an instruction set that is a subset of that of the top end of the family. This means that programs can move up but not down.
- **Similar or identical operating system:** The same basic operating system is available for all family members. In some cases, additional features are added to the higher-end members.

The characteristics of System/360 family are as follows:

- **Increasing speed:** The rate of instruction execution increases in going from lower to higher family members.
- **Increasing number of I/O ports**: The number of I/O ports increases in going from lower to higher family members.
- **Increasing memory size:** The size of main memory increases in going from lower to higher family members.
- **Increasing cost:** At a given point in time, the cost of a system increases in going from lower to higher family members.

DEC PDP-8 (Programmed Data Processor)

In 1964

- First minicomputer (after miniskirt!)
- Did not need air conditioned room
- Small enough to sit on a lab bench
- BUS STRUCTURE

DEC PDP-8

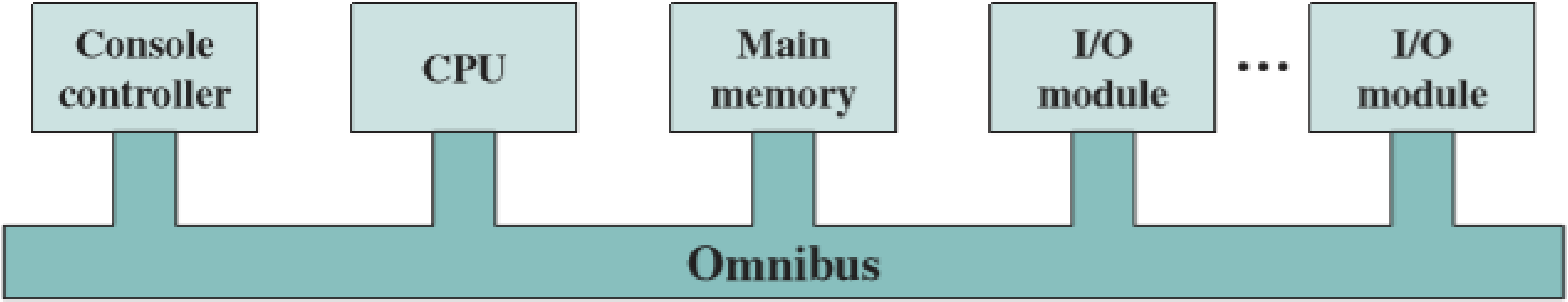


Fig. 6: DPD-8 Bus Structure

DEC PDP-8

Table 3: Evaluation of DPD-8

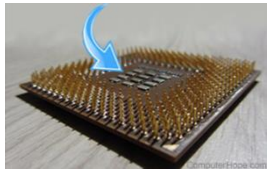
Model	First Shipped	Cost of Processor + 4K 12-bit Words of Memory ($1000s)	Data Rate from Memory (words/μs)	Volume (cubic feet)	Innovations and Improvements
PDP-8	4/65	16.2	1.26	8.0	Automatic wire-wrapping production
PDP-8/5	9/66	8.79	0.08	3.2	Serial instruction implementation
PDP-8/1	4/68	11.6	1.34	8.0	Medium-scale integrated circuits
PDP-8/L	11/68	7.0	1.26	2.0	Smaller cabinet
PDP-8/E	3/71	4.99	1.52	2.2	Omnibus
PDP-8/M	6/72	3.69	1.52	1.8	Half-size cabinet with fewer slots than 8/E
PDP-8/A	1/75	2.6	1.34	1.2	Semiconductor memory; floating-point processor

Semiconductor Memory

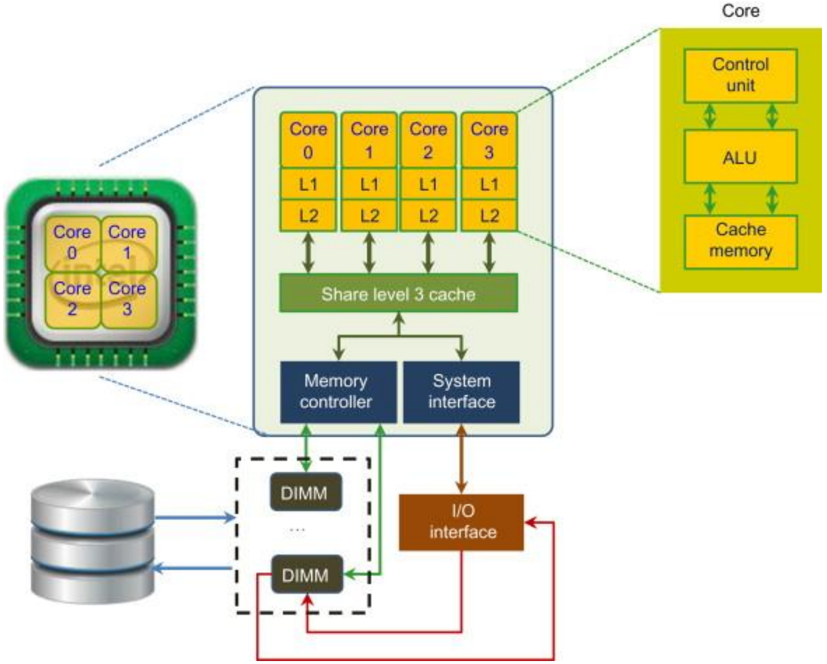
The first application of integrated circuit technology to computers was construction of the processor (the control unit and the arithmetic and logic unit) out of integrated circuit chips. But it was also found that this same technology could be used to construct memories.

- In 1970
- Fairchild

FAIRCHILD SEMICONDUCTOR®



- Size of a single core
  - i.e. 1 bit of magnetic core storage
  - a core is **a small CPU or processor built into a big CPU**
- Holds 256 bits
- Much faster than core
- Capacity approximately doubles each year



Semiconductor Memory

In 1974, a seminal event occurred: The price per bit of semiconductor memory dropped below the price per bit of core memory. Following this, there has been a continuing and rapid decline in memory cost accompanied by a corresponding increase in physical memory density. This has led the way to smaller, faster machines with memory sizes of larger and more expensive machines from just a few years earlier.

Semiconductor memory has been through 13 generations: 1K, 4K, 16K, 64K, 256K, 1M, 4M, 16M, 64M, 256M, 1G, 4G, and, as of this writing, 16 Gbits on a single chip (1K =2^{10}, 1M = 2^{20}, 1G = 2^{30}). Each generation has provided four times the storage density of the previous generation, accompanied by declining cost per bit and declining access time.

Microprocessor

- As time went on, more and more elements were placed on each chip, so that fewer and fewer chips were needed to construct a single computer processor.
- A breakthrough was achieved in ***1971***, when Intel developed its ***4004***. The 4004 was the first chip to contain all of the components of a CPU on a single chip: *The microprocessor was born*.
- The 4004 can add two 4-bit numbers and can multiply only by repeated addition. By today's standards, the 4004 is hopelessly primitive, but it marked the beginning of a continuing evolution of microprocessor capability and power.

Microprocessor Evaluation

- This evolution can be seen most easily in the number of bits that the processor deals with at a time.
- There is no clear-cut measure of this, but perhaps the best measure is the ***data bus width***: the number of bits of data that can be brought into or sent out of the processor at a time.
- Another measure is ***the number of bits in the accumulator or in the set of general-purpose registers***. Often, these measures coincide, but not always. For example, a number of microprocessors were developed that operate on 16-bit numbers in registers but can only read and write 8 bits at a time.
- In **1972** - **Intel 8008**, which was the first 8-bit microprocessor and was almost twice as complex as the 4004.

(a) 1970s Processors

	4004	8008	8080	8086	8088
Introduced	1971	1972	1974	1978	1979
Clock speeds	108 kHz	108 kHz	2 MHz	5 MHz, 8 MHz, 10 MHz	5 MHz, 8 MHz
Bus width	4 bits	8 bits	8 bits	16 bits	8 bits
Number of transistors	2300	3500	6000	29,000	29,000
Feature size (µm)	10		6	3	6
Addressable memory	640 Bytes	16 kB	64 kB	1 MB	1 MB

(b) 1980s Processors

	80286	386TM DX	386TM SX	486TM DX CPU
Introduced	1982	1985	1988	1989
Clock speeds	6 MHz–12.5 MHz	16 MHz–33 MHz	16 MHz–33 MHz	25 MHz–50 MHz
Bus width	16 bits	32 bits	16 bits	32 bits
Number of transistors	134,000	275,000	275,000	1.2 million
Feature size (µm)	1.5	1	1	0.8–1
Addressable memory	16 MB	4 GB	16 MB	4 GB
Virtual memory	1 GB	64 TB	64 TB	64 TB
Cache	—	—	—	8 kB

Microprocessor Evaluation

- In **1974** - **Intel 8080**, which was the first general-purpose microprocessor. Whereas the 4004 and the 8008 had been designed for specific applications, the 8080 was designed to be the CPU of a general-purpose microcomputer.
- Like the 8008, the 8080 is an 8-bit microprocessor. The 8080, however, is **faster**, has a **richer instruction set**, and has a **large addressing capability**.
- About the same time, 16-bit microprocessors began to be developed. However, it was not until the end of the 1970s that powerful, general-purpose 16-bit microprocessors appeared. One of these was the 8086.
- The next step in this trend occurred in 1981, when both Bell Labs and Hewlett-Packard developed 32-bit, s ingle-chip microprocessors.
- Intel introduced its own 32-bit microprocessor, the 80386, in 1985.

(c) 1990s Processors

	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

(d) Recent 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

Exercises

- Define: Moore's Law
- What are the generations of developing the computer, list them and the main differences between them.
- How the microprocessor performance can be improved.
- What are the general structure of the IAS computer list them and briefly describe them.
- Draw the general structure of the IAS computer
- Briefly describe the development of the Microprocessor.