Chapter 1

Basic Concepts and Computer Evolution

Computer Architecture

Computer Organization

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

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

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

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

Computer Architecture

- Computer architecture is the science and art of selecting and interconnecting hardware components to create a computer
- Computer architecture comes before computer organization
- Its like building the design and architecture of house takes maximum time and then organization is building house by bricks or by latest technology keeping the basic layout and architecture of house in mind

Computer organization

- It is concerned with the way the hardware are connected together to form the computer system
- It is those operational attributes are linked together and to realize the architectural specifications

Computer Architecture and Organization

- Computer Architecture =

  Instruction Set Architecture +

  Computer Organization

- Instruction Set Architecture (ISA)
  - What the computer does (logical view)
- Computer Organization
  - How the ISA is implemented (physical view)

- Many computer manufacturers offer a family of computer models, all with the same architecture but with differences in organization.
- Consequently, the different models in the family have different price and performance characteristics.
- Furthermore, a particular architecture may span many years and encompass a number of different computer models, its organization changing with changing technology.

- Architectural issue if a computer will have multiply instruction
- Organization is how to implement multiply instruction
  - By a special multiply unit
  - By a repeated use of addition unit
  - Decision is based on the anticipated frequency of use of multiply instruction, relative speed of the two approaches and the cost and physical size of the special multiply unit.

Structure and Function

- Computer is a complex system
- In order to understand it, we recognize the its hierarchical nature
  - 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

- There are four basic functions that a computer can perform:
  - Data processing
    - Data may take a wide variety of forms and the range of processing requirements is broad
  - Data storage
    - Short-term
    - Long-term
  - Data movement
    - Input-output (I/O) - when data are received from or delivered to a device (peripheral) that is directly connected to the computer
    - Data communications – when data are moved over longer distances, to or from a remote device
  - Control
    - A control unit manages the computer's resources and orchestrates the performance of its functional parts in response to instructions

Functions

Computer activity consists of input, processing, storage, and output

Function

Figure 1.1 A Functional View of the Computer

Possible computer operations

Figure 1.2 Possible Computer Operations

- Peripheral Devices
  - When data are received from or delivered to a device that is directly connected to the computer, the process is known as *input– output* (I/O), and the device is referred to as a *peripheral.*
- Data Communication
  - When data are moved over longer distances, to or from a remote device, the process is known as *data communications.*

The Computer

Figure 1.3 The Computer

Structure

We are concerned with the internal structure of the computer

There are four main structural components of the computer:

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

Structure of computer

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

Computer
I/O
System Bus
CPU
Memory
CPU
Registers
Arithmetic and Login Unit
Internal CPU Interconnection
Control Unit

Computer
I/O
System Bus
CPU
Memory
CPU
Registers
Arithmetic and Login Unit
Internal CPU Interconnection
Control Unit

Instruction Decode and Prefetch Unit
Code Cache
Branch Predictor
Core
Integer ALU
Registers
Execution Unit
32-bit Buses
To RAM
Floating Point Unit
Data Cache
64-bit Bus
Primary Cache
Bus Interface

Processor

- Central processing unit (CPU)
  - Portion of the computer that fetches and executes instructions
  - Consists of an ALU, a control unit, and registers
  - Referred to as a processor in a system with a single processing unit
- Core
  - An individual processing unit on a processor chip
  - May be equivalent in functionality to a CPU on a single-CPU system
  - Specialized processing units are also referred to as cores
- Processor
  - A physical piece of silicon containing one or more cores
  - Is the computer component that interprets and executes instructions
  - Referred to as a *multicore processor* if it contains multiple cores

Multi-core processors

- A single chip with 2 or more cores
- Multiple cores can run multiple instructions at the same time increasing overall speed for programs
- Parallel programming

Performance

Figure 1.2 Simplified View of Major Elements of a Multicore Computer

Figure 1.3

Motherboard with Two Intel Quad-Core Xeon Processors

Intel Core i9

Intel® Core™ i9-9900K Processor (16M Cache, up to 5.00 GHz)

- 16 MB SmartCache Cache
- 8 Cores
- 16 Threads
- 5.00 GHz Max Turbo Frequency
- K - Unlocked
- 9th Generation

Intel® Core™ i9-8950HK Processor (12M Cache, up to 4.80 GHz)

- 12 MB SmartCache Cache
- 6 Cores
- 12 Threads
- 4.80 GHz Max Turbo Frequency
- HK - High performance graphics, unlocked
- 8th Generation

★ ★ ★ ★ ☆
69 User Reviews

Intel® Core™ i9-7980XE Extreme Edition Processor

- 24.75 MB Cache
- 18 Cores
- 36 Threads
- 4.20 GHz Max Turbo Frequency
- XE - Extreme performance and mega-tasking, unlocked
- 7th Generation

Intel Core i9 X-Series

Product Name	Status	Launch Date	# of Cores	Max Turbo Frequency	Processor Base Frequency	Cache
Intel® Core™ i9-10900X X-series Processor (19.25M Cache, 3.70 GHz)	Launched	Q4'19	10	4.50 GHz	3.70 GHz	19.25 MB Intel® Smart Cache
Intel® Core™ i9-10980XE Extreme Edition Processor (24.75M Cache, 3.00 GHz)	Launched	Q4'19	18	4.60 GHz	3.00 GHz	24.75 MB Intel® Smart Cache
Intel® Core™ i9-10940X X-series Processor (19.25M Cache, 3.30 GHz)	Launched	Q4'19	14	4.60 GHz	3.30 GHz	19.25 MB Intel® Smart Cache
Intel® Core™ i9-10920X X-series Processor (19.25M Cache, 3.50 GHz)	Launched	Q4'19	12	4.60 GHz	3.50 GHz	19.25 MB Intel® Smart Cache

GPU – Graphics Processing Unit

- A specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display
- Modern GPUs are very efficient at manipulating computer graphics and image processing, and their highly parallel structure makes them more efficient than general-purpose CPUs for algorithms where the processing of large blocks of data is done in parallel
- Nvidia in 1999
- Video card, mother board, CPU die

- Nvidia GeForece RTX 2080 Ti
  - 4352 CUDA cores, 11 GB GDDR6, 1.1-1.3 GHz, 15W
- AMD's Radeon RX 480
  - 36 compute units and 2304 stream processors
  - 1.1 – 1.2 GHz
  - 150W
  - 5.8TFLOPs (Trillion Floating-point Operations Per Second)

- GeForce RTX 3090
  - NVIDIA CUDA Cores 10496
  - Boost Clock 1.70 GHz
  - Memory Size 24 GB
  - Memory Type GDDR6X

GPGPU

- GPGPU – General Purpose Graphical Processing Unit
- Parallel processing between one or more GPUs and CPUs
- While GPUs generally operate at lower frequencies, they usually have many times more cores to make up for it (up to hundreds at least) and can, thus operate on pictures and graphical data effectively much faster, dozens of even hundreds of times faster than a traditional CPU, migrating data into graphical form and then using the GPU to "look" at it and analyze it can result in profound speedup.

GPGPU

GPU Acceleration

Videos

- Adam and Jamie paint the Mona Lisa in 80 milliseconds
  - https://www.youtube.com/watch?v=WmW6SD-EHVY
- Multi-core
  - https://www.youtube.com/watch?v=S3I5WNHbnJ0
- CPUs vs GPUs
  - https://www.youtube.com/watch?v=1kypaBjJ-pg
  - https://www.youtube.com/watch?v=jznazEHJKiY
  - https://www.youtube.com/watch?v=pGN1na3F5do
- Core i3, Core i5 and Corei7
  - https://www.youtube.com/watch?v=GLSPub4ydiM

History of Computers

First Generation: Vacuum Tubes

- Vacuum tubes were used for digital logic elements and memory
- IAS computer – Stored-program computer
  - Fundamental design approach was the stored program concept
    - Attributed to the mathematician John von Neumann
    - First publication of the idea was in 1945 for the EDVAC
  - Design began at the Princeton Institute for Advanced Studies
  - Completed in 1952
  - Prototype of all subsequent general-purpose computers
  - IAS – Institute for Advanced Studies

Figure 1.6 IAS Structure

Figure 1.7 IAS Memory Formats

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

Table 1.1

The IAS Instruction Set

Instruction Type	Opcode	Symbolic Representation	Description
Data transfer	00001010	LOAD MQ	Transfer contents of register MQ to the accumulator AC
	00001001	LOAD MQ,M(X)	Transfer contents of memory location X to MQ
	00100001	STOR M(X)	Transfer contents of accumulator to memory location X
	00000001	LOAD M(X)	Transfer M(X) to the accumulator
	00000010	LOAD –M(X)	Transfer –M(X) to the accumulator
	00000011	LOAD |M(X)|	Transfer absolute value of M(X) to the accumulator
	00000100	LOAD –|M(X)|	Transfer –|M(X)| to the accumulator
Unconditional branch	00001101	JUMP M(X,0:19)	Take next instruction from left half of M(X)
	00001110	JUMP M(X,20:39)	Take next instruction from right half of M(X)
Conditional branch	00001111	JUMP+ M(X,0:19)	If number in the accumulator is nonnegative, take next instruction from left half of M(X)
		JUMP + M(X,20:39)	*If number in the accumulator is nonnegative, take next instruction from right half of M(X)*
Arithmetic	00000101	ADD M(X)	Add M(X) to AC; put the result in AC
	00000111	ADD |M(X)|	Add |M(X)| to AC; put the result in AC
	00000110	SUB M(X)	Subtract M(X) from AC; put the result in AC
	00001000	SUB |M(X)|	Subtract |M(X)| from AC; put the remainder in AC
	00001011	MUL M(X)	Multiply M(X) by MQ; put most significant bits of result in AC, put least significant bits in MQ
	00001100	DIV M(X)	Divide AC by M(X); put the quotient in MQ and the remainder in AC
	00010100	LSH	Multiply accumulator by 2; i.e., shift left one bit position
	00010101	RSH	Divide accumulator by 2; i.e., shift right one position
Address modify	00010010	STOR M(X,8:19)	Replace left address field at M(X) by 12 rightmost bits of AC
	00010011	STOR M(X,28:39)	Replace right address field at M(X) by 12 rightmost bits of AC

(Table can be found on page 17 in the textbook.)

M(X) = contents of memory location whose address is X
(i:j) = bits i through j

- 1950s – IBM and Sperry
- UNIVAC I and II, 1100, 1103
- IBM 701, 702, 700/7000

- With rare exceptions, all of today's computers have this same general structure and function and are thus referred to as **von Neumann machines**

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 transistor computers were commercially available

How a Transistor works?

- https://www.youtube.com/watch?v=IcrBqCFLHIY
- https://www.youtube.com/watch?v=7ukDKVHnac4

Table 1.2
Computer Generations

Generation	Approximate Dates	Technology	Typical Speed (operations per second)
1	1946–1957	Vacuum tube	40,000
2	1957–1964	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
    - Libraries perform common computations

Table 2.3 Example members of the IBM 700/7000 Series

Model Number	First Delivery	CPU Tech-nology	Memory Tech-nology	Cycle Time (μs)	Memory Size (K)	Number of Opcodes	Number of Index Registers	Hardwired Floating-Point	I/O Overlap (Channels)	Instruc-tion Fetch Overlap	Speed (relative to 701)
701	1952	Vacuum tubes	Electrostatic tubes	30	2–4	24	0	no	no	no	1
704	1955	Vacuum tubes	Core	12	4–32	80	3	yes	no	no	2.5
709	1958	Vacuum tubes	Core	12	32	140	3	yes	yes	no	4
7090	1960	Transistor	Core	2.18	32	169	3	yes	yes	no	25
7094 I	1962	Transistor	Core	2	32	185	7	yes (double precision)	yes	yes	30
7094 II	1964	Transistor	Core	1.4	32	185	7	yes (double precision)	yes	yes	50

Figure 1.9 An IBM 7094 Configuration

- Discrete components
  - Transistors, resistors, capacitors and so on

Transistor

Resistors

Capacitor

- *Problems with 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

History of Computers

Third Generation: Integrated Circuits

- 1958 – the invention of the integrated circuit
- The two most important members of the third gener
  were the IBM System/360 and the DEC PDP-8

- Only two fundamental types of components are required to manufacuter computers; gates and memory cells.
- A gate is a device that implements a simple Boolean or logical function.
- 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.

Figure 1.10 Fundamental Computer Elements

Fundamental components to computer functions

- **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. For example, a gate will have one or two data inputs plus a control signal input that activates the gate. When the control signal is ON, the gate performs its function on the data inputs and produces a data output. Conversely, when the control signal is OFF, the output line is null, such as the one produced by a high impedance state. Similarly, the memory cell will store the bit that is on its input lead when the WRITE control signal is ON and will place the bit that is in the cell on its output lead when the READ control signal is ON.

Integrated Circuit

Integrated Circuits

- 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 process of metallization to form circuits

Figure 1.11 Relationship Among Wafer, Chip, and Gate

Fabrication of IC

- https://www.youtube.com/watch?v=JDROPMoNZpk

Figure 1.12 Growth in Transistor Count on Integrated Circuits (DRAM memory)

Processor trends

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

Magnetic-core memory

- Used in 1950s to 1960s
- Expensive and bulky
- Destructive readout
  - Simple act of reading a core erased the data stored in it

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.

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

K, 4K, 16K, 64K, 256K, 1M, 4M, 16M, 64M, 256M, 1G, 4G, and, as of this writing, 16 Gbits on a single chip. Each generation has provided four times the storage density of the previous generation, accompanied by declining cost per bit and declining access time

Microprocessors

- The density of elements on processor chips continued to rise
  - More and more elements were placed on each chip so that fewer and fewer chips were needed to construct a single computer processor
- 1971 Intel developed 4004
  - First chip to contain all of the components of a CPU on a single chip. Could add two 4-bit numbers and multiply by addition.
  - Birth of microprocessor
- 1972 Intel developed 8008
  - First 8-bit microprocessor
- 1974 Intel developed 8080
  - First general purpose microprocessor
  - Faster, has a richer instruction set, has a large addressing capability

Videos

- https://www.youtube.com/watch?v=7MgHsgoilQ4
- https://www.youtube.com/watch?v=oBpaEmTcIRU
- https://www.youtube.com/watch?v=peX0rICizC0

Evolution of Intel Microprocessors

	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	2,300	3,500	6,000	29,000	29,000
Feature size (μm)	10	8	6	3	6
Addressable memory	640 Bytes	16 KB	64 KB	1 MB	1 MB

(a) 1970s Processors

Evolution of Intel Microprocessors

	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

(b) 1980s Processors

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

Evolution of Intel Microprocessors

	Pentium III	Pentium 4	Core 2 Duo	Core i7 EE 4960X
Introduced	1999	2000	2006	2013
Clock speeds	450 - 660 MHz	1.3 - 1.8 GHz	1.06 - 1.2 GHz	4 GHz
Bus width	64 bits	64 bits	64 bits	64 bits
Number of transistors	9.5 million	42 million	167 million	1.86 billion
Feature size (nm)	250	180	65	22
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 L2/15 MB L3
Number of cores	1	1	2	6

(d) Recent Processors

Videos

- https://www.youtube.com/watch?v=FPks89DvOms
- https://www.youtube.com/watch?v=utiat2AOxnA

The Evolution of the Intel x86 Architecture

- Two processor families are the Intel x86 and the ARM architectures
- Current x86 offerings represent the results of decades of design effort on complex instruction set computers (CISCs)
- An alternative approach to processor design is the reduced instruction set computer (RISC)
- ARM architecture is used in a wide variety of embedded systems and is one of the most powerful and best-designed RISC-based systems on the market

Highlights of the Evolution of the Intel Product Line:

8080

- World's first general-purpose microprocessor
- 8-bit machine, 8-bit data path to memory
- Was used in the first personal computer (Altair)

8086

- A more powerful 16-bit machine
- Has an instruction cache, or queue, that prefetches a few instructions before they are executed
- The first appearance of the x86 architecture
- The 8088 was a variant of this processor and used in IBM's first personal computer (securing the success of Intel

80286

- Extension of the 8086 enabling addressing a 16-MB memory instead of just 1MB

80386

- Intel's first 32-bit machine
- First Intel processor to support multitasking

80486

- Introduced the use of much more sophisticated and powerful cache technology and sophisticated instruction pipelining
- Also offered a built-in math coprocessor

Highlights of the Evolution of the Intel Product Line:

Pentium

- Intel introduced the use of superscalar techniques, which allow multiple instructions to execute in parallel

Pentium Pro

- Continued the move into superscalar organization with aggressive use of register renaming, branch prediction, data flow analysis, and speculative execution

Pentium II

- Incorporated Intel MMX technology, which is designed specifically to process video, audio, and graphics data efficiently

Pentium III

- Incorporated additional floating-point instructions
- Streaming SIMD Extensions (SSE)

Pentium 4

- Includes additional floating-point and other enhancements for multimedia

Core

- First Intel x86 micro-core

Core 2

- Extends the Core architecture to 64 bits
- Core 2 Quad provides four cores on a single chip
- More recent Core offerings have up to 10 cores per chip
- An important addition to the architecture was the Advanced Vector Extensions instruction set

AMD

DESKTOP GRAPHICS

Radeon Pro Duo
Radeon™ RX Series Graphics
AMD Radeon™ R9 Series Graphics
AMD Radeon™ R7 Series Graphics
AMD Radeon™ R5 Series Graphics
Preinstalled Graphics

DESKTOP PROCESSORS

AMD FX™ Processors
AMD Athlon™ X4 Processors
AMD A-Series APUs
AMD Athlon™ APUs
AMD Sempron™ APUs

EMBEDDED

AMD R-Series Processors

LAPTOP GRAPHICS

AMD Radeon™ R9 Series Graphics
AMD Radeon™ R7 Series Graphics
AMD Radeon™ R5 Series Graphics

LAPTOP PROCESSORS

CHIPSETS

MEMORY

SOLID STATE DRIVES (SSD)

PROFESSIONAL GRAPHICS

AMD FirePro™ Desktop Workstation Graphics
AMD FirePro™ Mobile Workstation Graphics
AMD FirePro™ Multi-Display Graphics
AMD FirePro™ for Servers

AMD PRO A-SERIES APUS

SERVER PROCESSORS

Based on x86 design

Embedded Systems

- The use of electronics and software within a product
- Billions of computer systems are produced each year that are embedded within larger devices

- Today many devices that use electric power have an embedded computing system

- Often embedded systems are tightly coupled to their environment
  - Constraints
    - Real-time
    - Power
    - Portability
    - Poor interface

Video

- https://www.youtube.com/watch?v=y70V0qHAFNQ

The Internet of Things (IoT)

- The expanding interconnection of smart devices, ranging from appliances to tiny sensors
- Is primarily driven by deeply embedded devices
- Generations of deployment culminating in the IoT:
  - Information technology (IT)
    - PCs, servers, routers, firewalls, and so on, bought as IT devices by enterprise IT people and primarily using wired connectivity
  - Operational technology (OT)
    - Machines/appliances with embedded IT built by non-IT companies, such as medical machinery, process control etc using wired connectivity
  - Personal technology
    - Smartphones, tablets, and eBook readers bought as IT devices by consumers exclusively using wireless connectivity
  - Sensor/actuator technology (IoT)
    - Single-purpose devices bought by consumers, IT, and OT people exclusively using wireless connectivity, generally of a single form, as part of larger systems

Videos

- https://www.youtube.com/watch?v=LVlT4sX6uVs
- https://www.youtube.com/watch?v=AhrB5ZZXnjM
- https://www.youtube.com/watch?v=SDJVFr4VUHA&list=PLcUDkVNXTfvRicfBYlclGZELQcFSAYrrT

ARM

Refers to a processor architecture that has evolved from RISC design principles and is used in embedded systems

Family of RISC-based microprocessors and microcontrollers designed by ARM Holdings, Cambridge, England

Chips are high-speed processors that are known for their small die size and low power requirements

Probably the most widely used embedded processor architecture and indeed the most widely used processor architecture of any kind in the world

Acorn RISC Machine/Advanced RISC Machine

ARM Products

Cortex-A

Highest performance
Optimized for rich operating systems

Cortex-R

Fast response
Optimized for high-performance, hard real-time applications

Cortex-M

Smallest/lowest power
Optimized for discrete processing and microcontroller

SecurCore

Tamper resistant
Optimized for security applications

Figure 1.16 Typical Microcontroller Chip Based on Cortex-M3

ARM Evolution

Table 2.8 ARM Evolution

Family	Notable Features	Cache	Typical MIPS @ MHz
ARM1	32-bit RISC	None	
ARM2	Multiply and swap instructions; Integrated memory management unit, graphics and I/O processor	None	7 MIPS @ 12 MHz
ARM3	First use of processor cache	4 kB unified	12 MIPS @ 25 MHz
ARM6	First to support 32-bit addresses; floating-point unit	4 kB unified	28 MIPS @ 33 MHz
ARM7	Integrated SoC	8 kB unified	60 MIPS @ 60 MHz
ARM8	5-stage pipeline; static branch prediction	8 kB unified	84 MIPS @ 72 MHz
ARM9		16 kB/16 kB	300 MIPS @ 300 MHz
ARM9E	Enhanced DSP instructions	16 kB/16 kB	220 MIPS @ 200 MHz
ARM10E	6-stage pipeline	32 kB/32 kB	
ARM11	9-stage pipeline	Variable	740 MIPS @ 665 MHz
Cortex	13-stage superscalar pipeline	Variable	2000 MIPS @ 1 GHz
XScale	Applications processor; 7-stage pipeline	32 kB/32 kB L1 512 kB L2	1000 MIPS @ 1.25 GHz

DSP = digital signal processor
SoC = system on a chip

Apple A series

A9	APL0898		14 nm FinFET (Samsung)[102]	96 mm²[103]	>2 billion	ARMv8-A	1.85 GHz dual-core Twister[104][105]	L1i: 64 KB L1d: 64 KB L2: 3 MB L3: 4 MB (Victim)[94][106]	PowerVR GT7600 (hexa-core) @ 450 MHz (172.8 GFLOPS)[107]	64-bit Single-channel 1600 MHz LPDDR4-3200[105][106] (25.6 GB/sec).[105]
	APL1022		16 nm FinFET (TSMC)[103]	104.5 mm²[103]						
A9X	APL1021		16 nm FinFET (TSMC)[108]	147 mm²[108]	>3 billion	ARMv8-A	2.16–2.26 GHz dual-core Twister[109][110]	L1i: 64 KB L1d: 64 KB L2: 3 MB L3: *none*[94][108]	PowerVR 7XT Series (dodeca-core custom design)[108](345.6 GFLOPS)	64-bit Dual-channel 1600 MHz LPDDR4-3200 (51.2 GB/sec).
A10 Fusion	APL1W24		16 nm FinFET (TSMC)[111]	125 mm²[111]	3.3 billion	ARMv8-A	2.34 GHz quad-core Hurricane (2+2 Fusion)[112]	L1i: 64 KB L1d: 64 KB L2: 3 MB L3: 4 MB	Hexa-core	LPDDR4

S Series for Apple Watches
W Series for Apple Air Pods

- **Apple A4** — introduced in the original iPad and also used in the iPhone 4, iPod touch (4th generation), and Apple TV (2nd generation).
- **Apple A5** — introduced in the iPad 2 and used in the iPhone 4S, iPod touch (5th generation) and 1st generation iPad mini.
- **Apple A5X** — introduced in the iPad (3rd generation).
- **Apple A6** — introduced in the iPhone 5 and later used in the iPhone 5C; contained a custom CPU designed internally at Apple (called "Swift") instead of one licensed from ARM.[10]
- **Apple A6X** — introduced in the iPad (4th generation).
- **Apple A7** — introduced in the iPhone 5S, the company's first 64-bit mobile processor. Also used in the 2nd and 3rd generation iPad minis and 1st generation iPad Air.[1]
- **Apple A8** — introduced in the iPhone 6 and 6 Plus; also used by the iPad mini 4, 6th generation iPod touch, and the HomePod.
- **Apple A8X** — introduced in the iPad Air 2.
- **Apple A9** — introduced in the iPhone 6S and 6S Plus, and later used in the 1st-generation iPhone SE and the 5th-generation iPad.
- **Apple A9X** — introduced in the 1st-generation iPad Pro, the company's first "desktop class" processor for mobile devices.[1]
- **Apple A10 Fusion** — introduced in the iPhone 7 and 7 Plus, and later used in the 6th and 7th-generation iPad. This was the first processor produced for Apple solely by TSMC.[9]
- **Apple A10X Fusion** — introduced in the 2nd-generation iPad Pro, and also used in the 1st-generation Apple TV 4K.
- **Apple A11 Bionic** — introduced in the iPhone 8, 8 Plus, and iPhone X. First implementation of Neural Engine machine learning.
- **Apple A12 Bionic** — introduced in the iPhone XR, XS, and XS Max, and also used in the 3rd-generation iPad Air, 5th-generation iPad mini and the 2nd-generation Apple TV 4K, and later used in the 8th-generation iPad.
- **Apple A12X Bionic** — introduced in the 3rd-generation iPad Pro.
- **Apple A12Z Bionic** — introduced in the 4th-generation iPad Pro, and used in the 2020 Mac mini-based Developer Transition Kit.
- **Apple A13 Bionic** — introduced in the iPhone 11, 11 Pro, 11 Pro Max; also used in the 2nd-generation iPhone SE, and later used in the 9th-generation iPad.
- **Apple A14 Bionic** — introduced in the 4th-generation iPad Air and also used in the iPhone 12 series.[11][12]
- **Apple A15 Bionic** — introduced in the iPhone 13 series and the 6th-generation iPad mini;[13] also adopted by the 3rd-generation iPhone SE in March 2022.[14]

Apple A5X processor

Apple A6X processor

Apple A12 processor

Videos

- https://www.youtube.com/watch?v=X4BxUiqWq8E
- https://www.youtube.com/watch?v=5IizbFrqN9g

Processor Companies

- IBM – Mainframe computers
- DEC – PDP – Minicomputers
- Intel – Microprocessor
- Hewlett-Packard
- Dell
- Sun -> Now known as Oracle
- AMD
- Nvidia
- Motorola
- Samsung
- Qualcomm
- Media Tek
- Apple