Computer Architecture
Lecture 1

Peter Mikulecký

peter.mikulecky@uhk.cz

Room 309

Introductory Remarks

- Lectures – Every Tuesday from 13:15 in J17
- Resources – Recommended by the Lecturer
- Two projects:
  - Introductory project
  - Final project
- Fulfilling both successfully means to pass the course

Introductory project

- Look for relevant resources on the Web, e.g., courses, videos, on-line books, etc.
  - Keywords: computer architecture, computer organization, history of computing, units of a computer, etc.
- Evaluate the resources according to their relevance to our course
- Create a Web page (or PowerPoint presentation) presenting the resources found with a short annotation about the particular resource.

Final project

- Everyone will be assigned by a particular theme at the end of March
- A paper (about 15 – 20 pages) plus a presentation (about 15 slides) will be required in order to pass the course successfully.
- The last lecture will be on May 7 (probably).

Some recommended resources

- http://www.karbosguide.com/books/pcarchitecture/chapter00.htm
- http://en.wikipedia.org/wiki/Computer_architecture
- http://en.wikipedia.org/wiki/Computer_hardware
- http://www.cs.iastate.edu/~prabhu/Tutorial/title.html
- http://www.ece.cmu.edu/~ece447/s13/doku.php?id=schedule
- https://www.youtube.com/watch?v=HEjPop-aK_w
- http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-823-computer-system-architecture-fall-2005/lecture-notes/

Some recommended resources

- http://computer.howstuffworks.com/
- http://www.computerhope.com/jargon/h/hardware.htm
- http://pcsupport.about.com/od/termshm/g/hardware.htm
- http://openbookproject.net/courses/intro2ict/hardware/internal.html
- http://www.coolnerds.com/newbies/hardware/hardware.htm
- http://www.computerhelpatoz.com/hardware.html

Computer

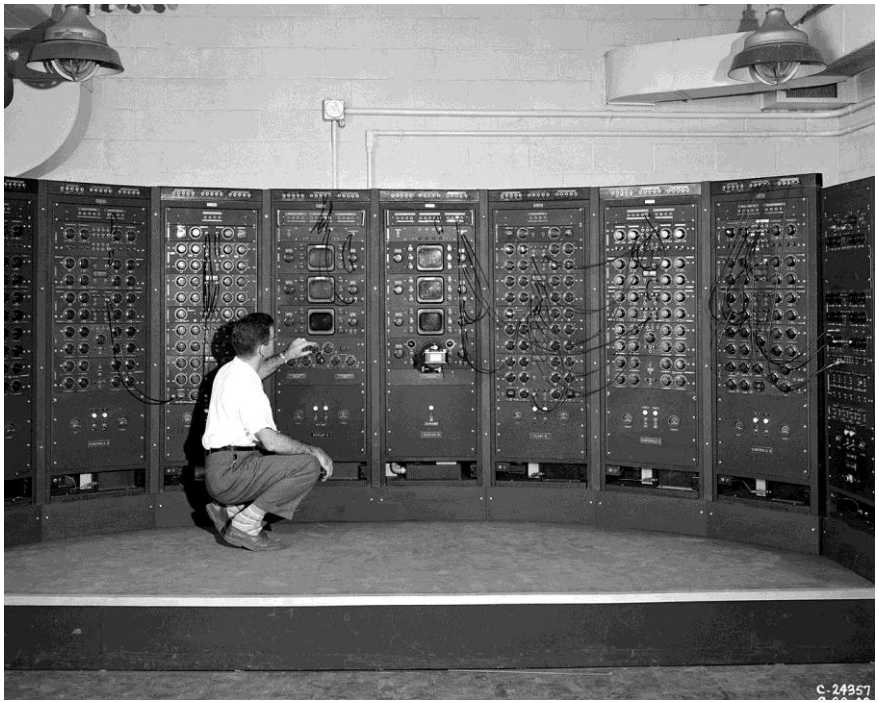
- A computer is a general purpose device that can be programmed to carry out a set of arithmetic or logical operations automatically. Since a sequence of operations can be readily changed, the computer can solve more than one kind of problem.
- Conventionally, a computer consists of at least one processing element, typically a **central processing unit** (CPU), and some form of **memory**. The processing element carries out arithmetic and logic operations, and a sequencing and control unit can change the order of operations in response to stored information. **Peripheral devices** allow information to be retrieved from an external source, and the result of operations saved and retrieved.

Analog and digital computers

- Mechanical analog computers started appearing in first century and were later used in the medieval era for astronomical calculations.
- In World War II, mechanical analog computers were used for specialized military applications.
- During this time the first electronic digital computers were developed. Originally they were the size of a large room, consuming as much power as several hundred modern personal computers (PCs)

Analog computers

- An analog computer is a form of computer that uses the continuously changeable aspects of physical phenomena such as electrical, mechanical, or hydraulic quantities to model the problem being solved.
- Analog computers were widely used in scientific and industrial applications where digital computers of the time lacked sufficient performance.
- The advent of digital computing and its success made analog computers largely obsolete in 1950s and 1960s, though they remain in use in some specific applications.



Analog computing machine at the Lewis Flight Propulsion Laboratory in 1949.

Modern computers

- Modern computers based on integrated circuits are millions to billions of times more capable than the early machines, and occupy a fraction of the space.
- Simple computers are small enough to fit into **mobile devices**, and mobile computers can be powered by small batteries.
- **Personal computers** in their various forms are icons of the Information Age and are what most people think of as "computers." However, the **embedded computers** found in many devices from MP3 players to fighter aircraft and from toys to industrial robots are the most numerous.

Computer generations - 1

- **First generation**: (mechanical/electromechanical)
  - Calculators
    - Pascal's calculator, Arithmometer, Difference engine, Quevedo's analytical machines
  - Programmable devices
    - Jacquard loom, Analytical engine, IBM ASCC/Harvard Mark I, Harvard Mark II, IBM SSEC, Z1, Z2, Z3

Computer generations - 2

- Second generation (vacuum tubes)
  - Calculators
    - Atanasoff–Berry Computer, IBM 604, UNIVAC 60, UNIVAC 120
  - Programmable devices
    - Colossus, ENIAC, Manchester Small-Scale Experimental Machine, EDSAC, Manchester Mark 1, Ferranti Pegasus, Ferranti Mercury, CSIRAC, EDVAC, UNIVAC I, IBM 701, IBM 702, IBM 650, Z22

Computer generations - 3

- Third generation (discrete transistors and SSI, MSI, LSI integrated circuits)
  - Mainframes
    - IBM 7090, IBM 7080, IBM System/360, BUNCH
  - Minicomputers
    - PDP-8, PDP-11, IBM System/32, IBM System/36

Mainframes

- Mainframe computers (colloquially referred to as "big iron") are computers used primarily by corporate and governmental organizations for critical applications, bulk data processing such as census, industry and consumer statistics, enterprise resource planning and transaction processing.
- The term originally referred to the large cabinets called "main frames" that housed the central processing unit and main memory of early computers. Later, the term was used to distinguish high-end commercial machines from less powerful units. Most large-scale computer system architectures were established in the 1960s, but continue to evolve.

Mainframes (contd..)

- Modern mainframe design is generally less defined by single-task computational speed (typically defined as MIPS rate or FLOPS in the case of floating point calculations), and more by:
  - Redundant internal engineering resulting in high reliability and security
  - Extensive input-output facilities with the ability to offload to separate engines
  - Strict backward compatibility with older software
  - High hardware and computational utilization rates through virtualization to support massive throughput.
- Their high stability and reliability enables these machines to run uninterrupted for decades.

Mainframes (contd..)

An IBM 704 mainframe (1964)

Minicomputers

- A minicomputer, or colloquially mini, is a class of smaller computers that developed in the mid-1960s and sold for much less than mainframe and mid-size computers from IBM and its direct competitors.
- In a 1970 survey, the New York Times suggested a consensus definition of a minicomputer as a machine costing less than 25,000 USD, with an input-output device such as a teleprinter and at least four thousand words of memory, that is capable of running programs in a higher level language, such as Fortran or BASIC.
- The class formed a distinct group with its own software architectures and operating systems. Minis were designed for control, instrumentation, human interaction, and communication switching as distinct from calculation and record keeping.
- During the two decade lifetime of the minicomputer class (1965-1985), almost 100 companies formed and only a half dozen remained.

Minicomputers (contd..)

Computer generations - 4

- Fourth generation (VLSI integrated circuits)
  - Minicomputers
    - VAX, IBM System i
  - 4-bit microcomputer
    - Intel 4004, Intel 4040
  - 8-bit microcomputer
    - Intel 8008, Intel 8080, Motorola 6800, Motorola 6809, MOS Technology 6502, Zilog Z80

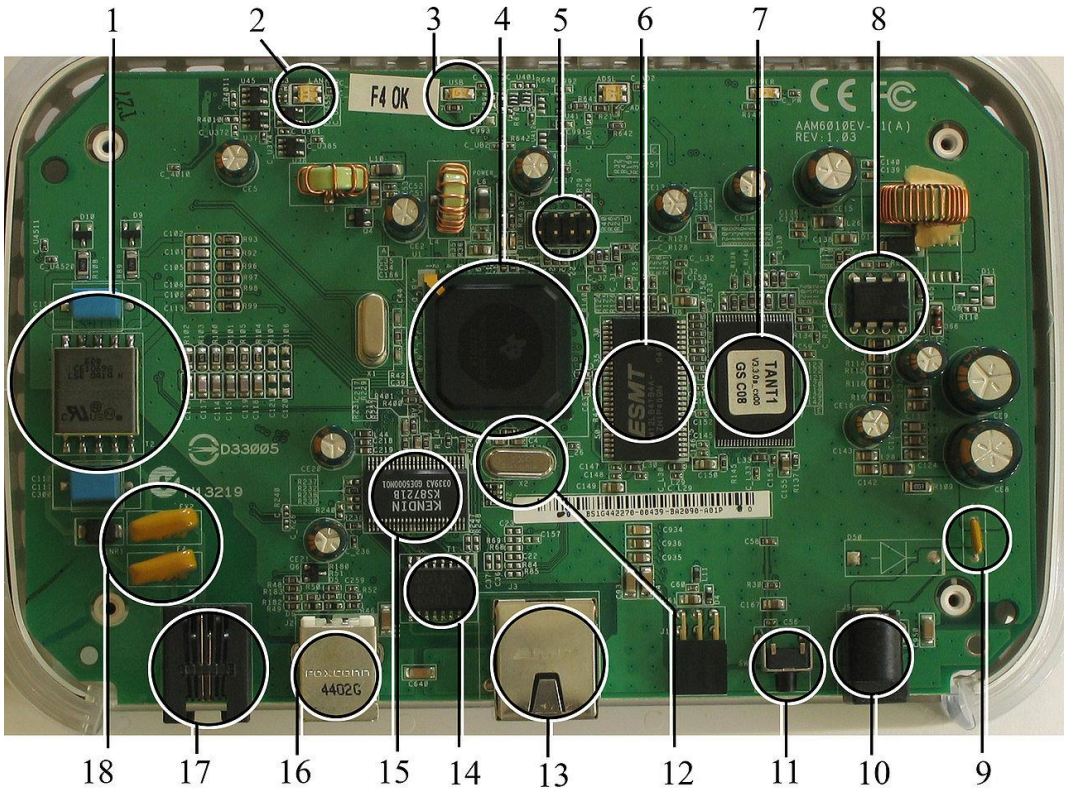
Computer generations – 4 (contd..)

- 16-bit microcomputer
  - Intel 8088, Zilog Z8000, WDC 65816/65802
- 32-bit microcomputer
  - Intel 80386, Pentium, Motorola 68000, ARM
- 64-bit microcomputer
  - Alpha, MIPS, PA-RISC, PowerPC, SPARC,
  - x86-64, ARMv8-A
- Embedded computers
  - Intel 8048, Intel 8051

Embedded systems

- An embedded system is a computer system with a dedicated function within a larger mechanical or electrical system, often with real-time computing constraints. It is embedded as part of a complete device often including hardware and mechanical parts. Embedded systems control many devices in common use today.
- Properties typical of embedded computers when compared with general-purpose ones are e.g. low power consumption, small size, rugged operating ranges and low per-unit cost.
- Embedded systems range from portable devices such as digital watches and MP3 players, to large stationary installations like traffic lights, factory controllers, and largely complex systems like hybrid vehicles, MRI, and avionics.
- Complexity varies from low, with a single microcontroller chip, to very high with multiple units, peripherals and networks mounted inside a large chassis or enclosure.

Embedded systems (contd..)



Picture of the internals of an ADSL modem/router. A modern example of an embedded system. Labelled parts include a microprocessor (4), RAM (6), and flash memory (7).

Computer generations – 4 (contd..)

- Personal computers
  - Desktop computer,
  - Home computer,
  - Laptop computer,
  - Personal digital assistant (PDA),
  - Portable computer,
  - Tablet PC,
  - Wearable computer

Desktop computer

- A desktop computer is a personal computer in a form intended for regular use at a single location desk/table due to its size and power requirements, as opposed to a laptop whose rechargeable battery and compact dimensions allow it to be regularly carried and used in different locations.
- The most common configuration is a computer monitor, keyboard and mouse, and a case that houses the main components of the PC, namely the power supply, motherboard, hard drive, optical drive, and previously also the floppy drive.

All-in-one desktop computers

- An all-in-one desktop computer integrates the system's internal components into the same case as the display, eliminating some connecting cables and allowing for a smaller footprint, sometimes giving a degree of portability, compared to the standard desktop configuration of the separate display monitor and computer system case.
- However the all-in-one form factor still requires an external power supply and must be deployed on a table or desk to use the (still separate) keyboard and mouse, making them less mobile than a laptop which can rely on power supplied by a rechargeable battery and provides a built-in keyboard plus pointing device for its user.

All-in-one desktop computers

Some examples of modern all-in-ones

Home computers

- Home computers were a class of microcomputers entering the market in 1977, and becoming common during the 1980s. They were marketed to consumers as affordable and accessible computers that, for the first time, were intended for the use of a single nontechnical user.
- These computers were a distinct market segment that typically cost much less than business, scientific or engineering-oriented computers of the time such as the IBM PC, and were generally less powerful in terms of memory and expandability.
- However, a home computer often had better graphics and sound than contemporary business computers. Their most common use was playing video games.

Laptop computers

- A laptop or a notebook is a portable personal computer with a clamshell form factor, suitable for mobile use.
- There was a difference between laptops and notebooks in the past, but nowadays it has gradually died away.
- Laptops are commonly used in a variety of settings, including at work, in education, and for personal multimedia.
- A laptop combines the components and inputs of a desktop computer, including display, speakers, keyboard and pointing device (such as a touchpad or a trackpad) into a single device. Most modern-day laptops also have an integrated webcam and a microphone. A laptop can be powered either from a rechargeable battery, or by mains electricity via an AC adapter.
- Laptop is a diverse category of devices and other more specific terms, such as rugged notebook or convertible, refer to specialist types of laptops, which have been optimized for specific uses.

Laptop examples

Personal Digital Assistant (PDA)

- A personal digital assistant (PDA), also known as a handheld PC, or personal data assistant, is a mobile device that functions as a personal information manager.
- The term evolved from Personal Desktop Assistant, a software term for an application that prompts or prods the user of a computer with suggestions or provides quick reference to contacts and other lists.
- PDAs were discontinued in early 2010s after the widespread adoption of smartphones.

Portable computers

- A portable computer is a computer that is designed to be moved from one place to another and includes a display and keyboard. Portable computers, by their nature, are generally microcomputers. Portable computers, because of their size, are also commonly known as 'Lunchbox' or 'Luggable' computers. They can also be called a 'Portable Workstation' or 'Portable PC'.
- These are distinct from desktop replacement computers in that they are usually constructed from full-specification desktop components, and do not incorporate features associated with laptops or mobile devices.
- The principal advantage of a portable computer versus a laptop or other mobile computing device is the use of standard motherboards or backplanes providing plug-in slots for add-in cards. This allows mission specific cards such as test, A/D, or communication protocol (IEEE-488, 1553) to be installed. Portable computers also provide for more disk storage by using standard 3-1/2" drives and providing for multiple drives.

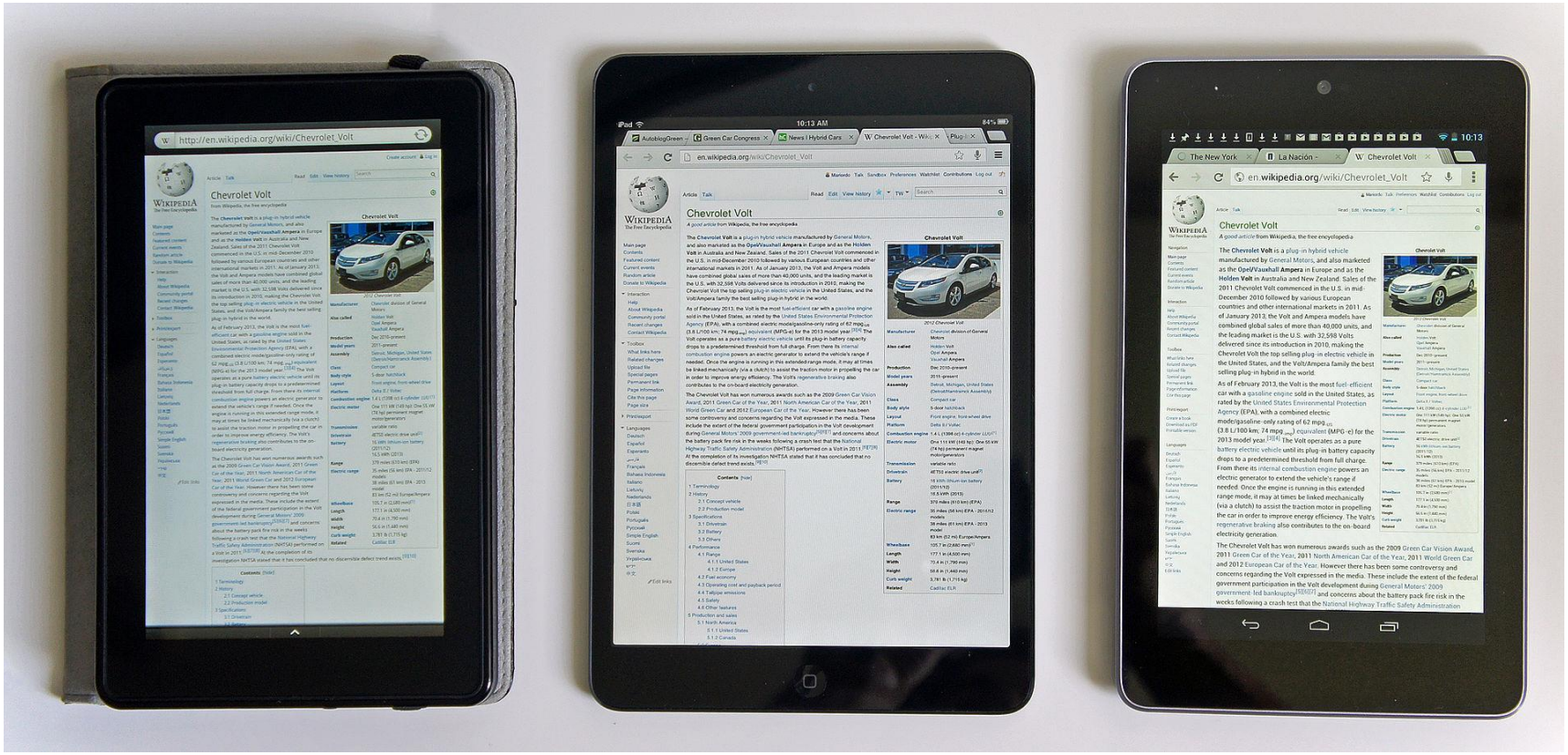
Portable computers

Contemporary portable computer with 1 20.1" LCD screen, EATX motherboard.

Tablet PC

- A tablet computer is a mobile computer with touch-screen display, circuitry and battery in a single unit. Tablets come equipped with sensors, including cameras, a microphone, an accelerometer and a touchscreen, with finger or stylus gestures substituting for the use of computer mouse and keyboard.
- Tablets may include physical buttons (for example: to control basic features such as speaker volume and power) and ports (for network communications and to charge the battery). They usually feature on-screen, pop-up virtual keyboards for typing.
- Tablets are typically larger than smartphones or personal digital assistants at 7 inches (18 cm) or larger, measured diagonally.
- One can classify tablets into several categories according to the presence and physical appearance of keyboards. Slates and booklets do not have a physical keyboard and typically feature text input performed through the use of a virtual keyboard projected on a touchscreen-enabled display. Hybrids and convertibles do have physical keyboards, although these devices typically also make virtual keyboards available.

Tablet examples

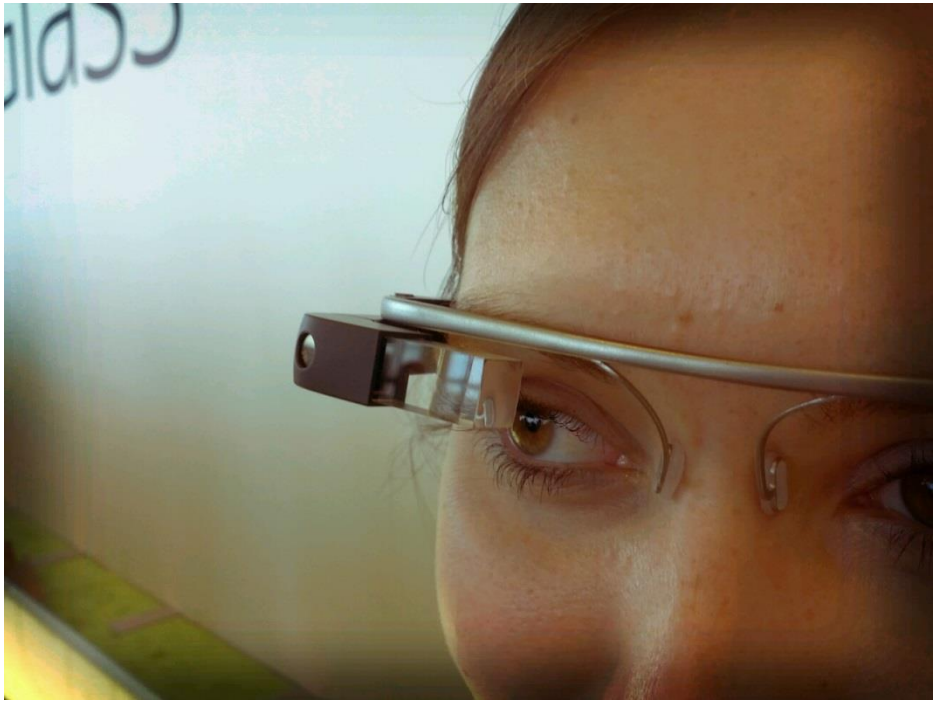


Comparison of several mini tablet computers: Amazon Kindle Fire (left), iPad Mini (center) and Google Nexus 7 (right).

Wearable computers

- Wearable computers, also known as body-borne computers or wearables are miniature electronic devices that are worn by the bearer under, with or on top of clothing. This class of wearable technology has been developed for general or special purpose information technologies and media development. Wearable computers are especially useful for applications that require more complex computational support than just hardware coded logics.
- If one is asked to give a simple, yet modern, example for wearable technology, that will be the Nike+ system which allows you to track your time, distance, pace and calories via a sensor in the shoe. Another example can be Google Glass, which combine innovative displays with some novel gestural movements for interaction.
- One of the main features of a wearable computer is consistency. There is a constant interaction between the computer and user, i.e. there is no need to turn the device on or off. Another feature is the ability to multi-task. It is not necessary to stop what you are doing to use the device; it is augmented into all other actions. These devices can be incorporated by the user to act like a prosthetic. It can therefore be an extension of the user's mind and/or body.

Examples



- Google Glass, Google's head-mounted display, which was launched in 2013 (right); the WIMM One, wearable computer (smartwatch) powered by Android (left).

Smartwatch

- A smartwatch (or smart watch) is a computerized wristwatch with functionality that is enhanced beyond timekeeping. While early models can perform basic tasks, such as calculations, translations, and game-playing, modern smartwatches are effectively wearable computers. Many smartwatches run mobile apps, while a smaller number of models run a mobile operating system and function as portable media players, offering playback of FM radio, audio, and video files to the user via a Bluetooth headset. Some smartwatches models, also called watch phones, feature full mobile phone capability, and can make or answer phone calls.
- Such devices may include features such as a camera, accelerometer, thermometer, altimeter, barometer, compass, chronograph, calculator, cell phone, touch screen, GPS navigation, Map display, graphical display, speaker, scheduler, watch, SDcards that are recognized as a mass storage device by a computer, and rechargeable battery. It may communicate with a wireless headset, heads-up display, insulin pump, microphone, modem, or other devices.

Theoretical/experimental computers

- Quantum computer,
- Chemical computer,
- DNA computing,
- Optical computer,
- Spintronics based computer

Quantum computing

- Quantum computing studies theoretical computation systems (quantum computers) that make direct use of quantum-mechanical phenomena, such as superposition and entanglement, to perform operations on data. Quantum computers are different from digital computers based on transistors. Whereas digital computers require data to be encoded into binary digits (bits), each of which is always in one of two definite states (0 or 1), quantum computation uses qubits (quantum bits), which can be in superpositions of states.
- A quantum Turing machine is a theoretical model of such a computer, and is the also known as the universal quantum computer. Quantum computers share theoretical similarities with non-deterministic and probabilistic computers.
- The field of quantum computing was first introduced by Yuri Manin in 1980, and Richard Feynman in 1982. A quantum computer with spins as quantum bits was also formulated for use as a quantum space–time in 1968.

Chemical computers

- A chemical computer, also called reaction-diffusion computer, BZ computer (stands for Belousov–Zhabotinsky computer) or gooware computer is an unconventional computer based on a semi-solid chemical "soup" where data are represented by varying concentrations of chemicals. The computations are performed by naturally occurring chemical reactions. So far it is still in a very early experimental stage, but may have great potential for the computer industry.
- The simplicity of this technology is one of the main reasons why it could in the future be turned into a serious competitor to machines based on conventional hardware. A modern microprocessor is an incredibly complicated device that can be destroyed during production by no more than a single airborne microscopic particle.
- In a BZ solution the waves are moving in all thinkable directions in all dimensions, across, away and against each other. These properties might make a chemical computer able to handle billions of times more data than a traditional computer.

DNA computing

- DNA computing is a branch of computing which uses DNA, biochemistry and molecular biology hardware, instead of the traditional silicon-based computer technologies.
- DNA computing, or, more generally, biomolecular computing, is a fast-developing interdisciplinary area. Research and development in this area concerns theory, experiments, and applications of DNA computing.
- This field was initially developed by Leonard Adleman of the University of Southern California, in 1994. Adleman demonstrated a proof-of-concept use of DNA as a form of computation which solved the seven-point Hamiltonian path problem. Since the initial Adleman experiments, advances have been made and various Turing machines have been proven to be constructible.
- DNA computing is a form of parallel computing in that it takes advantage of the many different molecules of DNA to try many different possibilities at once.

Optical computing

- Optical or photonic computing uses photons produced by lasers or diodes for computation. For decades, photons have promised to allow a higher bandwidth than the electrons used in conventional computers.
- Most research projects focus on replacing current computer components with optical equivalents, resulting in an optical digital computer system processing binary data. This approach appears to offer the best short-term prospects for commercial optical computing, since optical components could be integrated into traditional computers to produce an optical-electronic hybrid.
- However, optoelectronic devices lose 30% of their energy converting electrons into photons and back. This also slows transmission of messages. All-optical computers eliminate the need for optical-electrical-optical (OEO) conversions.

Spintronics based computers

- Spintronics (a blended word meaning "spin transport electronics"), also known as spinelectronics or fluxtronic, is an emerging technology exploiting both the intrinsic spin of the electron and its associated magnetic moment, in addition to its fundamental electronic charge, in solid-state devices.
- Spintronics differs from the older magnetoelectronics, in that the spins are not only manipulated by magnetic fields, but also by electrical fields.
- Non-volatile spin-logic devices to enable scaling beyond the year 2025 are being extensively studied. Spin-transfer torque-based logic devices that use spins and magnets for information processing have been proposed and are being extensively studied at Intel. These devices are now part of the ITRS exploratory road map and have potential for inclusion in future computers. Logic-in memory applications are already in the development stage at Crocus and NEC.

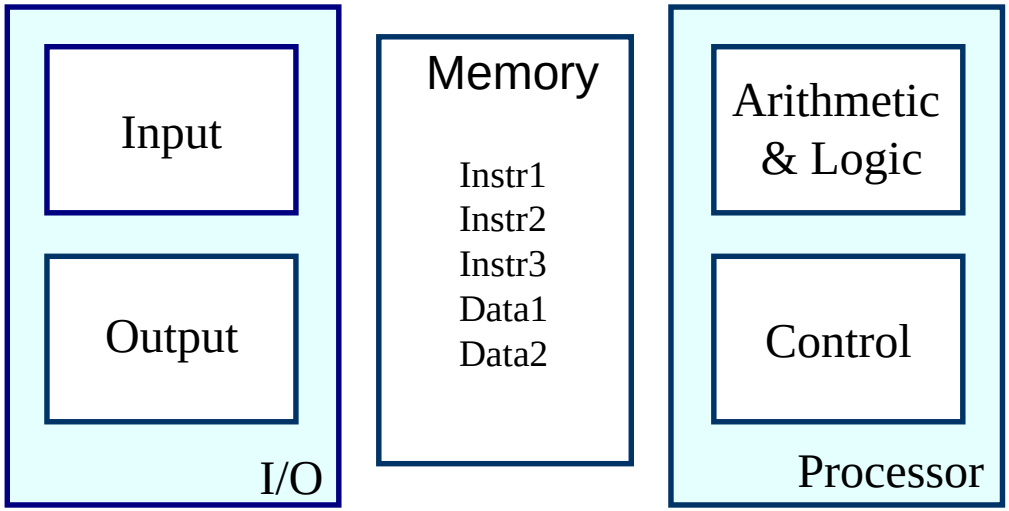
So, what is a computer?

- A computer is a sophisticated electronic calculating machine that:
  - Accepts input information,
  - Processes the information according to a list of internally stored instructions and
  - Produces the resulting output information.
- Functions performed by a computer are:
  - Accepting information to be processed as input.
  - Storing a list of instructions to process the information.
  - Processing the information according to the list of instructions.
  - Providing the results of the processing as output.
- What are the functional units of a computer?

Functional units of a computer

Input unit accepts information:
- Human operators,
- Electromechanical devices
- Other computers

Arithmetic and logic unit(ALU):
- Performs the desired operations on the input information as determined by instructions in the memory



Output unit sends results of processing:
- To a monitor display,
- To a printer

Stores information:
- Instructions,
- Data

Control unit coordinates various actions
- Input,
- Output
- Processing

Information in a computer: Instructions

- Instructions specify commands to:
  - Transfer information within a computer (e.g., from memory to ALU)
  - Transfer of information between the computer and I/O devices (e.g., from keyboard to computer, or computer to printer)
  - Perform arithmetic and logic operations (e.g., Add two numbers, Perform a logical AND).
- A sequence of instructions to perform a task is called a program, which is stored in the memory.
- Processor fetches instructions that make up a program from the memory and performs the operations stated in those instructions.
- What do the instructions operate upon?

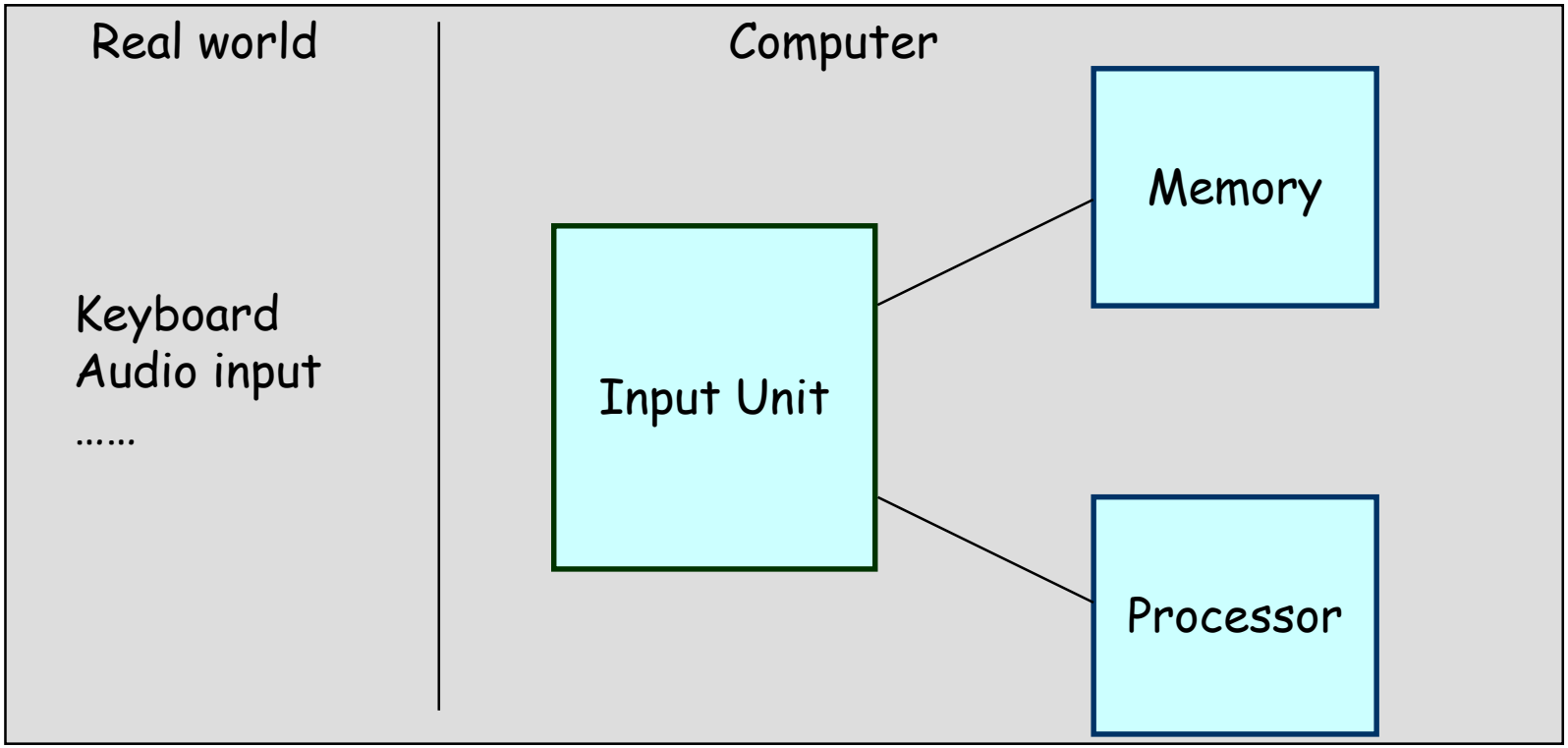
Information in a computer: Data

- Data are the “operands” upon which instructions operate.
- Data could be:
  - Numbers,
  - Encoded characters.
- Data, in a broad sense means any digital information.
- Computers use data that is encoded as a string of binary digits called bits.

Input unit

Binary information must be presented to a computer in a specific format. This task is performed by the input unit:

- Interfaces with input devices.
- Accepts binary information from the input devices.
- Presents this binary information in a format expected by the computer.
- Transfers this information to the memory or processor.



Memory unit

- Memory unit stores instructions and data.
  - Recall, data is represented as a series of bits.
  - To store data, memory unit thus stores bits.
- Processor reads instructions and reads/writes data from/to the memory during the execution of a program.
  - In theory, instructions and data could be fetched one bit at a time.
  - In practice, a group of bits is fetched at a time.
  - Group of bits stored or retrieved at a time is termed as “word”
  - Number of bits in a word is termed as the “word length” of a computer.
- In order to read/write to and from memory, a processor should know where to look:
  - “Address” is associated with each word location.

Memory unit (contd..)

- Processor reads/writes to/from memory based on the memory address:
  - Access any word location in a short and fixed amount of time based on the address.
  - Random Access Memory (RAM) provides fixed access time independent of the location of the word.
  - Access time is known as "Memory Access Time".
- Memory and processor have to "communicate" with each other in order to read/write information.
  - In order to reduce "communication time", a small amount of RAM (known as Cache) is tightly coupled with the processor.
- Modern computers have three to four levels of RAM units with different speeds and sizes:
  - Fastest, smallest known as Cache
  - Slowest, largest known as Main memory.

Memory unit (contd..)

- Primary storage of the computer consists of RAM units.
  - Fastest, smallest unit is Cache.
  - Slowest, largest unit is Main Memory.
- Primary storage is insufficient to store large amounts of data and programs.
  - Primary storage can be added, but it is expensive.
- Store large amounts of data on secondary storage devices:
  - Magnetic disks and tapes,
  - Optical disks (CD-ROMS).
  - Access to the data stored in secondary storage in slower, but take advantage of the fact that some information may be accessed infrequently.
- Cost of a memory unit depends on its access time, lesser access time implies higher cost.

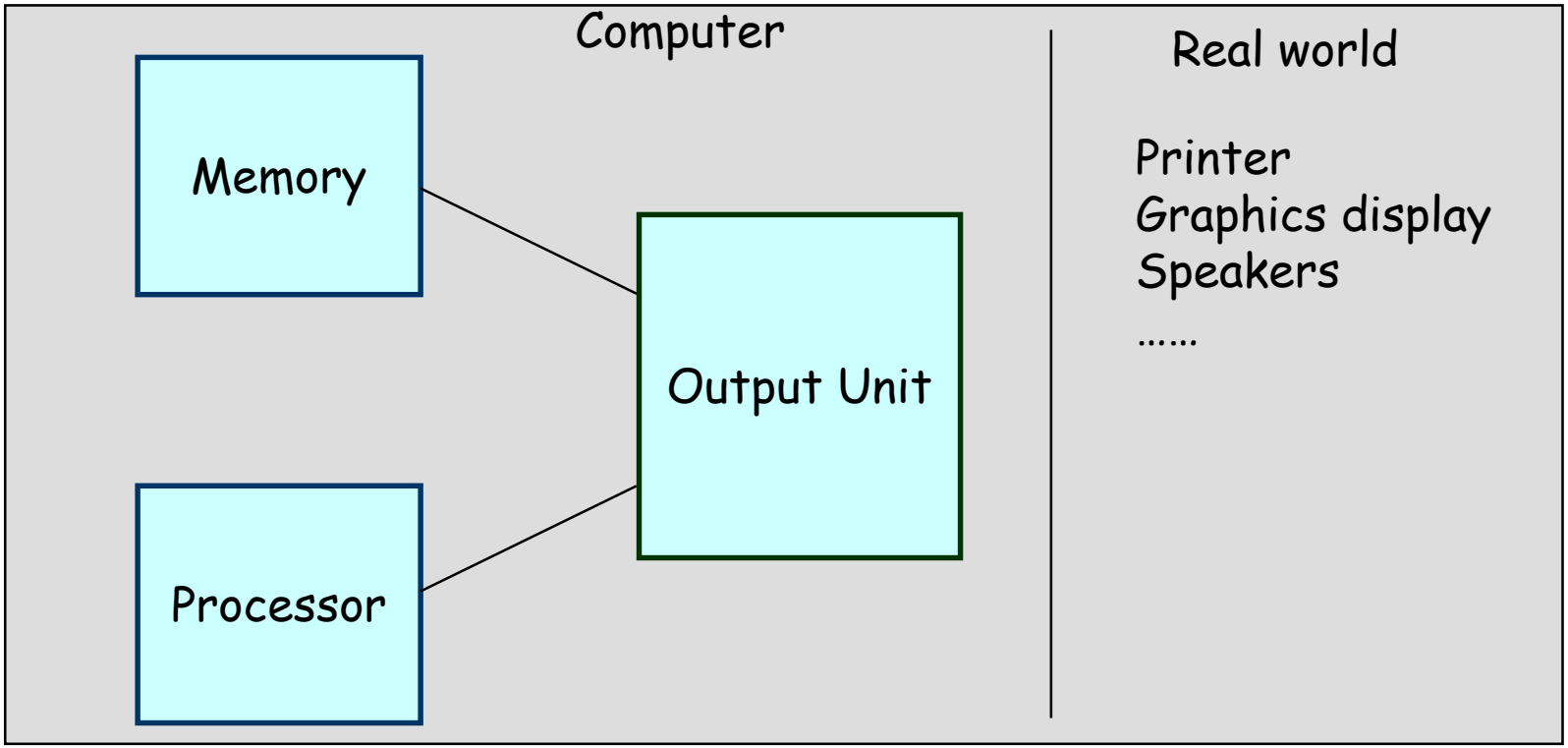
Arithmetic and logic unit (ALU)

- Operations are executed in the Arithmetic and Logic Unit (ALU).
  - Arithmetic operations such as addition, subtraction.
  - Logic operations such as comparison of numbers.
- In order to execute an instruction, operands need to be brought into the ALU from the memory.
  - Operands are stored in general purpose registers available in the ALU.
  - Access times of general purpose registers are faster than the cache.
- Results of the operations are stored back in the memory or retained in the processor for immediate use.

Output unit

Computers represent information in a specific binary form. Output units:

- Interface with output devices.
- Accept processed results provided by the computer in specific binary form.
- Convert the information in binary form to a form understood by an output device.

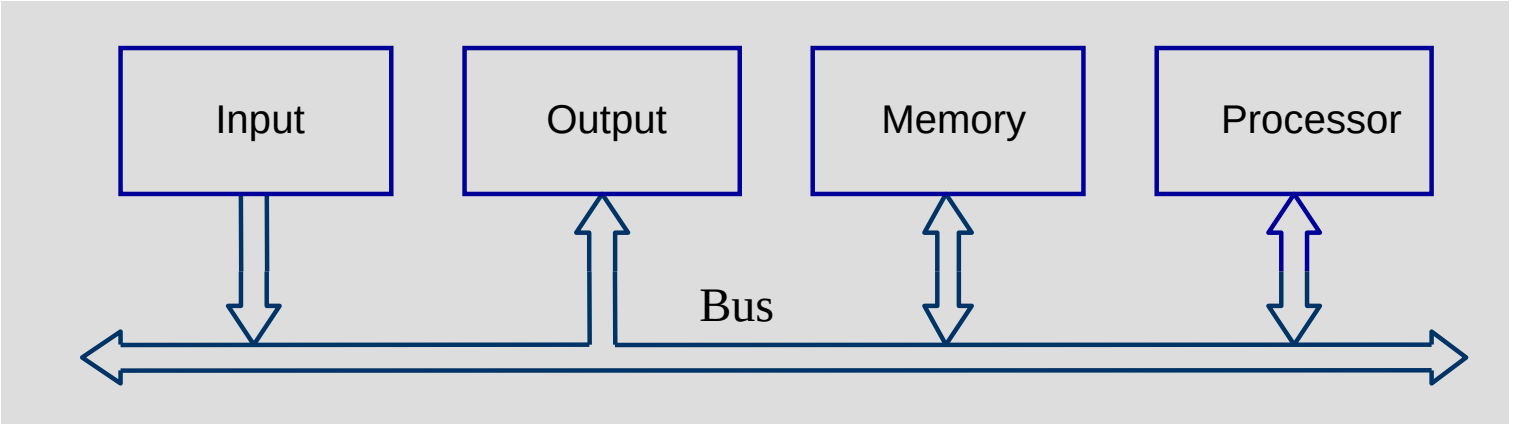


Control unit

- Operation of a computer can be summarized as:
  - Accepts information from the input units (Input unit).
  - Stores the information (Memory).
  - Processes the information (ALU).
  - Provides processed results through the output units (Output unit).
- Operations of Input unit, Memory, ALU and Output unit are coordinated by Control unit.
- Instructions control “what” operations take place (e.g. data transfer, processing).
- Control unit generates timing signals which determines “when” a particular operation takes place.

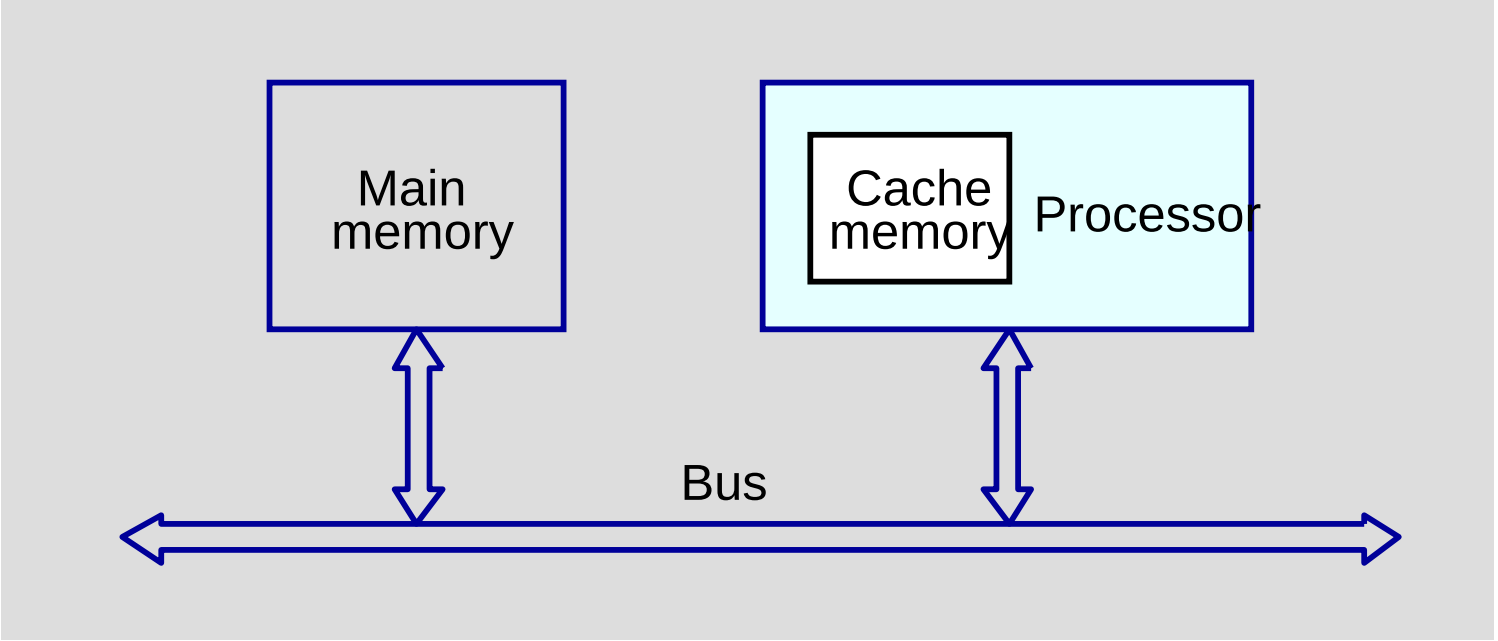
How are the functional units connected?

For a computer to achieve its operation, the functional units need to communicate with each other.
In order to communicate, they need to be connected.



Functional units may be connected by a group of parallel wires.
The group of parallel wires is called a bus.
Each wire in a bus can transfer one bit of information.
The number of parallel wires in a bus is equal to the word length of a computer.

Organization of cache and main memory



Why is the access time of the cache memory lesser than the access time of the main memory?