Moore's Law

Aniel Nieves-González

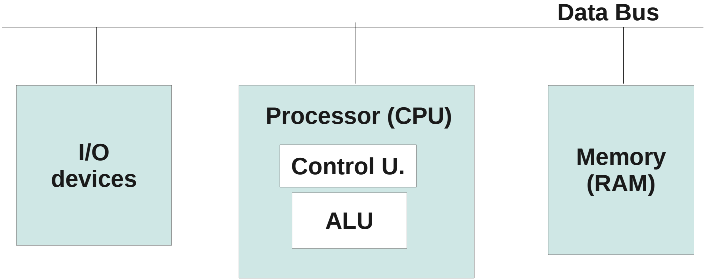
Institute of Statistics

Fall 2015

A bit of history I

- In the 1930's Alan Turing and Alonzo Church lay the foundation of the modern theory of algorithms and computer science.
- In the 1940's John von Neumann proposed a *computer architecture* that is still used today. From this architecture stem the idea that data and instruction are stored in memory.
- Another fundamental idea in von Neumann is the sequential execution of instructions: Fetch the instruction from memory and the central processing unit decodes and executes.

A bit of history II



- In the late 1940's Bardeen, Brattain, and Shockley developed the transistor. The development of the transistor fueled most of the subsequent improvements in hardware. From that point onwards computers grew in computing power and shrank in physical size.

What is a transistor? I

First, recall that:

- Data and instructions in a computer are represented using the binary number system.
- The construction of computer circuits is based upon a branch of mathematics called Boolean algebra.
- Boolean algebra consists of the set $\{0, 1\}$ with the operations complement ($'$), sum ($+$), and product ($\cdot$).
- We can rephrase the previous definition as: the set $\{\text{FALSE}, \text{TRUE}\}$ with the operations of negation (NOT, $\sim$), disjunction (OR, $\wedge$), and conjunction (AND, $\vee$).
- A Boolean variable is one that assumes one of the two aforementioned possible values.

What is a transistor? II

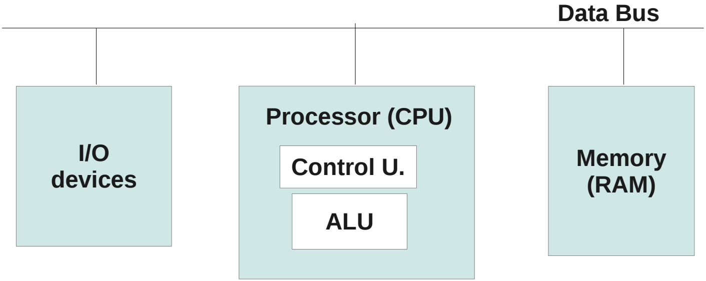
- A *logical gate* is an electronic device that operates on a collection of binary inputs and produces a binary output. That is, it maps from the set $\{0, 1\}$ (or a set of n-tuples $\{0, 1\} \times \ldots \times \{0, 1\}$) to $\{0, 1\}$.
- Logical gates are constructed using transistors.
- A *transistor* is akin to a switch (it can be in an OFF or in a ON state) made out of a semiconductor material.
- A semiconductor is a material with the property that under certain circumtances acts as a conductor and under other circumtances acts as an insulator. The branch of physics that study this type of material is solid state physics.
- Transistors have no moving parts and can be made to be very small.

Moore's law

- Gordon Moore proposed an idea in 1965 that became Moore's law.
- The law states that computing power will double for constant cost roughly once every two years.
- Observe that Moore's law implies exponential growth and that it refers to computing power specifically not to storage capacity etc.
- Many expect Moore's law to hit a brickwall as the size of transistors get closer to the size of atoms.
- Similar exponential growth has been observed in storage capacity.

A bit of computer architecture...

Recall the figure...



A bit of computer architecture... I

1. The CPU.
   - *Central Processing Unit* (CPU): it decodes and executes instructions given in a computer program. It has two components that have become integrated in what is called the Processor:
     - Arithmetic logic unit (ALU): It performs arithmetic and logic operations.
     - Control Unit (CU): It fetches, decodes, and execute instructions. These functions repeat until the last instruction of the program is executed.
   - The CPU has the ALU, the CU, registers (storage units), and interconnections between them.
   - The interconnections between the components are called the bus.
2. Memory.

A bit of computer architecture... II

- Random Access Memory (RAM): it stores data and instructions.
  - It is volatile.
  - It is made of fixed-size cells, each with a unique address.
  - The time to access a cell is uniform for all cells.
- Read only memory (ROM): Similar to RAM, but it has system information recorded during manufacture (it is nonvolatile).
- Cache memory: A special high-speed memory, physically close to the CPU. The rationale for the using the cache memory is the principle of locality (when the computer uses something, it will probably use it again very soon).

3. I/O devices.
   - A very heterogeneous set of devices. It includes display, keyboard, nonvolatile storage devices, etc.
   - Nonvolatile storage devices include:

A bit of computer architecture... III

  - Hard disk (HD). It is a rotating disc(s) made of a magnetic material.
  - Flash memory. It is also a semiconductor device (transistor based).
- Typically HDs are more prone to damage and are more slower in terms of access time than flash memory.

Bits...

Recall that everything in a modern computer is represented using the binary number system. The following definitions are important:

- Bit stands for binary digit.
- Byte stands for 8 binary digits. Historically, the definition of byte was architecture dependent, later on it was standarized.

	amount
1 kilobyte (KB)	$1024 = 2^{10}$
1 megabyte (MB)	2^{20}
1 gigabyte (GB)	2^{30}
1 terabyte (TB)	2^{40}
1 petabyte (PB)	2^{50}
1 exabyte (EB)	2^{60}

Limits of Moore's law

- Moore's law is a consequence of the ablity to pack more and more transistors (the building blocks of computer cicuitry) per unit area.
- As you make circuits smaller, the distance traveled by electrons (current is flow of electrons) is shorter.
- There are physical limits to how small you can go –you cannot go beyond subatomic particles. There are limits on how fast a signal can travel, etc.
- Possible workarounds for this limit, at least for the moment, are parallel computing and quantum computing.

Limits of von Neumann architecture

- Von Neumann bottleneck: a fundamental inability of von Neumann model to solve some large scale problems. What is that fundamental inability?

Limits of von Neumann architecture

- Von Neumann bottleneck: a fundamental inability of von Neumann model to solve some large scale problems. What is that fundamental inability?
- The sequential one-instruction-at-a-time mode of execution.

Limits of von Neumann architecture

- Von Neumann bottleneck: a fundamental inability of von Neumann model to solve some large scale problems. What is that fundamental inability?
- The sequential one-instruction-at-a-time mode of execution.
- Non-von Neumann models are more in in tune with the principle: "If you cannot build something to go twice as fast, do two things at once and the results will be identical".

Limits of von Neumann architecture

- Von Neumann bottleneck: a fundamental inability of von Neumann model to solve some large scale problems. What is that fundamental inability?
- The sequential one-instruction-at-a-time mode of execution.
- Non-von Neumann models are more in in tune with the principle: "If you cannot build something to go twice as fast, do two things at once and the results will be identical".
- Parallel computing and quantum computing are two examples of non-von Neumann architectures.

Parallel computing I

1. *Parallel processing* is the simultaneous processing by more than one processing unit of a single application.
2. In parallel processing the machine has several processors rather than just one. If each processor can be occupied with meaningful work then the von Neumann bottleneck can be overcome.
3. Some authors say that a *supercomputer* is the salesman name of a kind of parallel computer.
4. There are two fundamental approaches to designing parallel processing systems:
    1. *SIMD parallel processing.* SIMD stands for single-instruction-multiple-data system.

Parallel computing II

- In this model there is a single program whose instructions are processed in a sequential fashion by one CU. However, the ALU is replicated many times and each ALU has its own private memory. The CU broadcast the instructions to the ALUs (think about the game of bingo).
- This model is useful when dealing with data structures such as matrices and operations between them.
- Within the SIMD paradigm you can have *shared* or *distributed* memory schemes.

2. *MIMD parallel processing.* MIMD stands for multiple-instruction-multiple-data system.
   - This is also called cluster computing. In MIMD parallelism entire processors rather than just the ALU is replicated. In this case every processor is capable of executing its own separate program in its own private memory at its own rate.

Parallel computing III

- The processors are connected by an interconnection network.
- Each processor is a von Neumann machine that tackles a small part of the overall problem. The results or partial results are communicated via the inerconnection network.
- MIMD follows a distributed memory scheme.
- Massive parallel MIMD systems having 1000s of processors are used to solve very complex problems (check top500 supercomputer website).
- Multiple processors within an MIMD cluster do not have to be identical or even belong to a single administrative organization. Moreover, when computing resourses sitting idle around the world are used to solve massive problems we are talking about grid computing. Examples: SETI project, Folding@home project.

Parallel computing IV

5. The key to use parallel processing is to design algorithms that effectively utilize the available processors. The branch of science that occupies itself with that is parallel algorithms.

Quantum computing I

1. Quantum computers do not follow the von Neumann architecture.
2. It is a design that uses the principles of quantum mechanics.
3. Quantum mechanics is the branch of physics that describes the behavior of matter at the atomic and subatomic scale.
4. A quantum computer encodes information using some aspect of a quantum mechanical state such as electron spin, etc.
5. In contrast of a traditional bit, a quantum bit or qubit can be 0 and 1 at the same time. Thus theoretically a quantum computer can do different computations simultaneously.

Quantum computing II

6. There are algorithms design for quantum computers that would solve very hard problems relatively easily.
7. Quantum computers are very difficult to construct.

Downside of Moore's law: E-waste I

- Electronic waste or e-waste is the discarted tech junk.
- Recycling is a solution, but it raises another issues.
- There is a disconnect between consumers and managers who want to do good and those efforts that are actually doing good
- The following points show how difficult addressing this problem will be
  - The complexities of the modern value chain.
  - The vagaries of international law.
  - The nefarious actions of those willing to put profits above principle.
- The process of recycling is extremely labor intensive.

Downside of Moore's law: E-waste II

- Disregard of ethical recycling imperatives can tarnish a brand.
- Managers must consider and plan for the waste created by their products, services, and technology used by the organization.

References I

- Galla Textbook.
- Yuefang Deng
  Applied Parallel Computing
- Donald D. Knuth.
  The art of computer programming. Volume 1: Fundamental Algorithms.
  Third edition. Addison-Wesley, 1997.
- G. Michael Schneider and Judith Gersting.
  An Invitation to Computer Science: C++ Version.
  4th Edition. Thomson: Course Technology, 2006