Imperial College London

Topic 13

Introduction to Memories and Computer Architecture

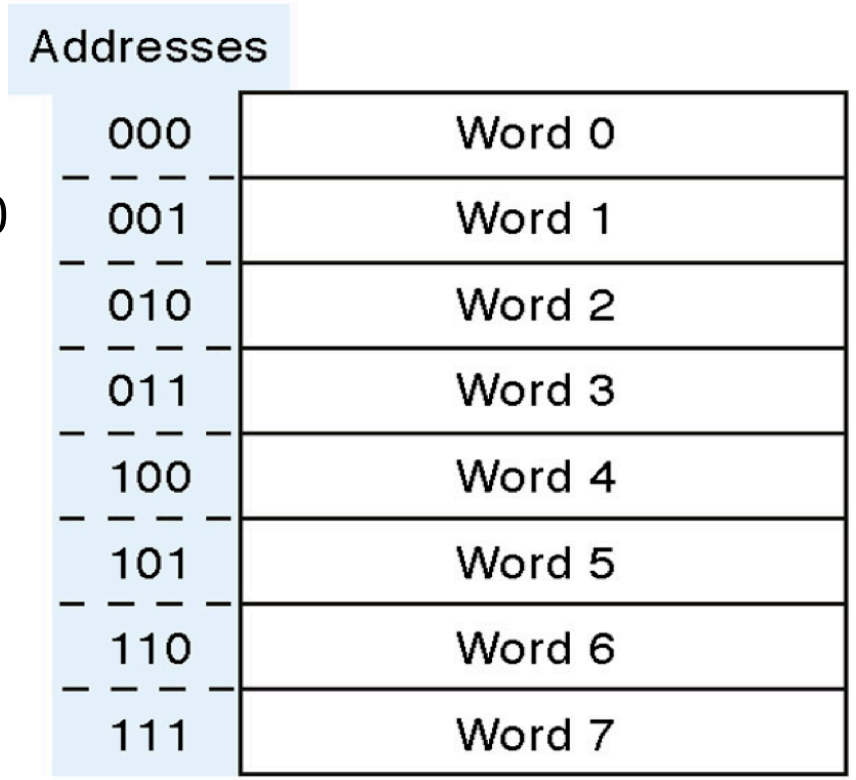
Professor Peter YK Cheung
Dyson School of Design Engineering

URL: www.ee.ic.ac.uk/pcheung/teaching/DE1_EE/
E-mail: p.cheung@imperial.ac.uk

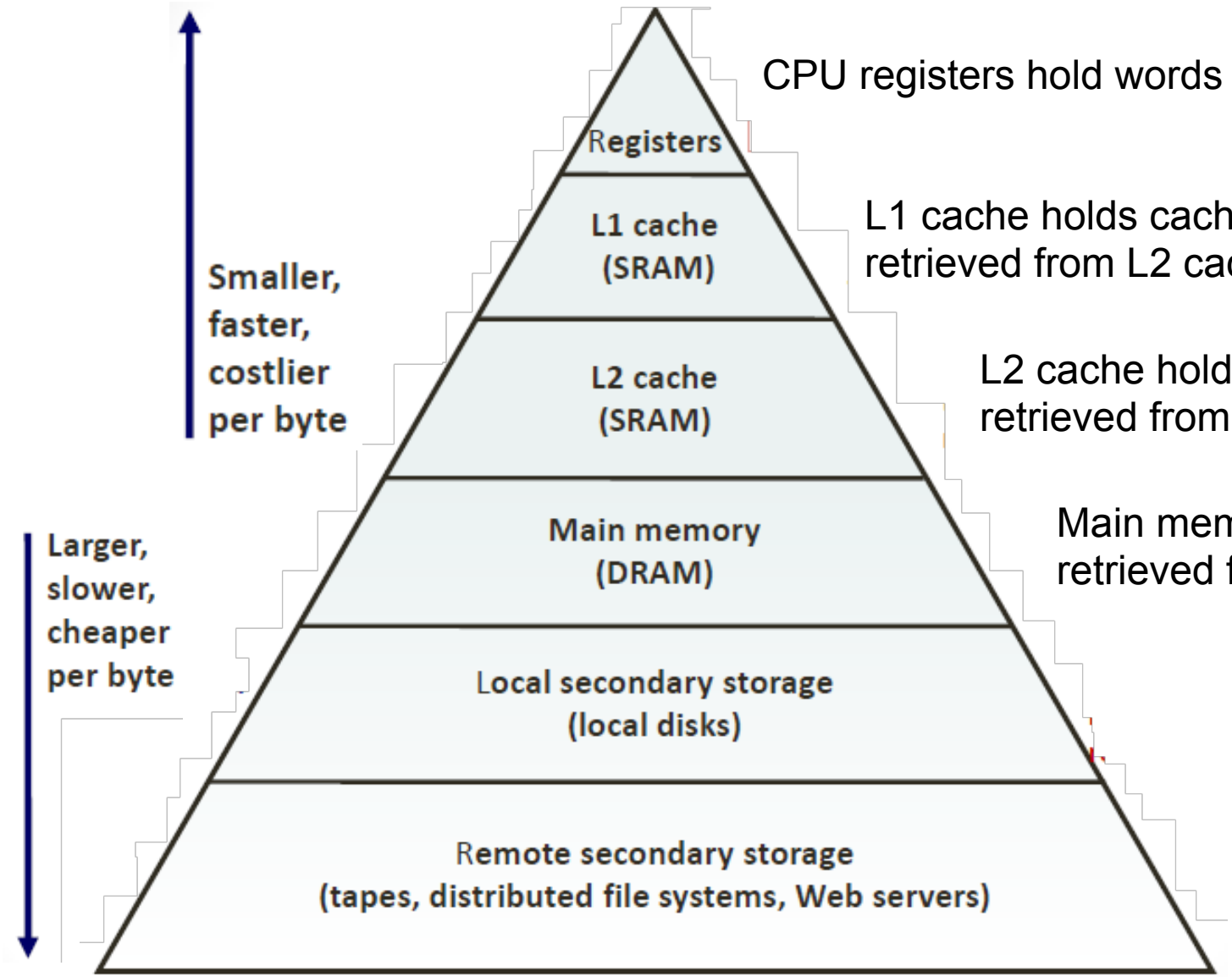
Memory Terminology

- Memory Cell: circuit that stores 1-bit of information
- Memory Word: 8 to 64 bits
- Byte: a group of 8 bits
- Capacity (=Density)
  - 4096 20-bit words
    - = 81,920 bits = 4096*20 = 4K*20
- Address
- Read Operation (=fetch operation)
- Write Operation (=store operation)

8 memory words



Memory Hierarchy



CPU registers hold words retrieved from L1 cache

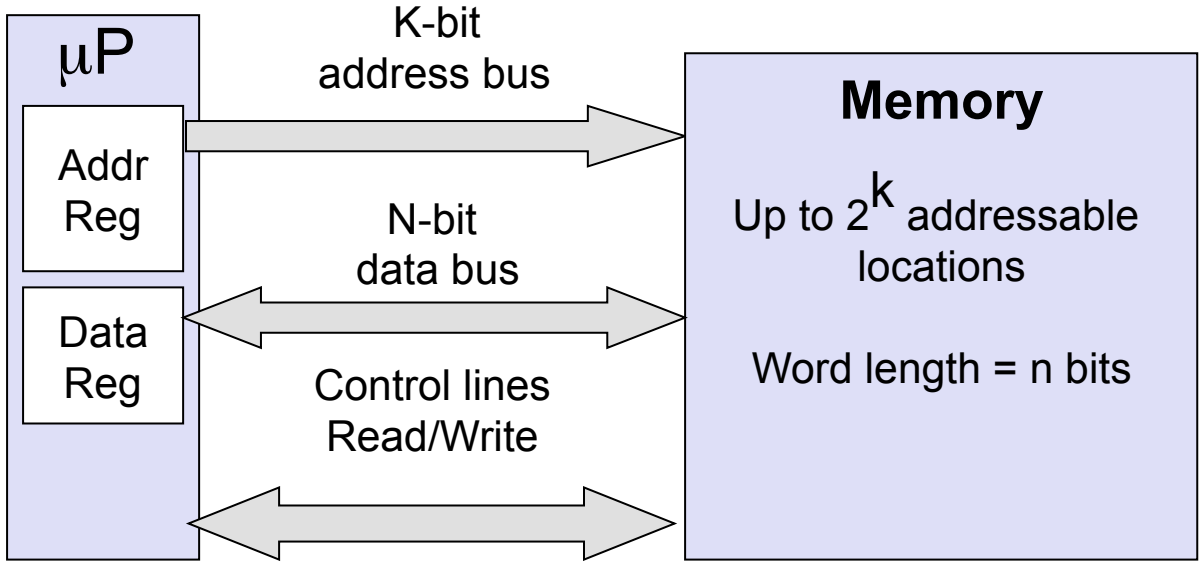
L1 cache holds cache lines retrieved from L2 cache

L2 cache holds cache lines retrieved from main memory

Main memory hold disk blocks retrieved from disks

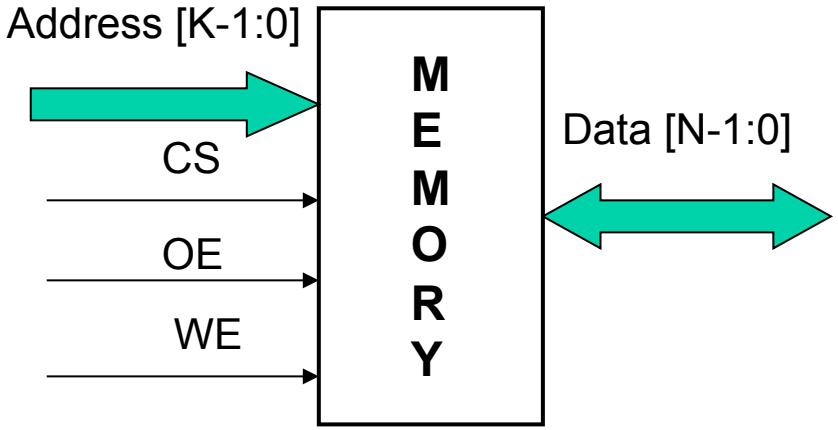
Local disks hold files retrieved from remote storage

Connecting Memory to Processor



- Interconnection between processor and memory is through three sets of wires known as buses
- Address bus establish the location of memory
- Data bus carries the data
- Control bus determine read/ write operations etc.

- **Chip Select** – must be asserted before Memory will respond to read or write operation. If negated, data bus is high impedance.
- **OE** – Output Enable: Asserted for read operation, Memory will drive data lines.
- **WE** – Write Enable: Asserted for a write operation (Memory inputs data from data pins, processor writes to memory).
- There may only be one control line (R/W).

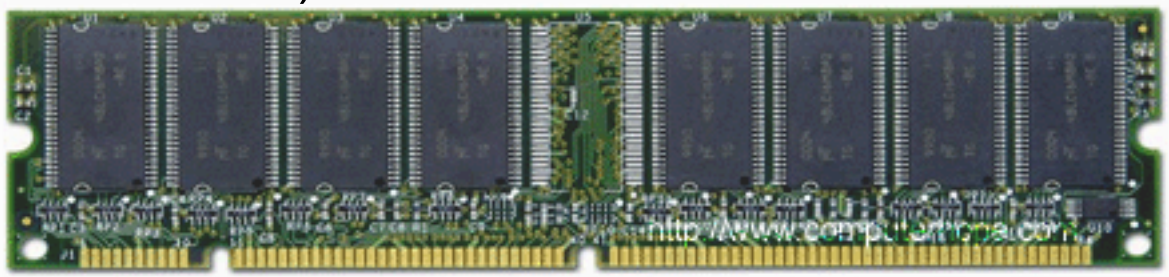


Main memory characteristics

- Most devices are 8-bits wide (Byte-addressable); some are 16-bits, others 1 bit wide.
- Chip organisation examples: 1k x 8 (capacity = 8kb), 1G x 16 (16Gb)

Characteristics

- Access Times (read, write, erase)
  - The time from a valid address being placed on the address bus until valid data appears on the data bus.
  - Faster is Better (varies from minutes to a few ns)
- Volatility
  - Ability to Retain Data After Power is Removed
- Power Consumption
  - Less is Better (mW to nW typical)
- Density
  - Larger is Better (bits/sq. transistors/bit)
- Cost
  - Less is Better



Types of Semiconductor Memories

- **ROM** – Read Only Memory - a type of memory that cannot be written, can only be read. Contents determined a manufacture time.
- **PROM** – Programmable ROM – a type of memory whose contents can be programmed by the user
  - OTP – One Time Programmable, a PROM is OTP if contents can be programmed only once.
- **RAM** – Random Access Memory
- Memory that can be both read and written during normal operation.
- Contents are volatile, i.e.will be lost on power off.
- Two types of RAM:

  Static RAM

  - Fast access time (used for off-processor cache)
  - Does not have to be refreshed

  Dynamic RAM

  - Slower access time
  - Must be refreshed
  - Much more dense

Static vs Dynamic RAM

Static RAM

- Fastest access time of all memory types. Typically the type of RAM used primarily as cache.
- Read, Write operations take equal amounts of time.
- Access to any ‘random’ location takes same amount of time.
- Basic memory cell is a latch (simple register), takes 6 transistors per memory bit.

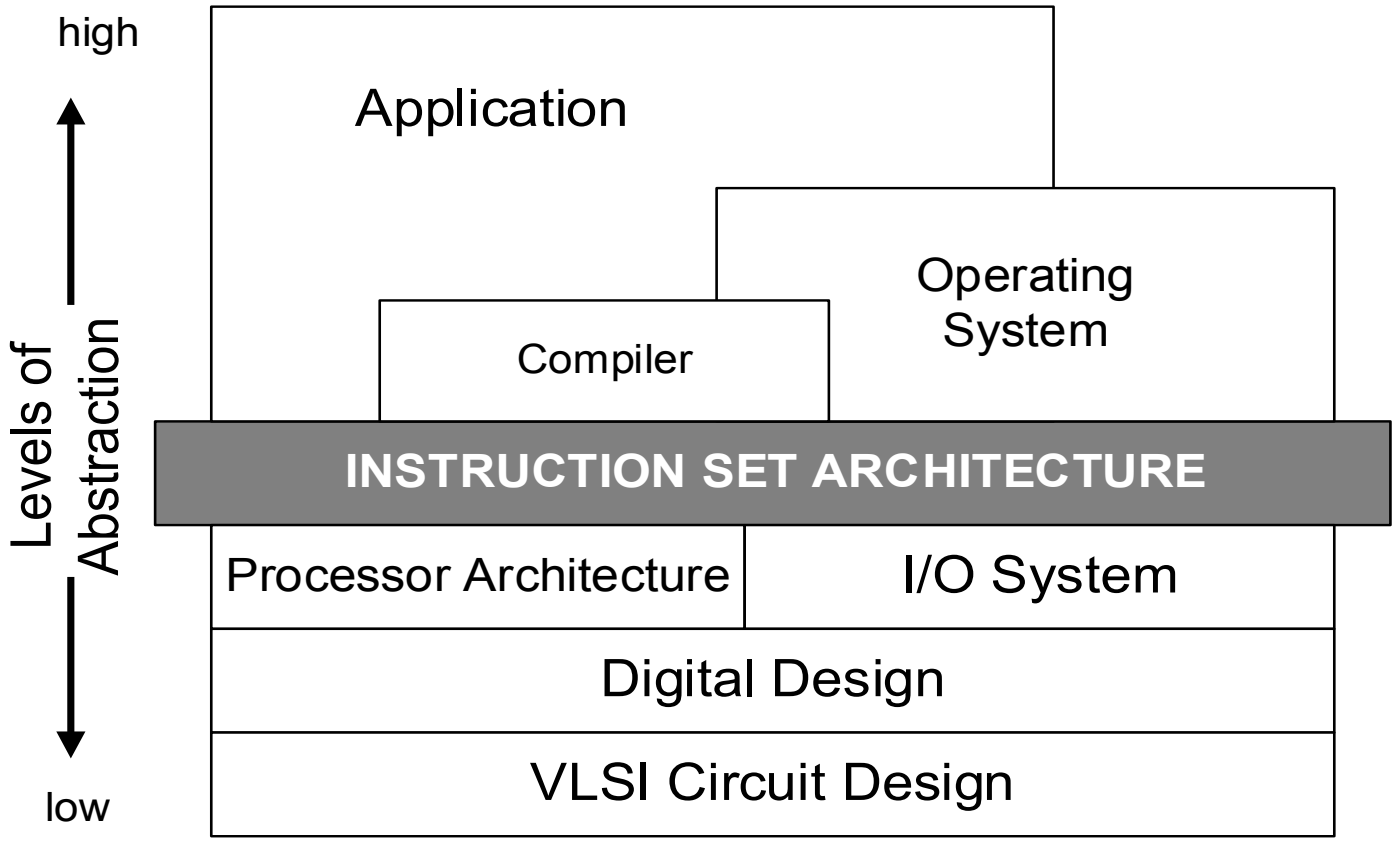
Dynamic RAM

- Must be refreshed within less than a millisecond
- Most main memory is dynamic RAM
- One transistor per memory cell (least expensive)
- SDRAM – Synchronous dynamic RAM – operates synchronously with system clock and data bus. Can handle 100MHz to >800MHz.
- DDR – Double Data Rate – can transmit data on both edges of the clock
- QDR – Quad Data Rate – twice as fast as DDR

Flash Memory

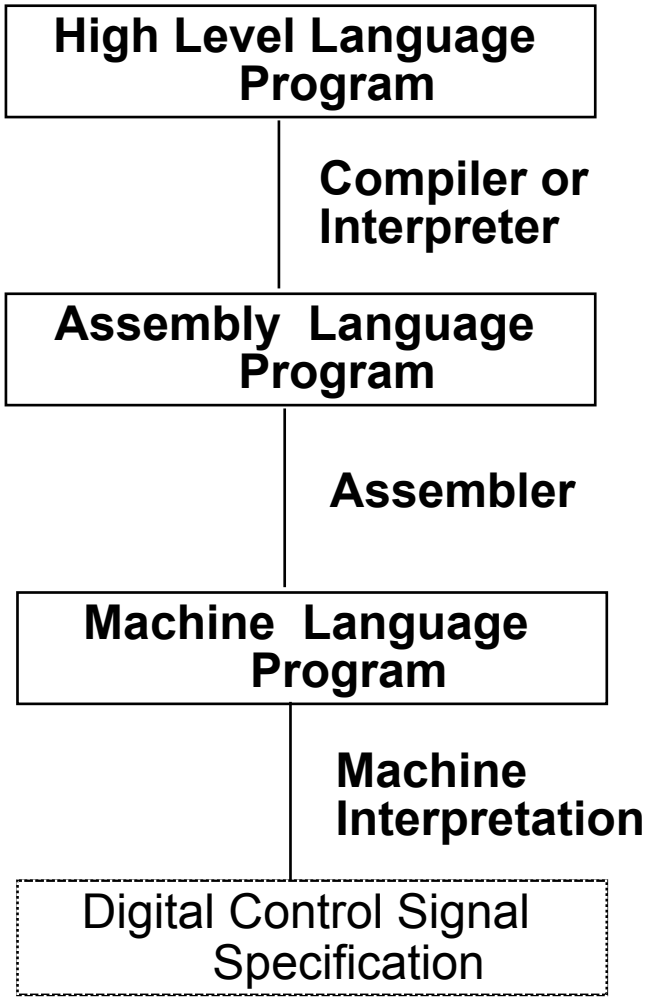
- Hybrid of RAM/ROM
- Memory parts can be electrically erased and reprogrammed without removing the chip.
- The entire chip (or block) must be erased at one time. Individual byte erasure is not possible.
- Many uses, e.g. Solid State Disks (SSD), Compact Flash (CF), Smart Media (SD cards), USB memory stick etc.
- Also embedded into microprocessors to make microcontrollers. Pyboard contains 1MB of flash memory on the chip.
- Flash memory has limited erase cycles – need smart erase algorithm with SSD drives.
- Write speed of flash memory usually is limited by erase speed – much slower than RAM and DRAM, but much faster than hard disks.

What is "Computer Architecture" ?



- Key: Instruction Set Architecture (ISA)
- Different levels of abstraction

Levels of representation in computers




```
Temp = v[k]
v[k] = v[k+1]
v[k+1] = temp
```

```
lw   $15, 0($2)
lw   $16, 4($2)
sw   $16, 0($2)
sw   $15, 4($2)
```

```
0000 1001 1100 0110 1010 1111 0101 1000
1010 1111 0101 1000 0000 1001 1100 0110
1100 0110 1010 1111 0101 1000 0000 1001
0101 1000 0000 1001 1100 0110 1010 1111
```

What is "Instruction Set Architecture (ISA)"?

- "... the attributes of a [computing] system as seen by the programmer, i.e. the conceptual structure and functional behavior, as distinct from the organization of the data flows and controls the logic design, and the physical implementation."
  - Amdahl, Blaaw, and Brooks, 1964

ISA includes:-

- Organization of Programmable Storage
- Data Types & Data Structures: Encodings & Representations
- Instruction Formats
- Instruction (or Operation Code) Set
- Modes of Addressing and Accessing Data Items and Instructions
- Exceptional Conditions

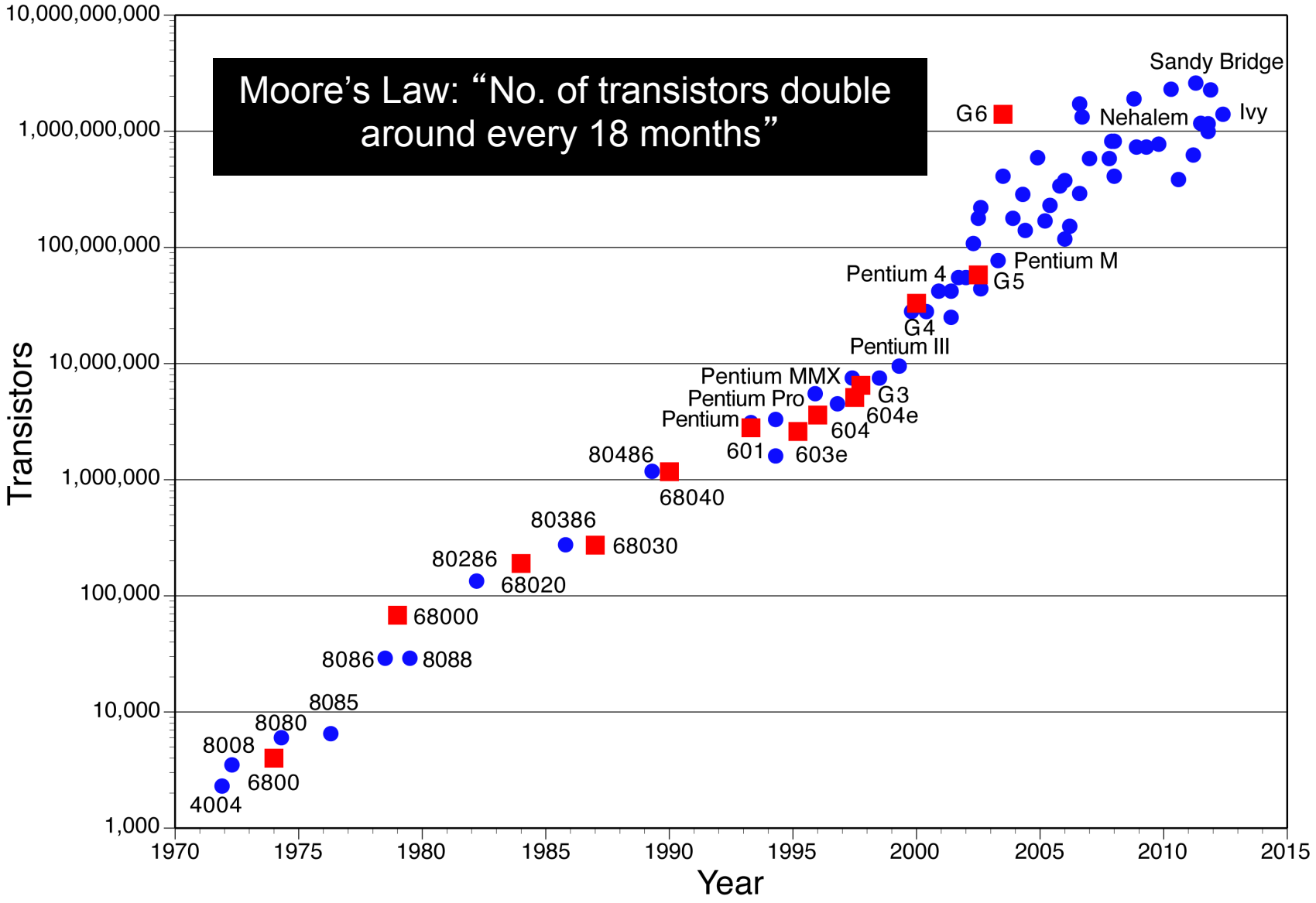
Instructure Set Architecture (ISA)

- A very important abstraction
  - interface between hardware and low-level software
  - standardizes instructions, machine language bit patterns, etc.
  - advantage: *different implementations of the same architecture*
  - disadvantage: *sometimes prevents using new innovations*

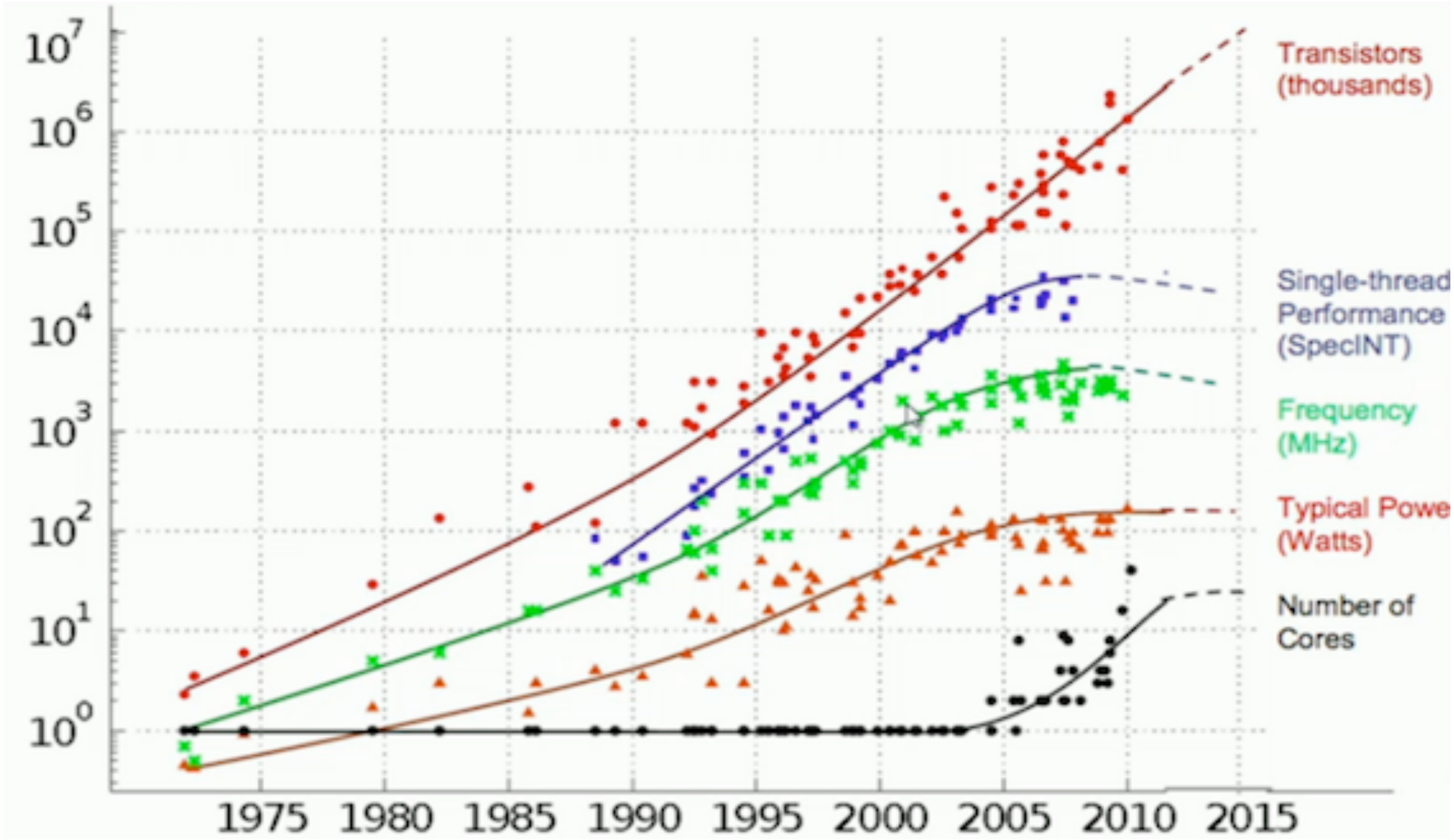
  True or False: Binary compatibility is extraordinarily important?

- Modern instruction set architectures:
  - ARM, 80x86/Pentium/K6, PowerPC, MIPS, Arduino, PIC

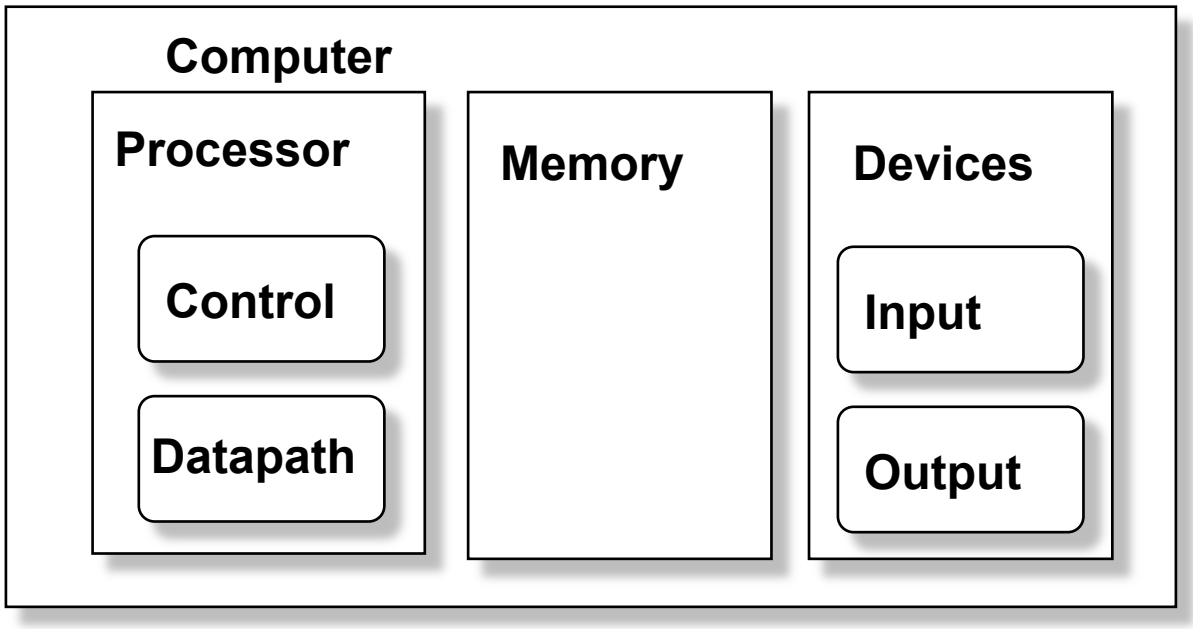
Technology: Logic Density (processors)



Technology: It is more than just transistor count

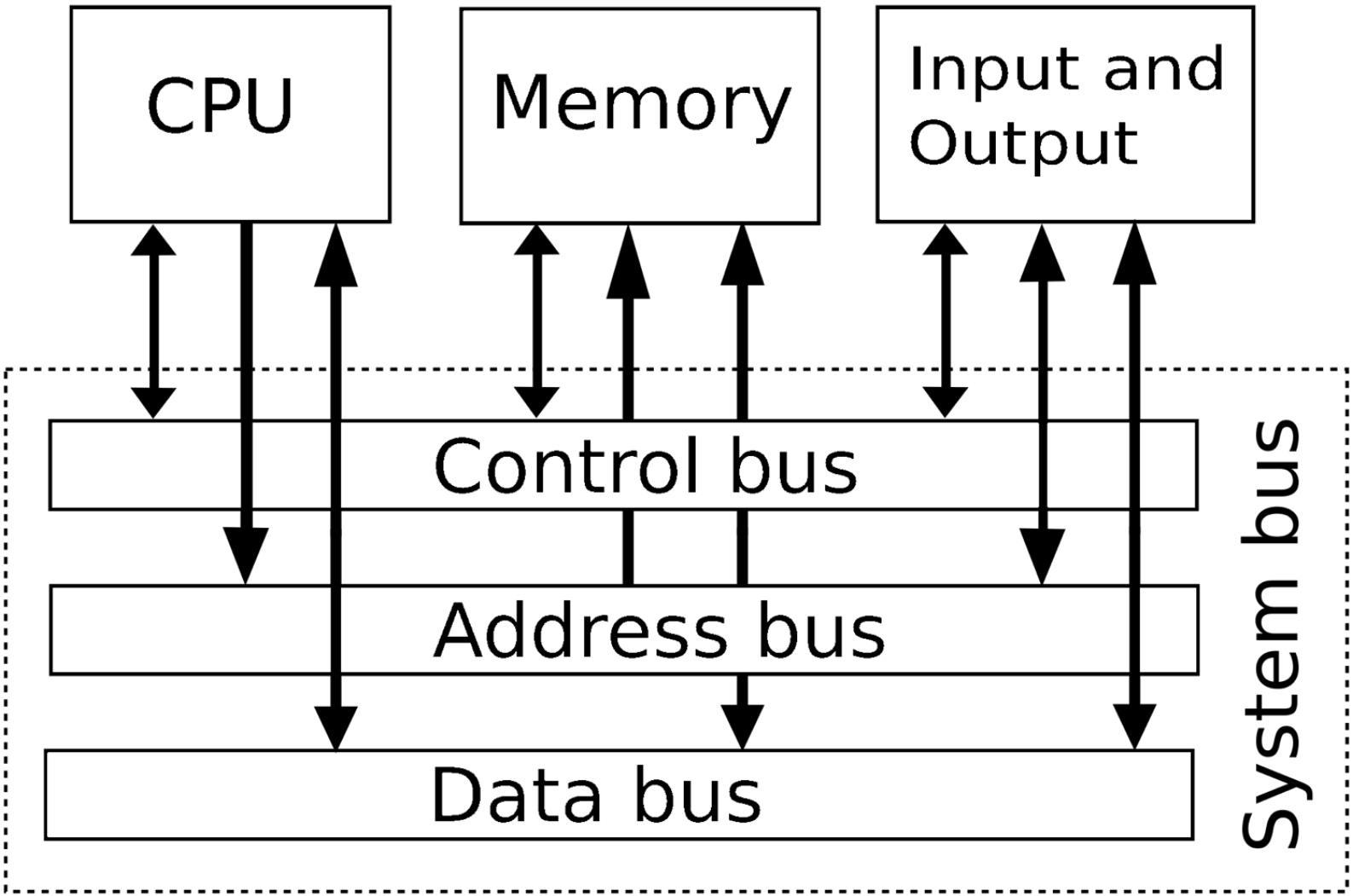


Internal Organisation



- Major components of Typical Computer System

A Typical Computer System with I/O



Summary

- All computers consist of five components
  - Processor: (1) datapath and (2) control
  - (3) Memory
  - (4) Input devices and (5) Output devices
- Not all “memory” are created equally
  - Cache: fast (expensive) memory are placed closer to the processor
  - Main memory: less expensive memory--we can have more
- Input and output (I/O) devices has the least regular organization
  - Wide range of speed: graphics vs. keyboard
  - Wide range of requirements: speed, standard, cost ... etc.
  - Least amount of research (so far)

A video on “How a CPU works?”