COMPUTER ARCHITECTURE (9 CFU)

Antonio Prete

a.prete@ing.unipi.it

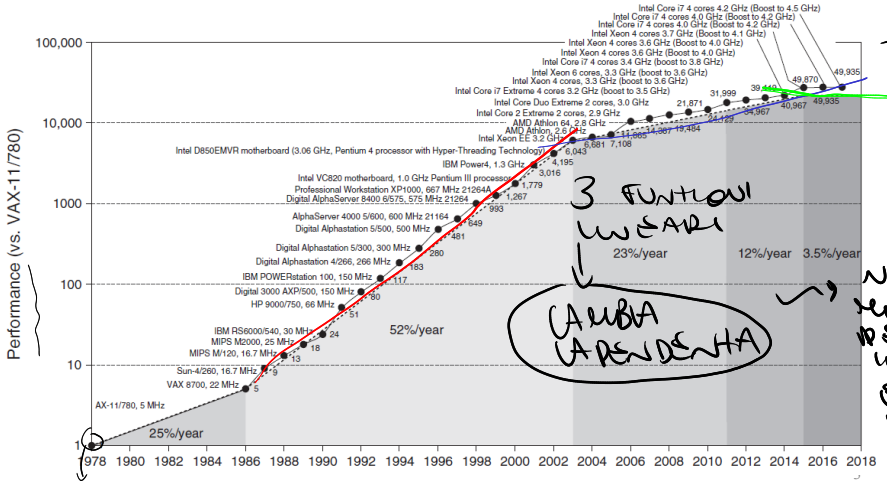
DOBBIAMO CAPIRE QUALI CLASSI DI COMPUTER CI SONO SUL MERCATO PER

- Classes of computers;
- Technology trends;

IN QUESTO MODO POSSIAMO OTTENERE I TREND PER IL FUTURO

Single processor performance vs VAX 11/780 (SPECbenchmarks)

Growth in processor performance over 40 years



TUTTE LE PERFORMANCE SONO RAPPORTATE AL VAX11/780

3 FUNZIONI LINEARI

CAMBIA LA PENDENZA

NON C'È PIÙ MODO MIGLIORARE LE PRESTAZIONI RIDUCENDO IL TEMPO DI ESECUZIONE DELLE ISTRUZIONI, DEVO PENSARE QUALCOS'ALTRO

PERFORMANCE WIZEMANN

3

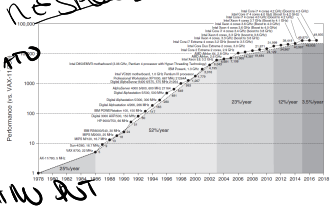
Single processor performance vs VAX 11/780

1.1 USARE LE RESPONSE TIME

- WIZEMANN
- MISURATO

1.2 THINK OUT

① MODO DI VALUTARE LE PERFORMANCE ② TIPO DI BENCHMARK USATO



This chart plots program performance relative to the VAX 11/780 as measured by the SPEC integer benchmarks.

The speed of Intel Core i7 (4 cores 4.2 GHz (Boost to 4.5 GHz)) is 50000 times the one of VAX 11/780.

4

When we say one computer is faster than another one is, what do we mean?

The user of Amazon.com may say a computer is faster when the browser spend less time to show an Amazon product,

Reduce response time - latency

the time between the start and the completion of an even

while

An Amazon.com administrator may say a computer is faster when it completes more transactions per hour.

Increase throughput

The total amount of work done in a given time

Cost (euro per transactions), power consumption (transactions per watt), space, …..

Elapsed time of the program or CPU time (?)

5

SPEC benchmarks

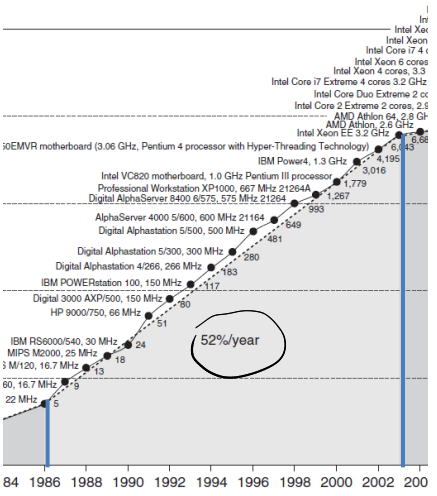
To evaluate a new system, user would compare the execution time of most used workloads (a set of programs).

Benchmark is the act of running a computer program, a set of programs, or other operations, in order to assess the relative **performance** of an object, normally by running several standard tests.

- **Embedded** Microprocessor Benchmark Consortium (EEMBC)
- **Standard** Performance Evaluation Corporation (SPEC), in particular their SPECint and SPECfp
- Transaction Processing Performance Council (TPC): **DBMS** benchmarks
  - TPC-A: measures performance in update-intensive database environments typical in **on-line transaction processing applications** (OLTP).
  - TPC-H: a **decision support** benchmark

6

1986-2003: performance at an annual rate of over 50%



Personal computers and workstations emerged with the availability of the **microprocessor**.

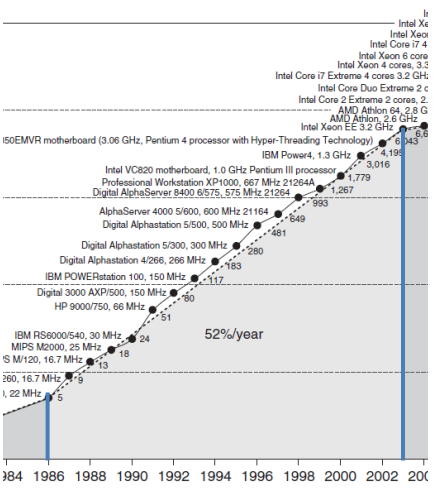
Unix and **C** made it possible to develop successfully a new set of architectures with simpler instructions, called RISC (**Reduced Instruction Set Computer**) architectures.

The RISC-based machines focused on two critical performance techniques:

- the **exploitation of instruction-level parallelism** (pipelining and later through multiple instruction issue) and
- the **use of caches**.

The technological evolution allowed to increase the clock frequency up to 2Ghz.

1986-2003: performance at an annual rate of over 50%



Main facts of the market

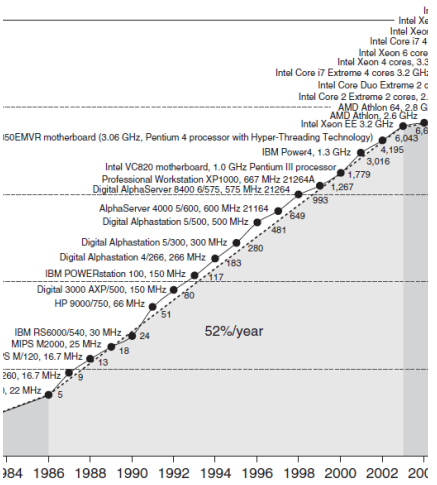
1. **Personal computers** and **workstations** emerged.

2. Even mainframe computers and high-performance supercomputers are **all collections of microprocessors**.

3. **Smart phones** and **tablet** computers, which many people are using as their primary computing platforms instead of PCs.

4. **ARM**, becomes dominant in low-end applications, such as phones.

1986-2003: performance at an annual rate of over 50%

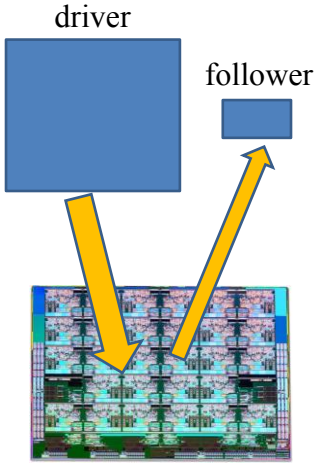


Main facts of the market

5. These **mobile client devices** are increasingly using the Internet to access warehouses containing 100,000 servers, which are being designed as if they were **a single gigantic computer**.

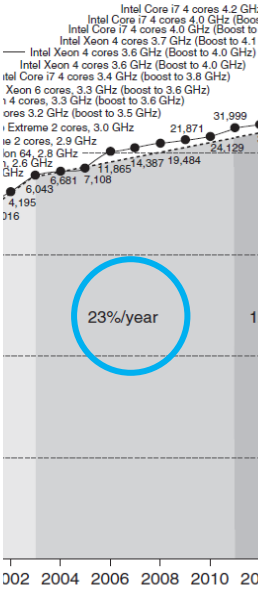
6. The nature of applications is also changing. **Speech, sound, images, and video** are becoming increasingly important, along with predictable response time that is so critical to the user experience.

Trends in micro-architecture



- Most **widespread applications** on the market finance the **research** of the microprocessors, influencing the **characteristics** of the micro-architectures.
- Application with limited market undergo / exploit the evolution.
- **First important design rule**: use in the project of a product with a limited market the components adopted in large markets.
- Trends in major markets influence the design of the remaining smaller markets-

2003-2011: Growth in processor performance over 40 years



It is no longer possible to increase the microprocessor frequency to obtain performance increases.

This situation forced the microprocessor industry to use **multiple efficient processors** or cores instead of a single inefficient processor.

In 2004 Intel canceled its high-performance uniprocessor projects.

This milestone signaled a historic switch:

- from only instruction-level parallelism (ILP),
- to data-level parallelism (DLP) and thread-level parallelism (TLP).

DLP, TLP e RLP: Higher costs in software design

Microarchitectures and compilers work to exploit ILP **without** the programmer's **effort**.

DLP, data-level parallelism

TLP, thread-level parallelism

RLP, request level parallelism

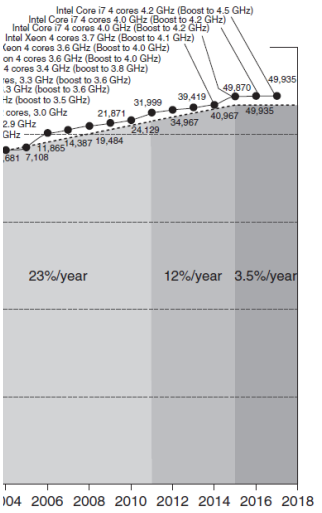
are **explicitly** parallel, requiring the restructuring of the application so that the programmer can exploit explicit parallelism.

Design effort in software

- In some instances, this is easy; in many, it is a major new burden for programmers. This depends on the actual sharing of data.
- Problems:
  - Link the solution to the parallelism model of the machine used in the development phase;
  - It is not easy to guarantee the scalability of the solution.

13

Growth in processor performance over 40 years Performance-cost-power



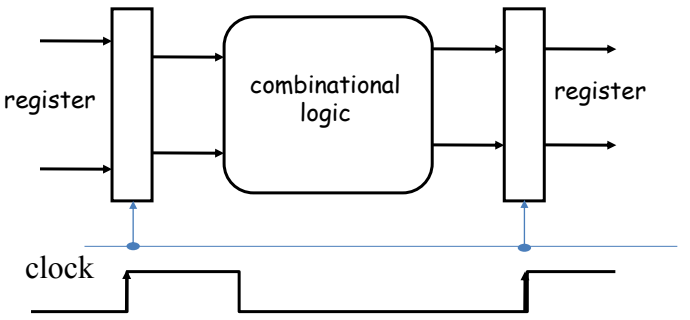
The only way to improve energy-performance-cost is specialization:

- Future microprocessors will include several domain-specific cores.
- We consider five computing markets, each characterized by different applications, requirements, and computing technologies:
  - Internet of Things/Embedded Computers
  - Personal Mobile Device (PMD)
  - Desktop Computing
  - Servers
  - Clusters/Warehouse-Scale Computers

14

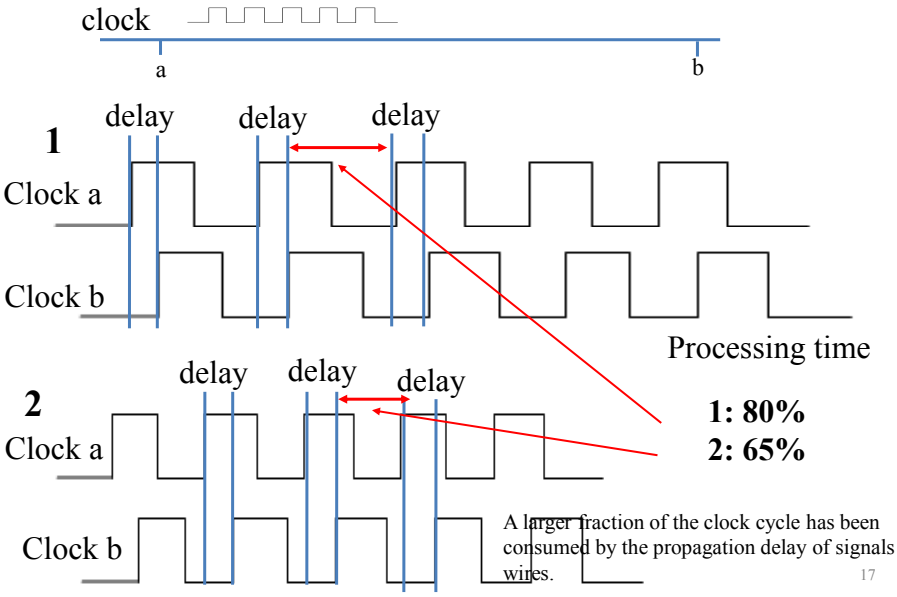
The wire delay problem

What's a Clock Cycle?



- Old days: 10 levels of gates
- Today: determined by numerous time-of-flight issues + gate delays
  - clock propagation, wire lengths, drivers

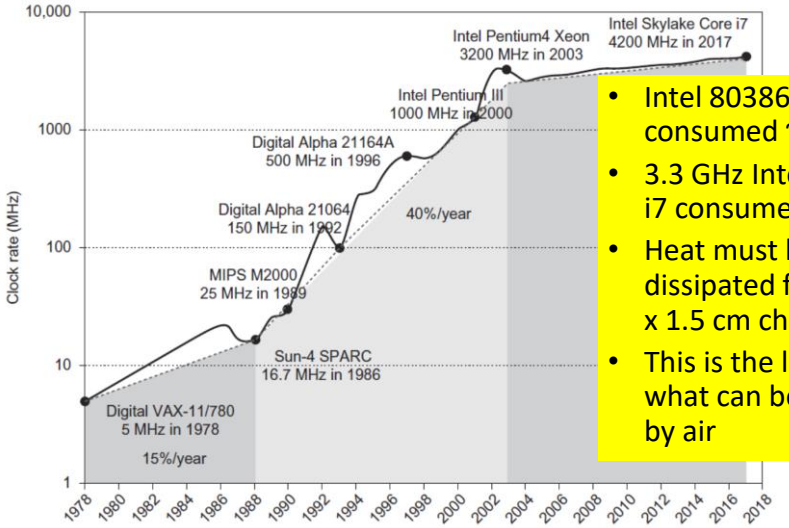
Propagation delay of clock



A larger fraction of the clock cycle has been consumed by the propagation delay of signals on wires.

17

Power



- Intel 80386 consumed ~ 2 W
- 3.3 GHz Intel Core i7 consumes 130 W
- Heat must be dissipated from 1.5 x 1.5 cm chip
- This is the limit of what can be cooled by air

Classes of Computers

- Internet of Things/Embedded Computers
- Personal Mobile Device (PMD)
- Desktop Computing
- Servers
- Clusters/Warehouse-Scale Computers

Internet of Things/Embedded Computers (I)

Embedded computers are present in everyday used machines: microwaves, washing machines, most of printers, most of networking switches, and all cars contain simple embedded microprocessors.

The ability to run third-party software is the line between no-embedded and embedded computers.

Internet of Things/Embedded Computers (II)

- Internet of Things (IoT), refers to embedded computers that are connected to the Internet, typically wirelessly.
- When augmented with sensors and actuators, IoT devices collect useful data and interact with the physical world, leading to a wide variety of "smart" applications,
  - such as smart watches, smart thermostats, smart speakers, smart cars, smart homes, smart grids, and smart cities.
- Internet of Things (IoT)
  - offers useful data and local services to remote applications
  - uses remote information and resources
  - allows us to update the software.

Internet of Things/Embedded Computers (III)

- Embedded computers have a range of processing power and of cost very large:
  - 8-bit and 16-bit processors (> a dime),
  - 32-bit microprocessors (>$5),
  - high-end processors for network switches ($100).
- **The main goal is to reach the required performance at the minimum price.**
- Microcontroller, a chip with off-the-shelf microprocessors and other special-purpose hardware.
- Projections for the number of IoT devices in 2022 range from 50 to 100 billion.

Personal Mobile Device (PMD) (I)

Wireless devices with multimedia user interfaces such as cell phones, tablet computers.

- The price for the consumer complete product is a few hundred dollars.
- Power consumption must be limited because of:
  - use of batteries
  - use of plastic packaging
  - absence of a fan cooling.
- Energy and size requirements lead to use of **Flash memory for storage**.

23

Personal Mobile Device (PMD) (II)

- Applications are often Web-based and media-oriented,
- Responsiveness and predictability are key characteristics for media applications.
- Key characteristics in many PMD applications are:
  - the need of an efficient use of energy (problems in heat dissipation) and
  - the need to minimize memory (is a big portion of system cost), in particular, the code size.

24

Desktop Computing

Desktop Computing spans from low-end netbooks to high-end workstations.

- The trend of desktop market is to be driven by *price-performance* optimization
- The increasing use of Web-centric, interactive applications launches new challenges to performance evaluation.
- The current strong trend of being the access point to data, services and applications available remotely via the cloud.
- This allows you to use different platforms over time to always use the same resources.

Servers

- Servers have become the backbone of large-scale enterprise computing, **replacing the traditional mainframe**.
  - **Availability**, Servers must operate seven days a week, 24 hours a day.
  - **Scalability**, the ability to scale up the computing capacity, the memory, the storage, and the *bandwidth is crucial.*
  - **Efficient throughput**, the *overall performance, in terms of transactions per minute or Web* pages served per second, is crucial.

Clusters/Warehouse-Scale Computers (I)

Clusters are collections of desktop computers or servers connected by local area networks acting as a single larger computer.

- *Warehouse-scale computers (WSCs) are designed so- that tens of* thousands of servers can act as one.
- Each node runs its own operating system, and nodes communicate using a networking protocol.

Clusters/Warehouse-Scale Computers (II)

Price-performance and power are critical.

- 80% of the cost of a $90M warehouse is associated to power and cooling of the computers inside.
- The computers and networking must be replaced every few years.

- *Supercomputers* are different because they emphasize floating-point performance and by running large, communication-intensive batch programs that can run for weeks at a time.

Classes of Computers
Critical system design issues

- Personal Mobile Device (PMD), phones, tablet computers
  - Emphasis on **energy** efficiency and software real-time
  - Media performance and responsiveness
- Desktop Computing
  - Emphasis on price-performance
  - **Energy** efficiency and graphics performance
- Servers
  - Emphasis on availability, scalability, throughput
  - **Energy** efficiency
- Clusters / Warehouse Scale Computers, "Software as a Service (SaaS)"
  - Emphasis on availability and price-performance
  - Sub-class: Supercomputers, emphasis: floating-point performance and fast internal networks
  - **Energy** proportionality
- Embedded Computers
  - Emphasis: price, **energy**
  - Application-specific performance

Classes of Computers:
Prices and features

Feature	Personal mobile device (PMD)	Desktop	Server	Clusters/warehouse-scale computer	Internet of things/ embedded
Price of system	$100–$1000	$300–$2500	$5000–$10,000,000	$100,000–$200,000,000	$10–$100,000
Price of microprocessor	$10–$100	$50–$500	$200–$2000	$50–$250	$0.01–$100
Critical system design issues	Cost, energy, media performance, responsiveness	Price-performance, energy, graphics performance	Throughput, availability, scalability, energy	Price-performance, throughput, energy proportionality	Price, energy, application-specific performance

Classes of Computers:
Number of the market

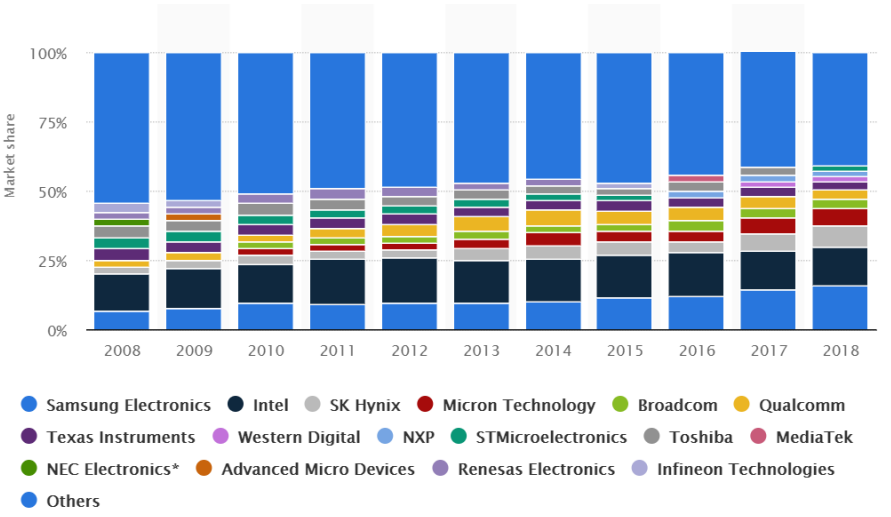
Feature	Personal mobile device (PMD)	Desktop	Server	Clusters/warehouse-scale computer	Internet of things/ embedded
Price of system	$100–$1000	$300–$2500	$5000–$10,000,000	$100,000–$200,000,000	$10–$100,000
Price of microprocessor	$10–$100	$50–$500	$200–$2000	$50–$250	$0.01–$100
Critical system design issues	Cost, energy, media performance, responsiveness	Price-performance, energy, graphics performance	Throughput, availability, scalability, energy	Price-performance, throughput, energy proportionality	Price, energy, application-specific performance

Sales in 2015:

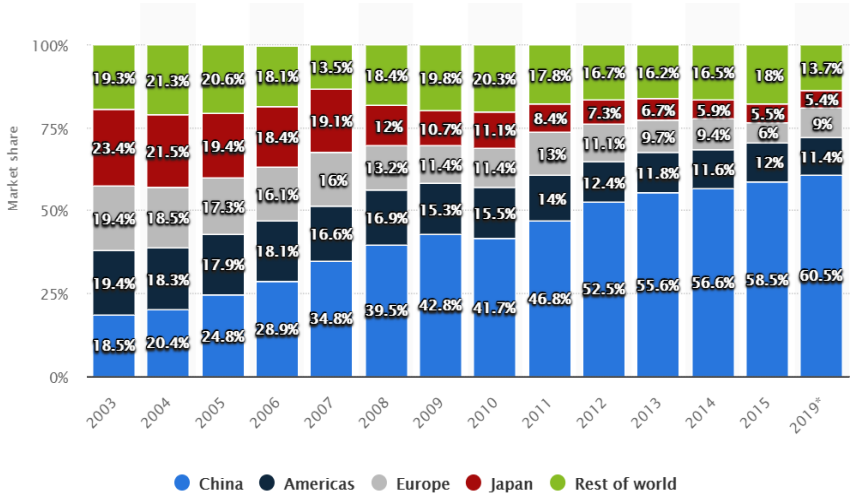
- 1.6 billion PMDs (90% cell phones),
- 275 million desktop PCs,
- 15 million servers,
- embedded processors sold was nearly **19 billion**.

14.8 billion ARM-technology-based chips were shipped.

Market share held by semiconductor vendors worldwide from 2008 to 2018



Semiconductor consumption market share worldwide, from 2003 to 2019 , by region



33

For servers, availability is critical

Application	Cost of downtime per hour	Annual losses with downtime of 1% (87.6 h/year)	0.5% (43.8 h/year)	0.1% (8.8 h/year)
Brokerage service	$4,000,000	$350,400,000	$175,200,000	$35,000,000
Energy	$1,750,000	$153,300,000	$76,700,000	$15,300,000
Telecom	$1,250,000	$109,500,000	$54,800,000	$11,000,000
Manufacturing	$1,000,000	$87,600,000	$43,800,000	$8,800,000
Retail	$650,000	$56,900,000	$28,500,000	$5,700,000
Health care	$400,000	$35,000,000	$17,500,000	$3,500,000
Media	$50,000	$4,400,000	$2,200,000	$400,000

Feature	Personal mobile device (PMD)	Desktop	Server	Clusters/warehouse-scale computer	Internet of things/ embedded
Price of system	$100–$1000	$300–$2500	$5000–$10,000,000	$100,000–$200,000,000	$10–$100,000
Price of microprocessor	$10–$100	$50–$500	$200–$2000	$50–$250	$0.01–$100
Critical system design issues	Cost, energy, media performance, responsiveness	Price-performance, energy, graphics performance	Throughput, availability, scalability, energy	Price-performance, throughput, energy proportionality	Price, energy, application-specific performance

34

FUNCTIONAL REQUIREMENTS

Defining Computer Architecture (I)

- The tasks of a computer designer are:
  - To list the most important features;
  - To design a computer **maximizing the performance in case of a set of applications** and energy efficiency respecting the **cost**, **power**, and **availability** constraints.

Computer Architecture

- Instruction set,
- functional organization and logic design, and
- implementation.
  - The implementation may encompass integrated circuit design, packaging, power, and cooling.

Optimizing the design requires familiarity with a very wide range of technologies, from compilers and operating systems to logic design and packaging.

Functional requirements

Market and product competition

The requirements may be specific features inspired by the **market**.

Application software typically drives the choice of certain functional requirements by determining how the computer will be used.

If a large body of software exists for a particular instruction set architecture, the architect may decide that a new computer should implement an **existing instruction set**.

The presence of a large market for **a particular class of applications** might encourage the designers to incorporate requirements that would make the **computer competitive in that market**.

Defining Computer Architecture (II)

Functional requirements

- Personal mobile device
  - Real-time performance for a range of tasks, including interactive performance for graphics, video, and audio; energy
- General-purpose desktop
  - Balanced performance for a range of tasks, including interactive performance for graphics, video, and audio

Defining Computer Architecture (III)
Functional requirements

- Servers
  - Support for databases and transaction processing; enhancements for reliability and availability; support for scalability
- Clusters/warehouse-scale computers
  - Throughput performance for many independent tasks; error correction for memory; energy proportionality

Defining Computer Architecture (IV)
Functional requirements

- Embedded computing
  - Often requires special support for graphics or video (or other application-specific extension); power limitations and power control may be required; real-time constraints

The most popular RISC processor come from ARM (Advanced Risc Machine) which were in **14.8 billion chips shipped in 2015**, or roughly **50 times** as many chips that shipped with 80x86 processors.

Level of software compatibility

Determines amount of existing software for computer

- At programming language
  - Most flexible for designer; need new compiler
- Object code or binary compatible
  - Instruction set architecture is completely defined—little flexibility—but no investment needed in software or porting programs

An instruction set must be designed to survive rapid changes in computer technology.

Operating system requirements

Operating system requirements: necessary features to support (chosen) OS

- Size of address space
  - Very important feature,
    - may limit applications
- Memory management
  - Required for modern OS
    - may be paged or segmented
- Protection
  - Different OS and application needs: page versus segment
    - virtual machines
- *Virtualization*
  - *Memory management, Protection and Interrupt*

Standards

Certain standards may be required by marketplace

Floating point

- Format and arithmetic: IEEE 754 standard, special arithmetic for graphics or signal processing

I/O interfaces

- For I/O devices: Serial ATA, Serial Attached SCSI, PCI Express

Operating systems

- UNIX, Windows, Linux, CISCO IOS

Networks

- Support required for different networks: Ethernet, Infiniband

Programming languages

- Languages (ANSI C, C++, Java, Fortran) affect instruction set

Market drivers

- Embedded systems in the Automotive Industry
- Wearable devices
- Embedded systems for smart home
- Embedded systems in smart cities
- Embedded systems for healthcare equipment

Opportunities

- Artificial intelligence technologies expand the functionality of existing applications, allow to create new applications and simplify human-machine interaction.
- Use of Internet Of Things (IOT) feature in embedded systems.
- Impact of use of multicore in Industry Applications.

PERFORMANCE

45

Flynn's Taxonomy

Flynn (1966) looked at the parallelism in the instruction and data streams.

- Single instruction stream, single data stream (SISD)
  - uniprocessor
- Single instruction stream, multiple data streams (SIMD)
  - Vector architectures
  - Multimedia extensions
  - Graphics processor units
- Multiple instruction streams, single data stream (MISD)
  - No commercial implementation
- Multiple instruction streams, multiple data streams (MIMD)
  - Tightly-coupled MIMD
  - Loosely-coupled MIMD

46

Parallel Architectures

Classes of architectural parallelism:

- Instruction-Level Parallelism (ILP)
  - pipeline and speculative execution
- Vector architectures/Graphic Processor Units (GPUs)
  - by applying a single instruction to a collection of data in parallel
- Thread-Level Parallelism
  - Management of parallel threads
- Request-Level Parallelism
  - Management of tasks

Multiple instruction streams, multiple data streams (MIMD)

- **Tightly coupled MIMD** architectures exploit thread-level parallelism where *multiple cooperating threads operate in parallel*.

- **Loosely coupled MIMD architectures**—specifically, clusters and warehouse-scale computers— exploit request-level parallelism, where *many independent tasks can proceed in parallel naturally with little need for communication or synchronization*.

Technology Trends (I)

The evolution of technologies has a big impact on the design of a computer.

- *Integrated circuit logic technology*
  - The transistor counts on a chip about 40% to 55% per year, or doubling every 18 to 24 months.
- *Semiconductor DRAM* (dynamic random-access memory) is the foundation of **main memory**
  - Capacity per DRAM chip has increased of about 25% to 40% per year, doubling roughly every two to three years.

49

Technology Trends (II)

- *Semiconductor Flash*

  the standard storage device in PMDs.

  Capacity per Flash chip has been increased of about 50% to 60% per year recently, doubling roughly every two years.

- *Magnetic disk technology*.

  Since 2004, it has dropped back to about 40% per year, or doubled every three years.

  This technology is central to server and warehouse scale storage.

- *Network technology-Network*

  Performance depends both on the performance of switches and on the performance of the transmission system.

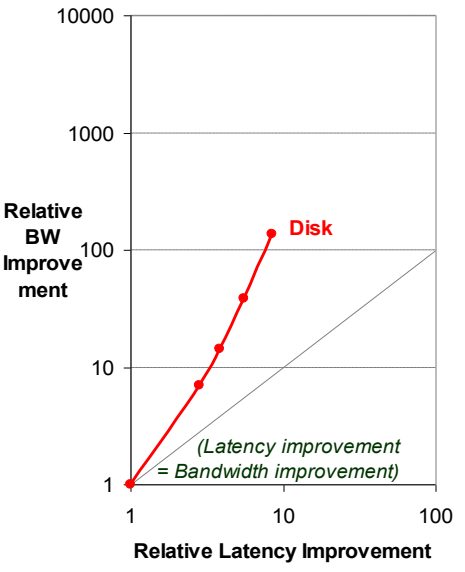
50

Bandwidth and Latency (performance)

- Bandwidth or throughput
  - Total work done in a given time
  - 10,000-25,000X improvement for processors
  - 300-1200X improvement for memory and disks

- Latency or response time
  - Time between start and completion of an event
  - 30-80X improvement for processors
  - 6-8X improvement for memory and disks

Latency Lags Bandwidth (for last ~20 years)



Bandwidth or throughput
Total work done in a given time
Mbytes/second

300-1200X improvement for disks

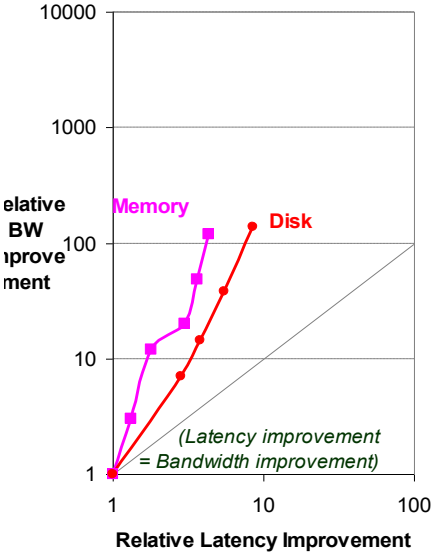
Latency or response time
Time between start and completion of an event
average disk access time in milliseconds

6-8X improvement for disks

Performance milestone for disks

Hard disk	3600 RPM	5400 RPM	7200 RPM	10,000 RPM	15,000 RPM	15,000 RPM
Product	CDC WrenI 94145-36	Seagate ST41600	Seagate ST15150	Seagate ST39102	Seagate ST373453	Seagate ST600MX0062
Year	1983	1990	1994	1998	2003	2016
Capacity (GB)	0.03	1.4	4.3	9.1	73.4	600
Disk form factor	5.25 in.	5.25 in.	3.5 in.	3.5 in.	3.5 in.	3.5 in.
Media diameter	5.25 in.	5.25 in.	3.5 in.	3.0 in.	2.5 in.	2.5 in.
Interface	ST-412	SCSI	SCSI	SCSI	SCSI	SAS
Bandwidth (MBytes/s)	0.6	4	9	24	86	250
Latency (ms)	48.3	17.1	12.7	8.8	5.7	3.6

Latency Lags Bandwidth (last ~20 years)



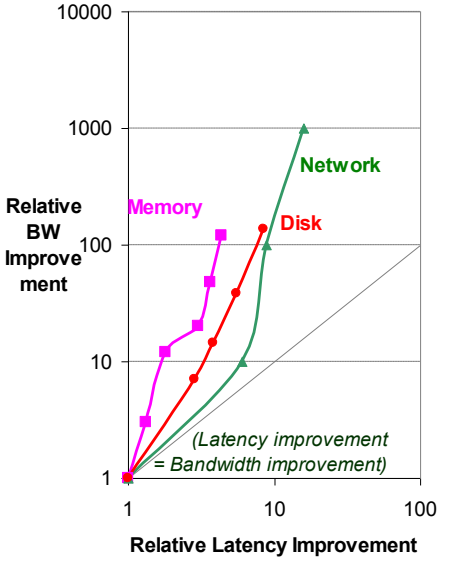
Performance Milestones

- Memory Module: 16bit plain DRAM, Page Mode DRAM, 32b, 64b, SDRAM, DDR SDRAM

Performance milestones for memory

Memory module	DRAM	Page mode DRAM	Fast page mode DRAM	Fast page mode DRAM	Synchronous DRAM	Double data rate SDRAM	DDR4 SDRAM
Module width (bits)	16	16	32	64	64	64	64
Year	1980	1983	1986	1993	1997	2000	2016
Mbits/DRAM chip	0.06	0.25	1	16	64	256	4096
Die size (mm^2)	35	45	70	130	170	204	50
Pins/DRAM chip	16	16	18	20	54	66	134
Bandwidth (MBytes/s)	13	40	160	267	640	1600	27,000
Latency (ns)	225	170	125	75	62	52	30

Latency Lags Bandwidth (last ~20 years)



Performance Milestones

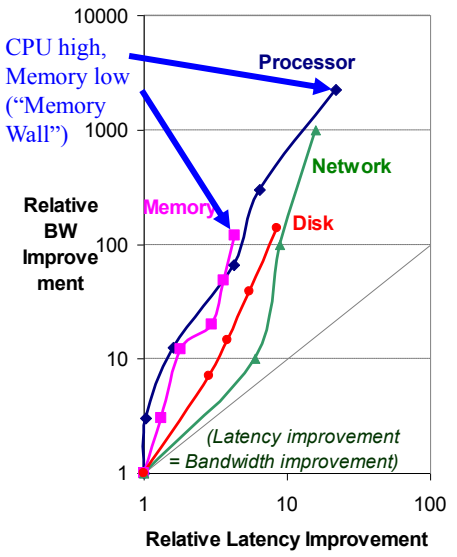
- Ethernet: 10Mb, 100Mb, 1000Mb, 10000 Mb/s (16x,1000x)
- Memory Module: 16bit plain DRAM, Page Mode DRAM, 32b, 64b, SDRAM, DDR SDRAM (4x,120x)
- Disk: 3600, 5400, 7200, 10000, 15000 RPM (8x, 143x)

(latency = simple operation w/o contention
BW = best-case)

Performance milestone for networks

Local area network	Ethernet	Fast Ethernet	Gigabit Ethernet	10 Gigabit Ethernet	100 Gigabit Ethernet	400 Gigabit Ethernet
IEEE standard	802.3	803.3u	802.3ab	802.3ac	802.3ba	802.3bs
Year	1978	1995	1999	2003	2010	2017
Bandwidth (Mbits/seconds)	10	100	1000	10,000	100,000	400,000
Latency (μs)	3000	500	340	190	100	60

Latency Lags Bandwidth (last ~20 years)



Performance milestones for microprocessors

Microprocessor	16-Bit address/ bus, microcoded	32-Bit address/ bus, microcoded	5-Stage pipeline, on-chip I & D caches, FPU	2-Way superscalar, 64-bit bus	Out-of-order 3-way superscalar	Out-of-order superpipelined, on-chip L2 cache	Multicore OOO 4-way on chip L3 cache, Turbo
Product	Intel 80286	Intel 80386	Intel 80486	Intel Pentium	Intel Pentium Pro	Intel Pentium 4	Intel Core i7
Year	1982	1985	1989	1993	1997	2001	2015
Die size (mm^2)	47	43	81	90	308	217	122
Transistors	134,000	275,000	1,200,000	3,100,000	5,500,000	42,000,000	1,750,000,000
Processors/chip	1	1	1	1	1	1	4
Pins	68	132	168	273	387	423	1400
Latency (clocks)	6	5	5	5	10	22	14
Bus width (bits)	16	32	32	64	64	64	196
Clock rate (MHz)	12.5	16	25	66	200	1500	4000
Bandwidth (MIPS)	2	6	25	132	600	4500	64,000
Latency (ns)	320	313	200	76	50	15	4

Intel i7 (1)

Intel Core i7 can generate two data memory references per core each clock cycle

– with four cores and a 3.2 GHz clock rate, the i7 can generate a peak of 25.6 billion 64-bit data memory references per second,

- to a peak instruction demand (for four cores) of about 12.8 billion 128-bit instruction references per second;
- total peak bandwidth of 409.6 GB/sec

Intel i7 (2)

- In contrast, the peak bandwidth to DRAM main memory is only 6% of this
  - 25 GB/sec

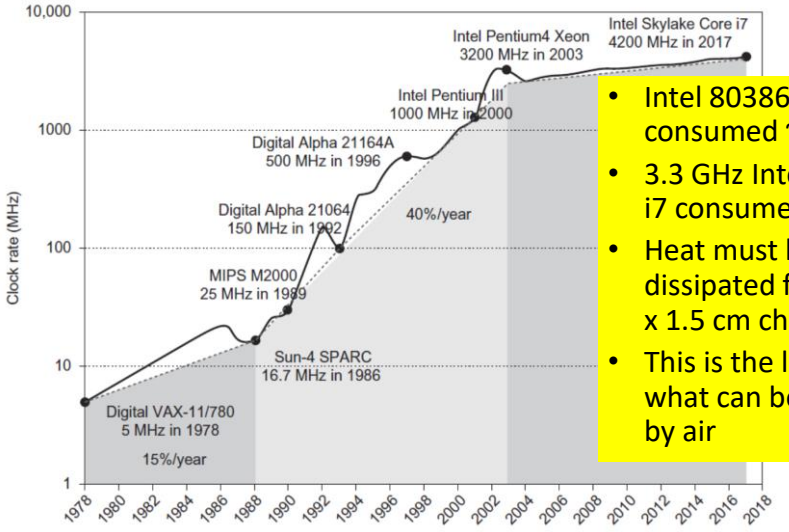
61

Intel i7 (3)

- This incredible bandwidth is achieved
  - by multiporting and pipelining the cache accesses;
  - by the use of multiple levels of caches,
  - by using a separate instruction and data cache at the first level,
  - By using separate first- and sometimes second-level caches per core.

62

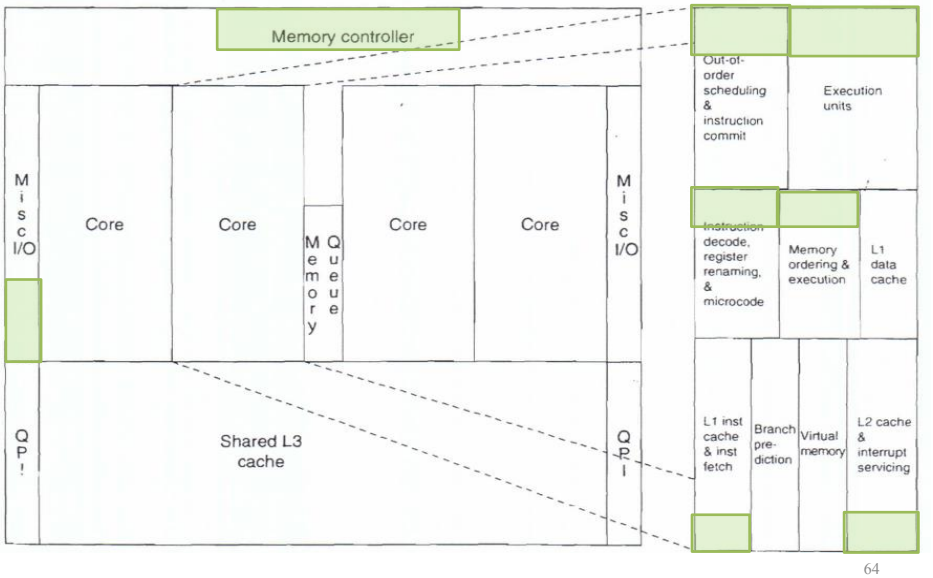
Power



- Intel 80386 consumed ~ 2 W
- 3.3 GHz Intel Core i7 consumes 130 W
- Heat must be dissipated from 1.5 x 1.5 cm chip
- This is the limit of what can be cooled by air

63

Intel core i7



64

Dependability

Service Level Agreement (SLA)

- One difficult question is deciding when a system is operating properly.
- Infrastructure providers started offering service level agreements (SLAs) or service level objectives (SLOs) to guarantee that their networking or power service would be dependable.
- For example, they would pay a penalty to the customer if they did not meet an agreement more than some hours per month.
- Thus, an SLA could be used to decide whether the system is up or down.

- The system has additional resources and software features for:
  - Error detection
  - System reconfiguration
  - System recovery

Fault-tolerant system design (1)

Error detection

A fault causes an incorrect operation of the system. There is a temporal and spatial distance between error and incorrect operation.

Temporal. As long as a component in error is not used, the fault cannot be discovered.

- If I don't use the failed memory bit, I can't know it's broken.

Periodic memory test.

Spatial. The incorrect functioning of a component can be due to an error in another component.

- Word opens in a file stored in a memory with faults.

Encapsulation, verification of the result at the component level for each individual operation.

Reconfiguration and system recovery (additional software and resources)

Reconfiguration involves using additional resources to get a new working system configuration.

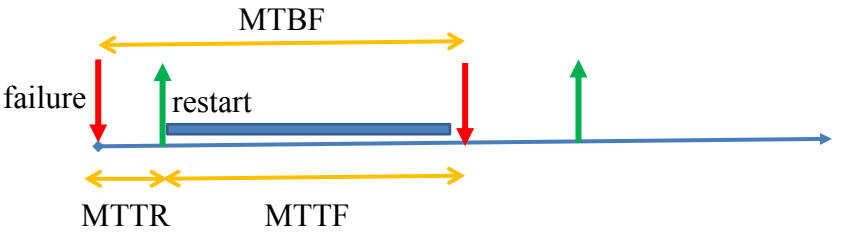
- It can be done **hot**, that is, without completely interrupting the service.
- **Cold**, stopping the service.

System recovery involves finding a set of data that represents a consistent state of the system.

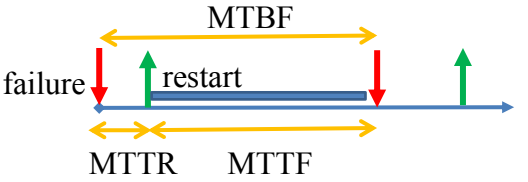
- Roll back the system by redoing the unfinished operations (atomic transactions, commit protocol, ...).

Main items in dependability

- Module reliability
  - Mean time to failure (MTTF)
  - Mean time to repair (MTTR)
  - Mean time between failures (MTBF) = MTTF + MTTR
  - Availability = MTTF / MTBF



An example (1)

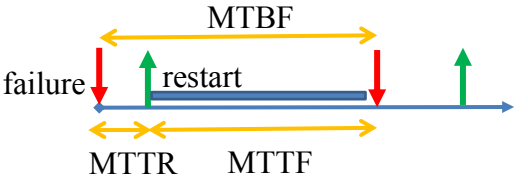


Assume a disk subsystem with the following components and MTTF:

- 10 disks, each rated at 1,000,000-hour MTTF
- 1 ATA controller, 500,000-hour MTTF
- 1 power supply, 200,000-hour MTTF
- 1 fan, 200,000-hour MTTF
- 1 ATA cable, 1,000,000-hour MTTF

Using the simplifying assumptions that the lifetimes are exponentially distributed and that failures are independent, compute the MTTF of the system as a whole.

An example (2)



The sum of the failure rates is (failures are independent)

$$\text{Failure rate}_{\text{system}} = 10 \times \frac{1}{1{,}000{,}000} + \frac{1}{500{,}000} + \frac{1}{200{,}000} + \frac{1}{200{,}000} + \frac{1}{1{,}000{,}000}$$

$$= \frac{10+2+5+5+1}{1{,}000{,}000 \text{ hours}} = \frac{23}{1{,}000{,}000} = \frac{23{,}000}{1{,}000{,}000{,}000 \text{ hours}}$$

$$\text{MTTF}_{\text{system}} = \frac{1}{\text{Failure rate}_{\text{system}}} = \frac{1{,}000{,}000{,}000 \text{ hours}}{23{,}000} = 43{,}500 \text{ hours}$$

1 power supply, 200,000-hour MTTF
1 fan, 200,000-hour MTTF

$$\text{MTTF}_{\text{system}} < \frac{(\text{MTTF}_{\text{power\_supply}} + \text{MTTF}_{\text{fan}})}{2}$$

Benefits of redundancy (1)

The primary way to cope with failure is redundancy either:

in time (repeat the operation to see if it still is erroneous) or

in resources (have other components to take over from the one that failed).

Once the component is replaced and the system

is fully repaired, the dependability of the system is assumed to be as good as new.

Improve system dependability (1)

We can use redundant power supplies to improve dependability.

- We assume that the lifetimes of the components are exponentially distributed and that there is no dependency between the component failures.

MTTF for our redundant power supplies is the mean time until one power supply fails divided by the chance that the other will fail before the first one is replaced.

- Since we have two power supplies and independent failures, the mean time until one supply fails is $MTTF_{power\_supply}/2$.
- A good approximation of the probability of a second failure is MTTR over the mean time until the other power supply fails.

Improve system dependability (2)

$$\text{MTTF}_{\text{power supply pair}} = \frac{\text{MTTF}_{\text{power supply}}/2}{\frac{\text{MTTR}_{\text{power supply}}}{\text{MTTF}_{\text{power supply}}}} = \frac{\text{MTTF}^2_{\text{power supply}}/2}{\text{MTTR}_{\text{power supply}}} = \frac{\text{MTTF}^2_{\text{power supply}}}{2 \times \text{MTTR}_{\text{power supply}}}$$

If we assume it takes on average 24 hours to replace the power supply, and that

$\text{MTTF}_{\text{power\_supply}}$= 200,000-hour

The reliability of the fault tolerant pair of power supplies is

$$\text{MTTF}_{\text{power supply pair}} = \frac{\text{MTTF}^2_{\text{power supply}}}{2 \times \text{MTTR}_{\text{power supply}}} = \frac{200{,}000^2}{2 \times 24} \cong 830{,}000{,}000$$

4150 times more reliable than a single power supply

75