

IT Infrastructure Architecture

Infrastructure Building Blocks
and Concepts

Compute

Midrange systems

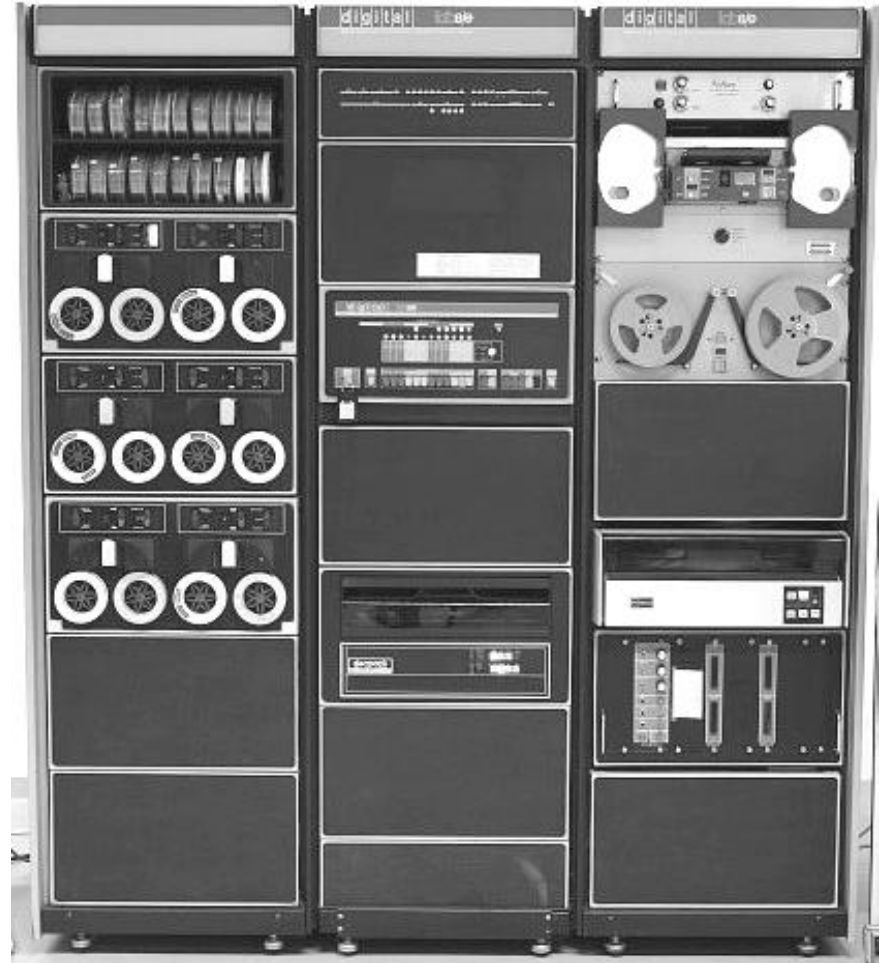
- The midrange platform is positioned between the mainframe platform and the x86 platform
- Built using parts from only one vendor, and run an operating system provided by that same vendor
- This makes the platform:
 - Stable
 - High available
 - Secure

Midrange systems

- Today midrange systems are produced by three vendors:
 - IBM
 - Power Systems series
 - Operating system: AIX UNIX, Linux and IBM i
 - Hewlett-Packard
 - HP Integrity systems
 - Operating system: HP-UX UNIX and OpenVMS
 - Oracle
 - Sun Microsystems's based SPARC servers
 - Operating system: Solaris UNIX

Midrange systems - History

- Minicomputer evolved in the 1960s as small computers that became possible with the use of IC and core memory technologies
- Small was relative
 - A single minicomputer typically was housed in a few cabinets the size of a 19" rack
- The first commercially successful minicomputer was DEC PDP-8, launched in 1964



Midrange systems - History

- Minicomputers became powerful systems
 - They ran full multi-user, multitasking operating systems like OpenVMS and UNIX
- In the 1980s, minicomputers (a.k.a. midrange systems) became less popular
 - A result of the lower cost of PCs, and the emergence of LANs
- Still used in places where high availability, performance, and security are very important

Midrange systems - Architecture

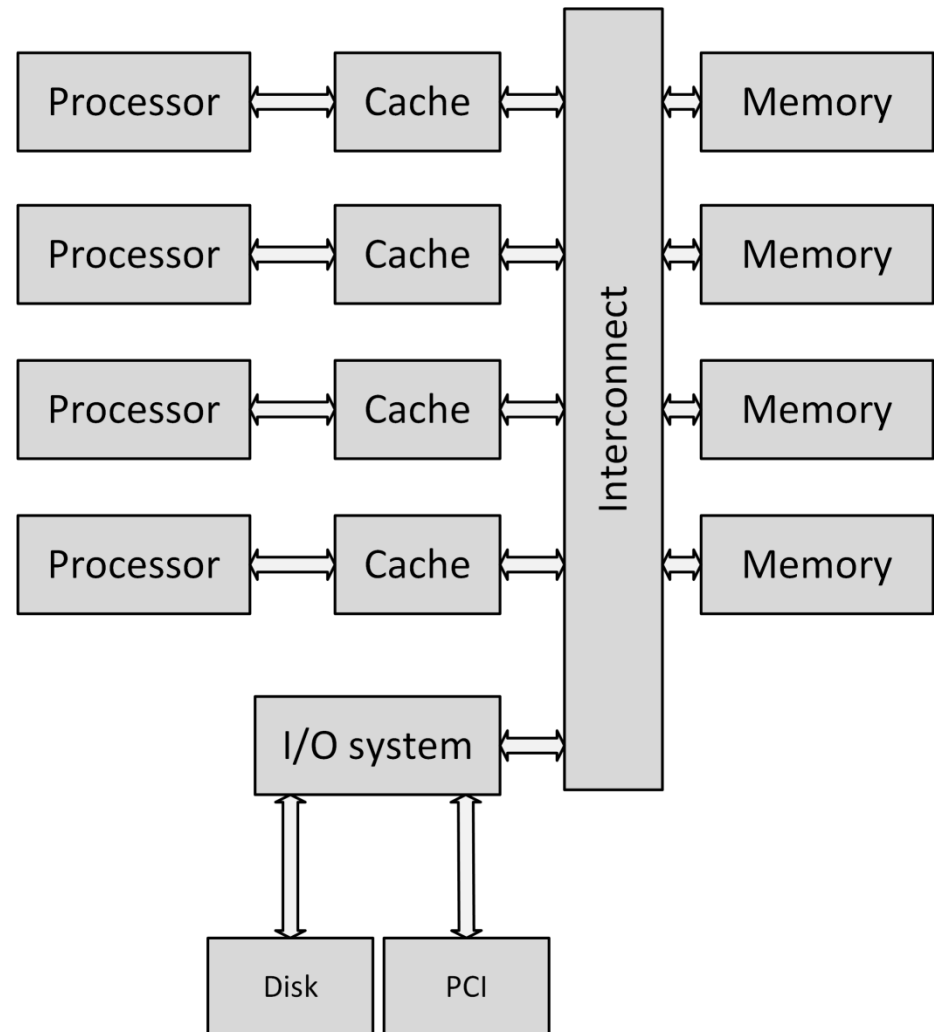
- The architecture of most midrange systems:
 - Uses multiple CPUs
 - Based on a shared memory architecture
- In a shared memory architecture, all CPUs in the system can access all installed memory blocks
 - Changes made in memory by one CPU are immediately seen by all other CPUs
 - A shared bus connects all CPUs and all RAM
 - The I/O system is also connected to the interconnection network

Midrange systems - Architecture

- Shared memory architectures come in two flavors:
 - Uniform Memory Access (UMA)
 - Non-Uniform Memory Access (NUMA)
 - Cache coherence is needed
 - If one CPU writes to a location in shared memory, all other CPUs must update their caches to reflect the changed data
 - Cache coherent versions are known as ccUMA and ccNUMA

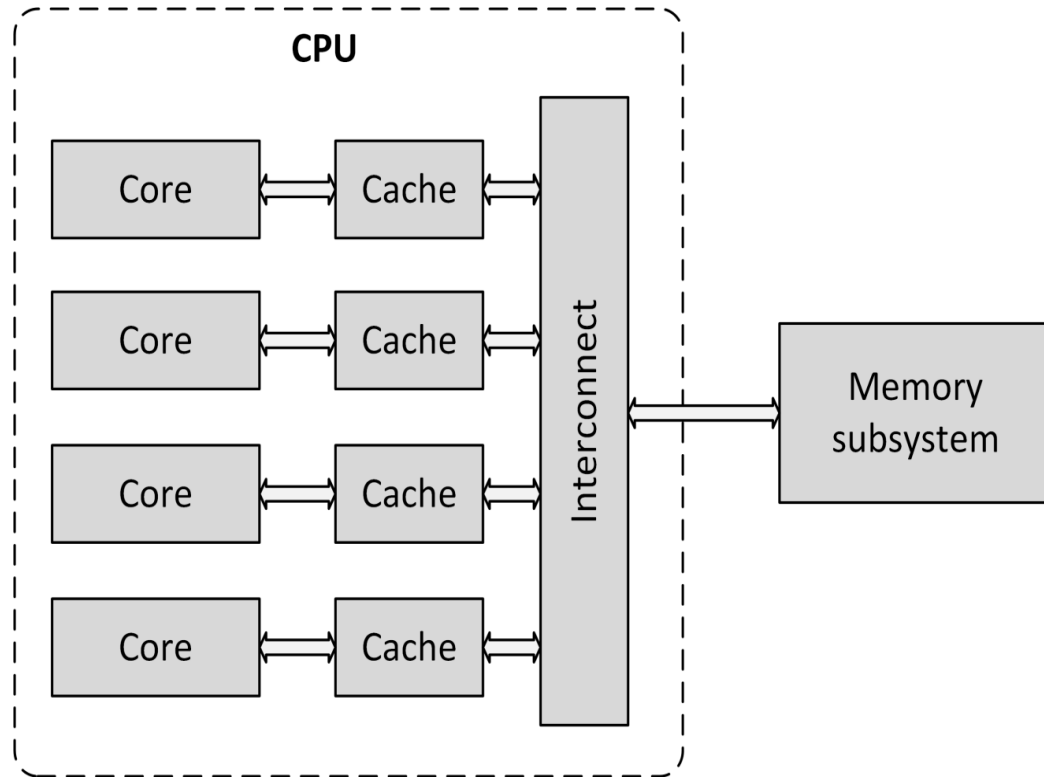
Midrange systems – UMA architecture

- The UMA architecture is one of the earliest styles of multi-CPU architectures, typically used in systems with no more than 8 CPUs
- The machine is organized into a series of nodes containing either a processor, or a memory block
- Nodes are interconnected, usually by a shared bus



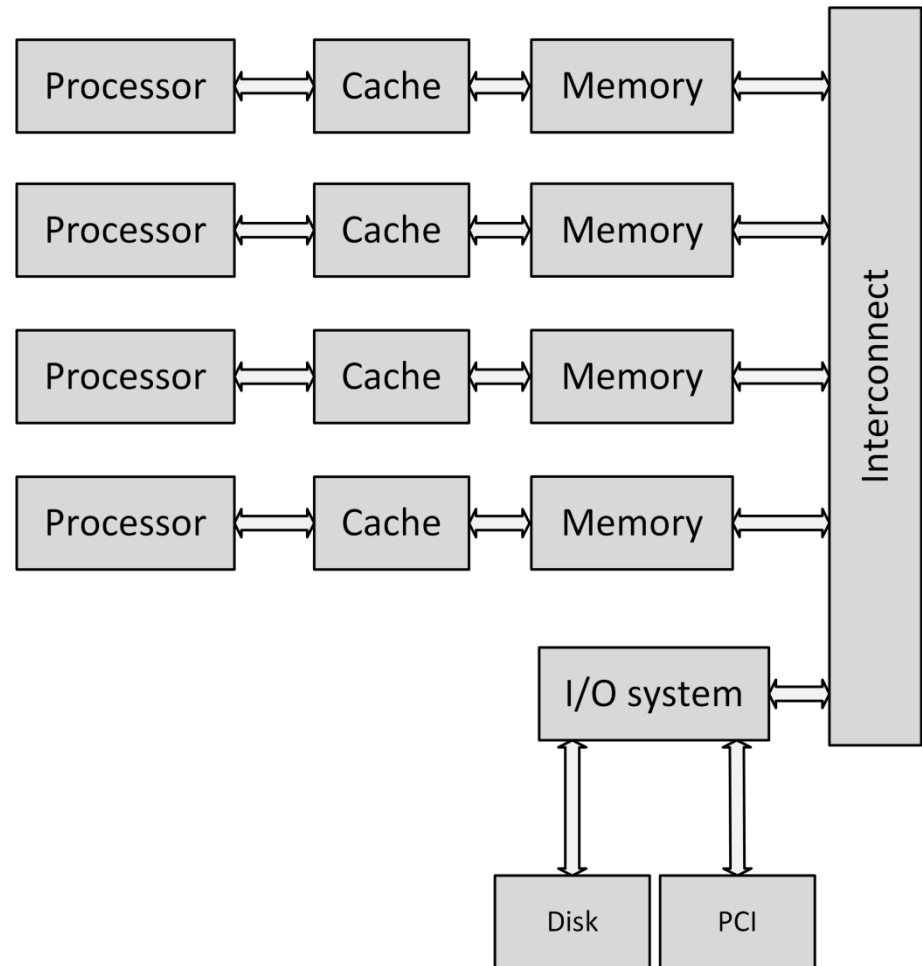
Midrange systems – SMP architecture

- UMA systems are also known as Symmetric Multi-Processor (SMP) systems
- SMP is used in x86 servers as well as early midrange systems
- Can be Implemented inside multi-core CPUs
 - The interconnect is implemented on-chip in the CPU
 - A single path to the main memory is provided between the chip and the memory subsystem



Midrange systems – NUMA architecture

- NUMA is a computer architecture in which the machine is organized into a series of nodes
- Each node contains processors and memory
- Nodes are interconnected using a crossbar interconnect
- When a processor accesses memory not within its own node, data must be transferred over the interconnect
 - Slower than accessing local memory
 - Memory access times are **non-uniform**



Midrange virtualization

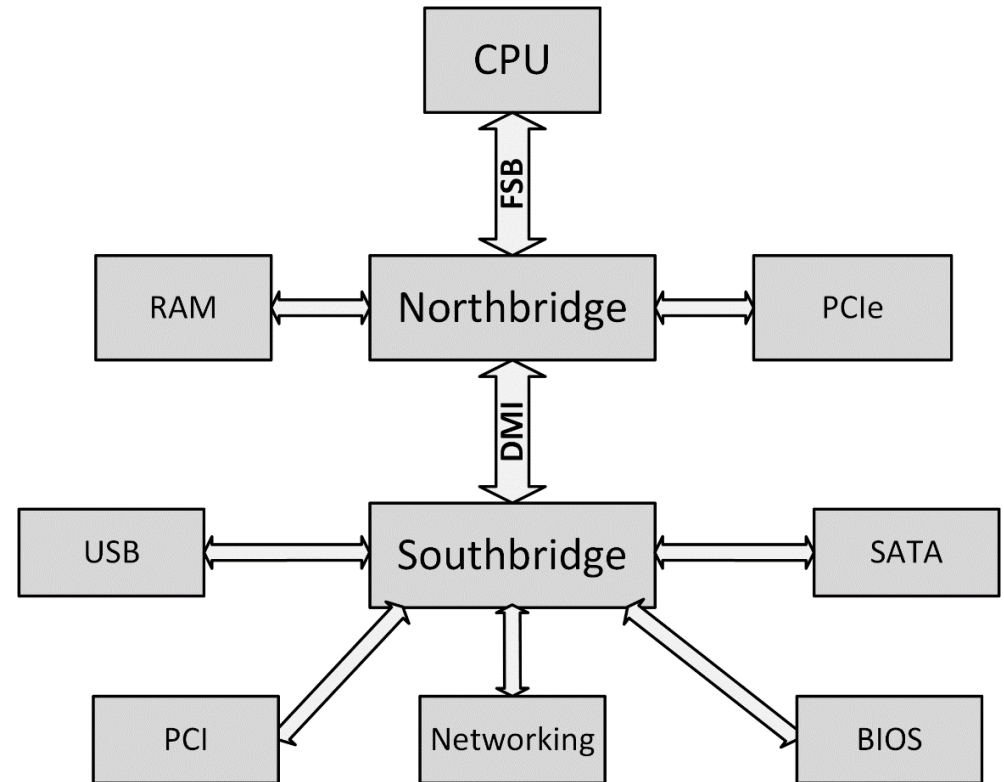
- Most midrange platform vendors provide virtualization based on LPARs
- LPARS are a type of hardware partitioning
 - IBM AIX: Workload/Working Partitions (WPARs)
 - HP: nPARs
 - Oracle Solaris: zones and containers

x86 servers

- The x86 platform is the most dominant server architecture today
- In the 1990s, x86 servers first started to appear.
 - They were basically big PCs, housed in 19” racks without dedicated keyboards and monitors
- x86 servers are produced by many vendors, like:
 - HP
 - Dell
 - Lenovo
 - HDS (Hitachi Data Systems)
 - Huawei
- Implementation of the platform is highly dependent on the vendor and the components available at a certain moment
- Most used operating systems are Microsoft Windows and Linux

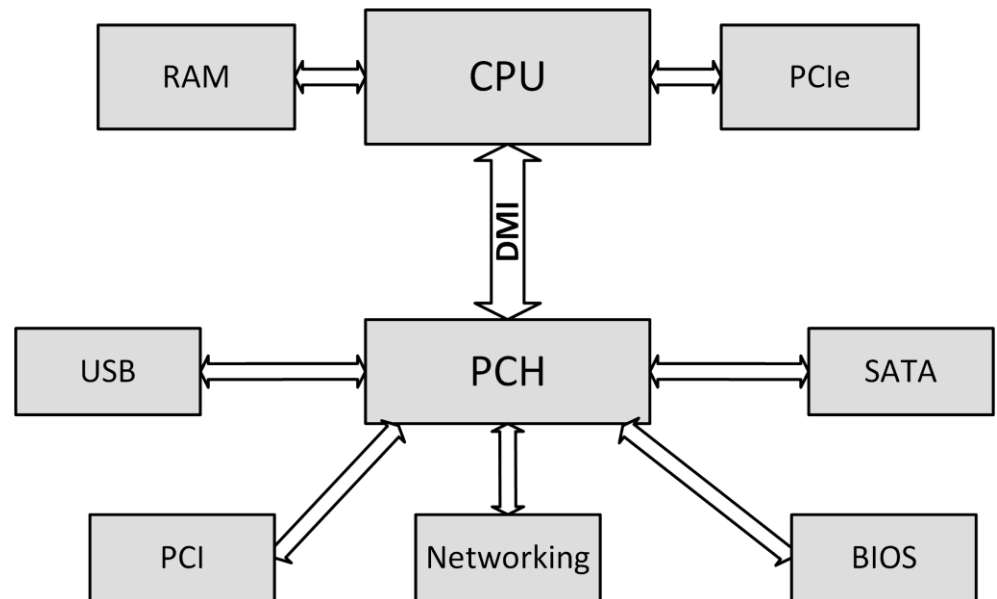
x86 servers - Architecture

- x86 architectures are defined by a CPU from the x86 family, and building blocks, integrated in a number of specialized chips, known as an x86 chipset
- Earlier x86 systems utilized a Northbridge / Southbridge architecture
 - Front Side Bus (FSB)
 - Direct Media Interface (DMI)



x86 servers - Architecture

- In the PCH architecture, the RAM and PCIe data paths are directly connected to the CPU
- The Northbridge integrated in the CPU
- Intel introduces new architectures and chipsets roughly every two years
- Now full system on a chip (SoC)
- SOC's directly expose:
 - PCIe lanes
 - SATA
 - USB
 - High Definition Video



x86 virtualization

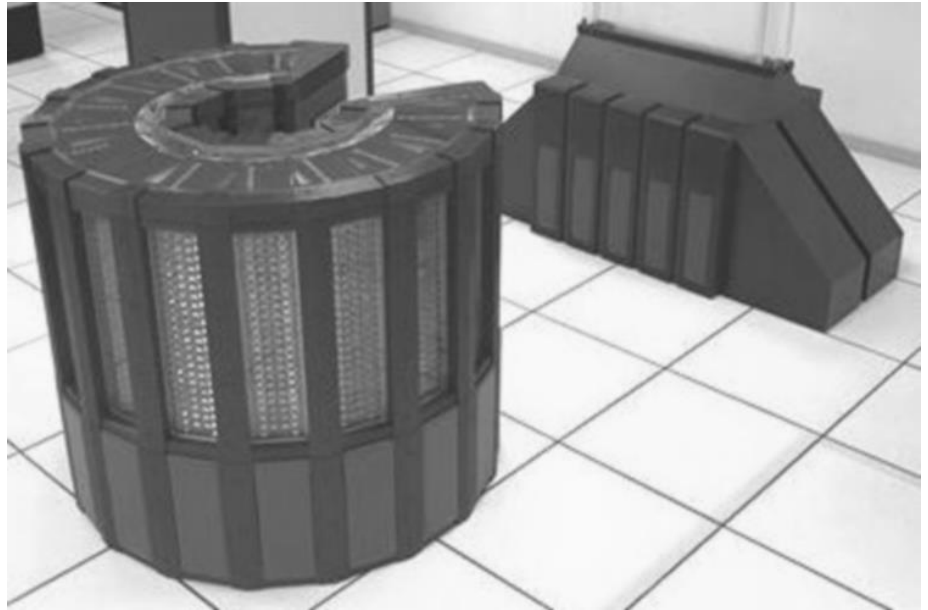
- On x86 platforms, most servers only run one application each
 - A Windows server running Exchange will probably not also run SharePoint
 - A Linux server running Apache will probably not also run MySQL
 - This is the main reason x86 systems use their hardware much less efficient than midrange systems
- By running multiple operating systems – each in one virtual machine – on a large x86 server, resource utilization can be improved
- The most used products for virtualization on the x86 platform are:
 - VMware vSphere
 - Microsoft's Hyper-V
 - Citrix XenServer
 - Oracle VirtualBox
 - Red Hat RHEV

Supercomputers

- A supercomputer is a computer architecture designed to maximize calculation speed
 - This in contrast with a mainframe, which is optimized for high I/O throughput
- Supercomputers are the fastest machines available at any given time
- Used for highly compute-intensive tasks requiring floating point calculations, like:
 - Weather forecast calculations
 - Oil reservoir simulations
 - Rendering of animation movies

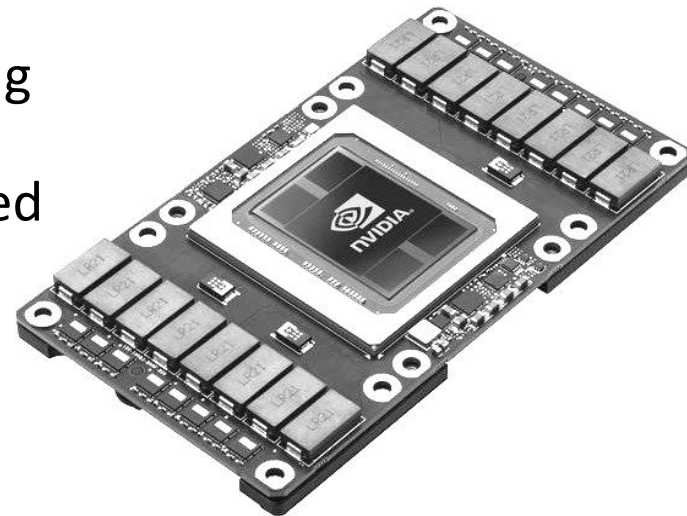
Supercomputers

- Originally, supercomputers were produced primarily by a company named Cray Research
 - Cray-1 (1976): 250 MFLOPS (Million Floating Point Operations per second)
 - Cray-2 (1985): 1,900 MFLOPS



Supercomputers

- Nowadays high performance computing is done mainly with large arrays of x86 systems
- Intel's Core i7 5960X CPU has a peak performance of 354,000 MFLOPS
- The fastest computer array is a cluster with more than 10,000,000 CPU cores, calculating at 125,435,000,000 MFLOPS, running Linux
- Graphics processing units (GPUs) can be used together with CPUs to accelerate specific calculations
- A GPU has a massively parallel architecture consisting of thousands of small, efficient cores designed for handling multiple tasks simultaneously
 - The NVIDIA Tesla GP100 GPU, introduced in 2016, has 3840 cores



Compute availability

Hot swappable components

- Hot swappable components are server components that can be installed, replaced, or upgraded while the server is running
 - Memory
 - CPUs
 - Interface cards
 - Power supplies
- The virtualization and operating systems using the server hardware must be aware that components can be swapped on the fly
 - For instance, the operating system must be able to recognize that memory is added while the server operates and must allow the use of this extra memory without the need for a reboot

Parity memory

- To detect memory failures, parity bits can be used as the simplest form of error detecting code
- Parity bits enable the detection of data errors
- They cannot correct the error, as it is unknown which bit has flipped

DATA	PARITY
1001 0110	0
1011 0110	1
0001 0110	0 -> ERROR: parity bit should have been 1!

ECC memory

- ECC memory not only detects errors, but is also able to correct them
- ECC Memory chips use Hamming Code or Triple Modular Redundancy (TMR) as the method of error detection and correction
- Memory errors are proportional to the amount of RAM in a computer as well as the duration of operation
 - Since servers typically contain many GBs of RAM and are in operation 24 hours a day, the likelihood of memory errors is relatively high and hence they require ECC memory

Lockstepping

- Lockstepping is an error detection and correction technology for servers
- Multiple systems perform the same calculation, and the results of the calculations are compared
 - If the results are equal, the calculations were correctly performed
 - If there are different outcomes, one of the servers made an error
- Very expensive technology
 - Only used in systems that require extremely high reliability

Virtualization availability

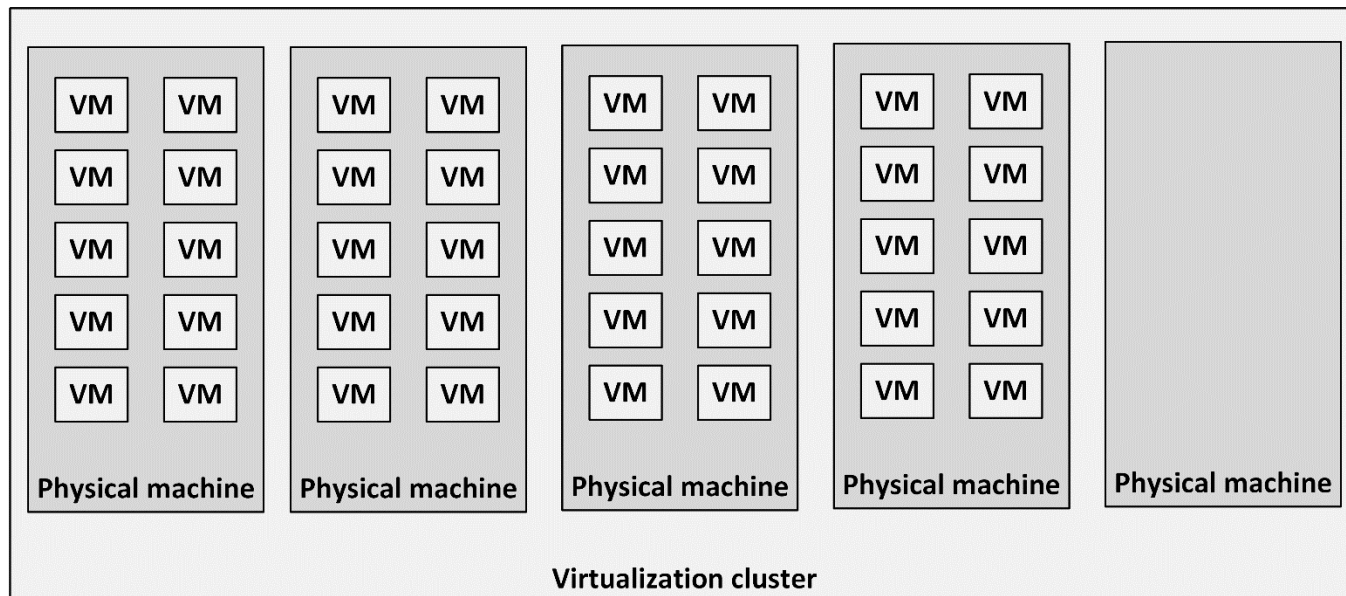
- All virtualization products provide failover clustering
 - When a physical machine fails, the virtual machines running on that physical machine can be configured to restart automatically on other physical machines
 - When a virtual machine crashes, it can be restarted automatically on the same physical machine

Virtualization availability

- The virtualization layer has no knowledge of the applications running on the virtual machine's operating system
 - Failover clustering on virtualization level can only protect against two situations:
 - A physical hardware failure.
 - An operating system crash in a virtual machine

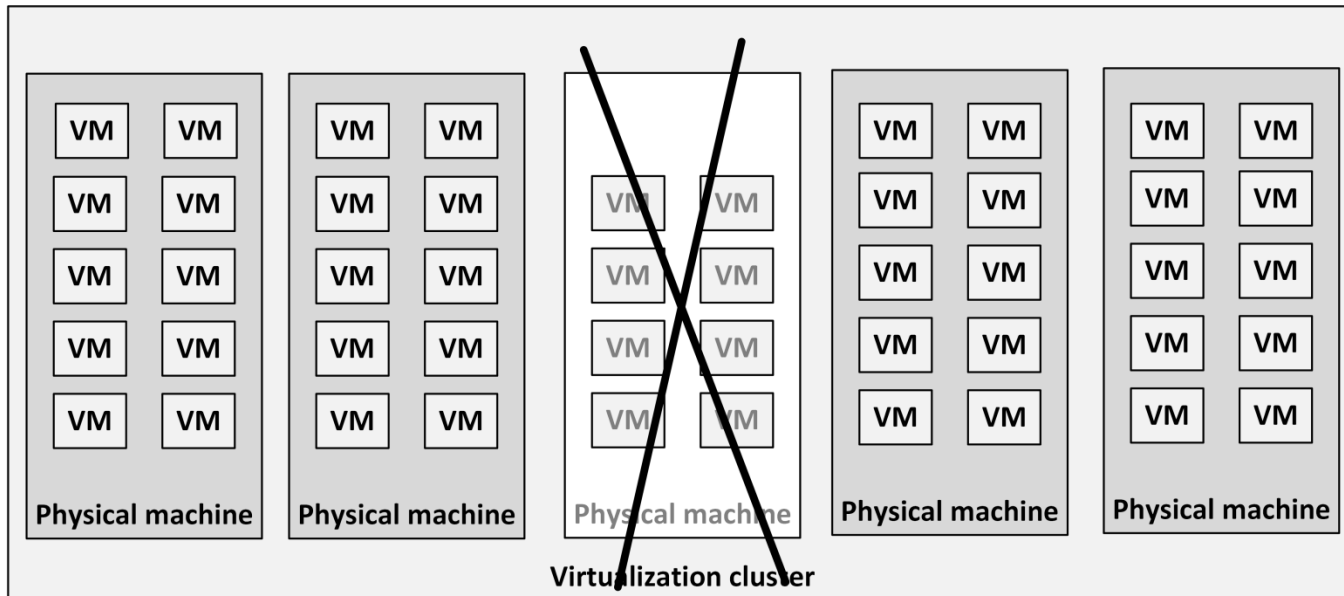
Virtualization availability

- To cope with the effects of a failure of a physical machine, a spare physical machine is needed
 - All hypervisors are placed in a virtualization cluster
 - The hypervisors on the physical machines check the availability of the other hypervisors in the cluster
 - One physical machine is running as a spare to take over the load of any failing physical machine



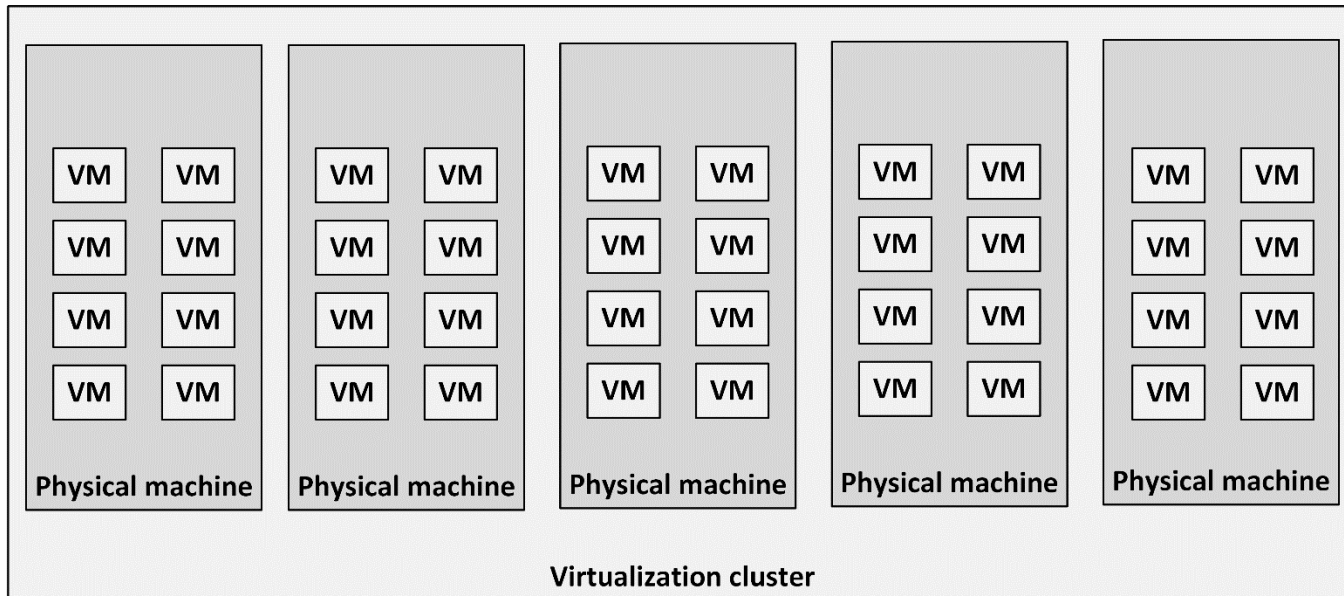
Virtualization availability

- When a physical machine fails, the virtual machines running on it are automatically restarted on the spare physical machine



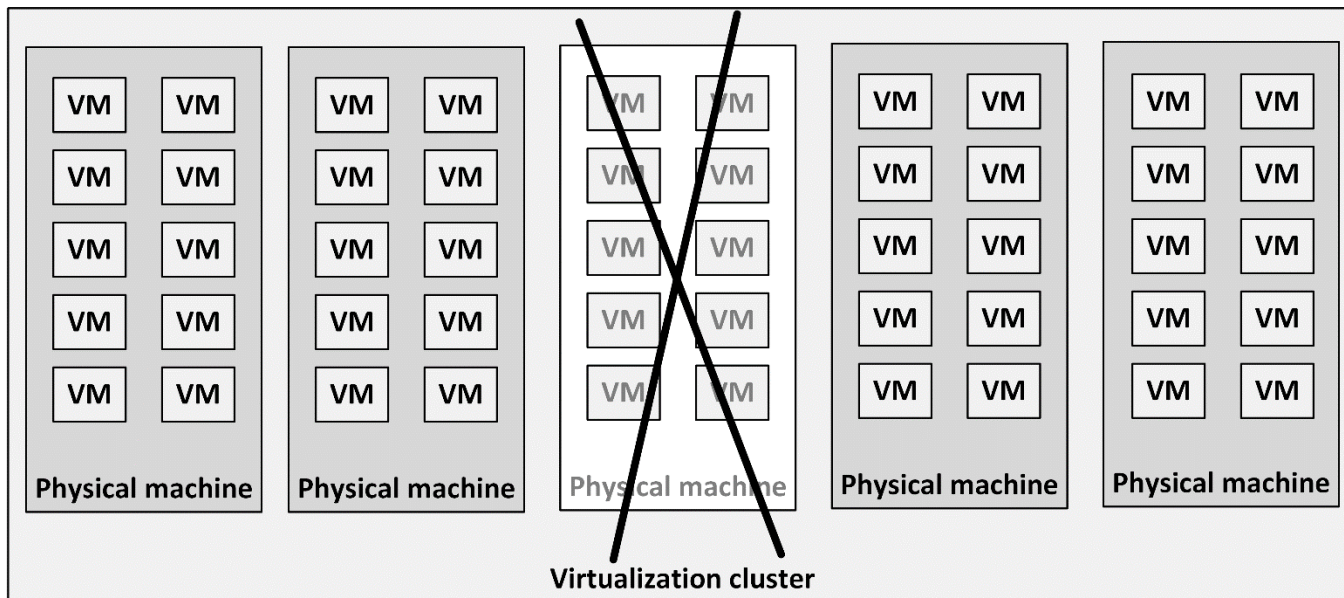
Virtualization availability

- An alternative is to have all physical machines running at lower capacity



Virtualization availability

- When a physical machine fails, the virtual machines running on it are automatically restarted on the other physical machine
- All machines now run on full capacity



Compute performance

CPU: Moore's law

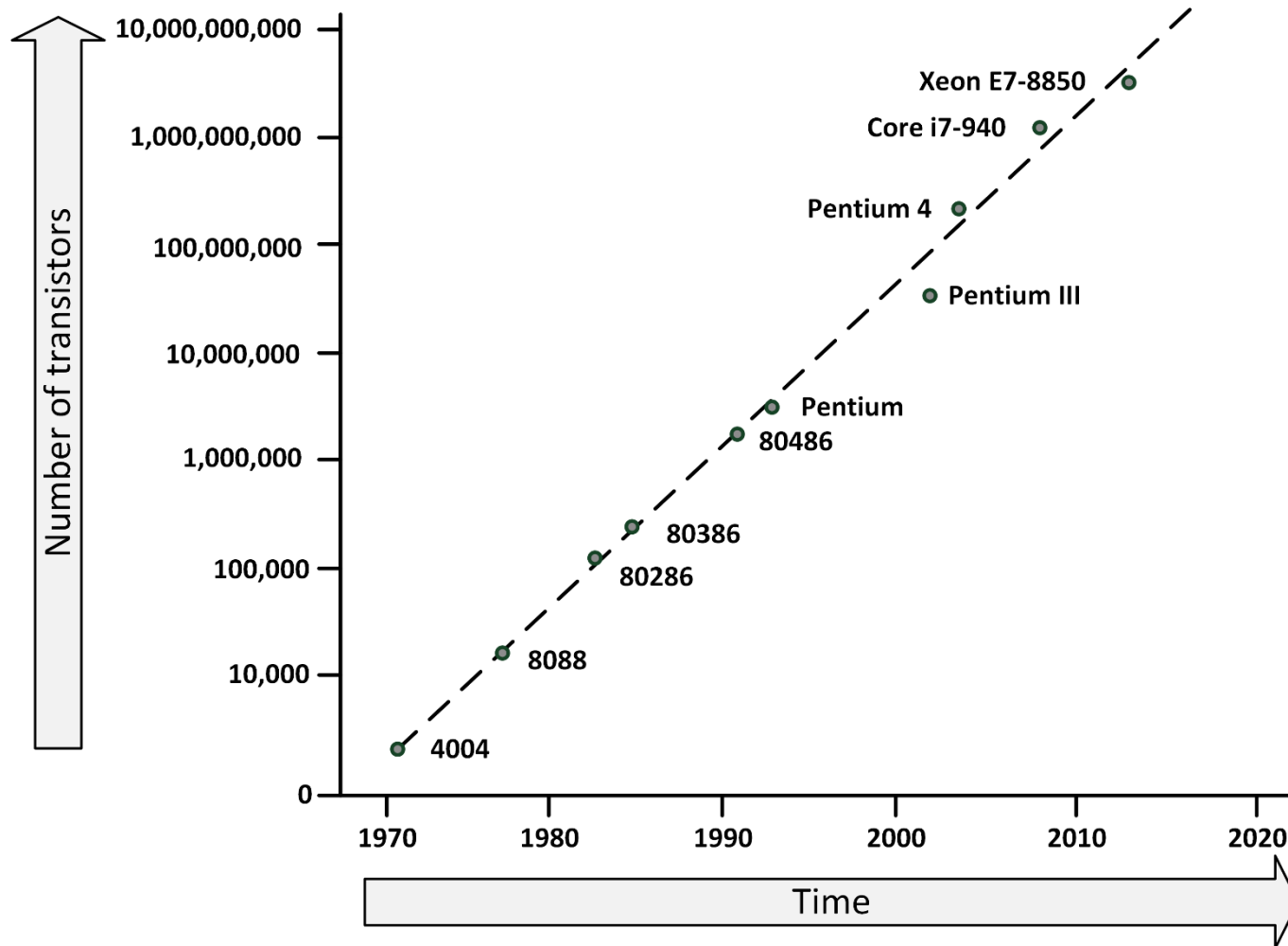
- In 1971, Intel released the world's first universal microprocessor, the 4004
 - Contained 2,300 transistors
 - Could perform 60,000 instructions per second
 - About as much as the ENIAC computer that filled a complete room and weighed several tons
- Since the introduction of the first CPU in 1971, the power of CPUs has increased exponentially



CPU: Moore's law

- Moore's law states:
 - The number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years
- This trend has continued for more than half a century now
- An Intel Broadwell-EP Xeon in 2017 contains 7,200,000,000 transistors
- An 3,100,000-fold increase in 45 years' time!

CPU: Moore's law



Please note that the vertical scale is logarithmic instead of linear, showing a 10-fold increase of the number of transistors in each step

CPU: Moore's law

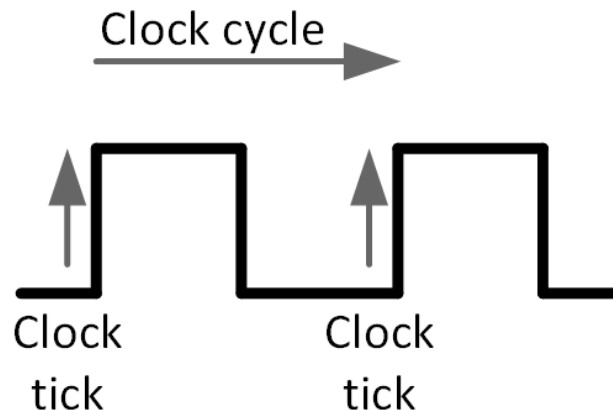
- Moore's law only speaks of the number of transistors; not the performance of the CPU
- The performance of a CPU is dependent on a number of variables, like:
 - Clock speed
 - Use of caches and pipelines
 - Width of the data bus
- Moore's law cannot continue forever, as there are physical limits to the number of transistors a single chip can hold
 - In 2017, connections used inside a high-end CPU had a physical width of 14 nm (nanometer), the size of 140 atoms
 - in 2020, 5 nm CPUs will be produced; traces on the chip are just 50 atoms wide

CPU: Increasing CPU and memory performance

- Various techniques have been invented to increase CPU performance, like:
 - Increasing the clock speed
 - Caching
 - Prefetching
 - Branch prediction
 - Pipelines
 - Use of multiple cores

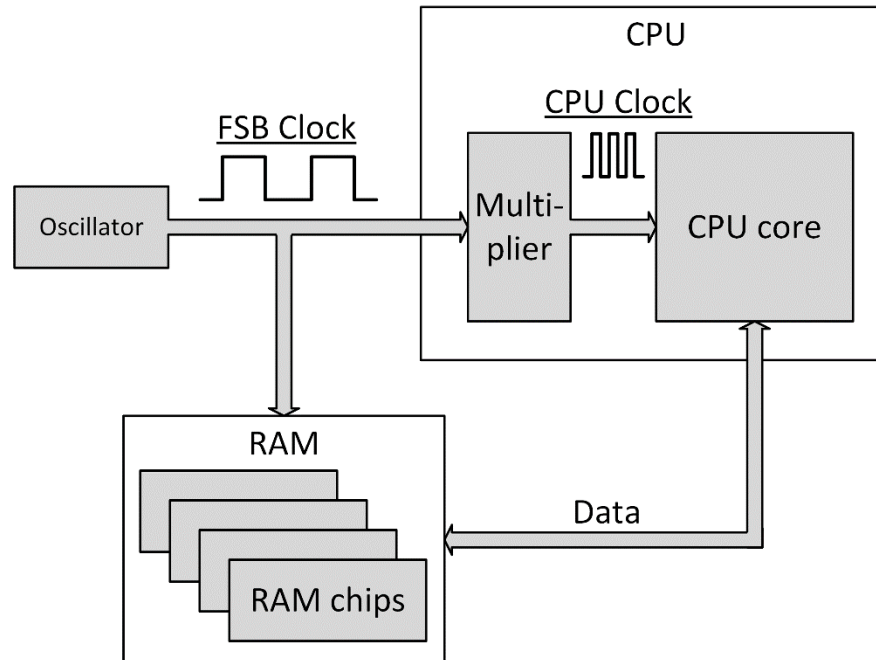
CPU: Increasing clock speed

- CPU instructions need to be fetched, decoded, executed, and the result must often be written back to memory
- Each step in the sequence is executed when an external clock signal is changed from 0 to 1 (the clock tick)
- The clock signal is supplied to the CPU by an external oscillator
- In the early years, CPU clock speed was the main performance indicator
- CPU clock speed is measured in Hertz (Hz) – clock ticks or cycles per second



CPU: Increasing clock speed

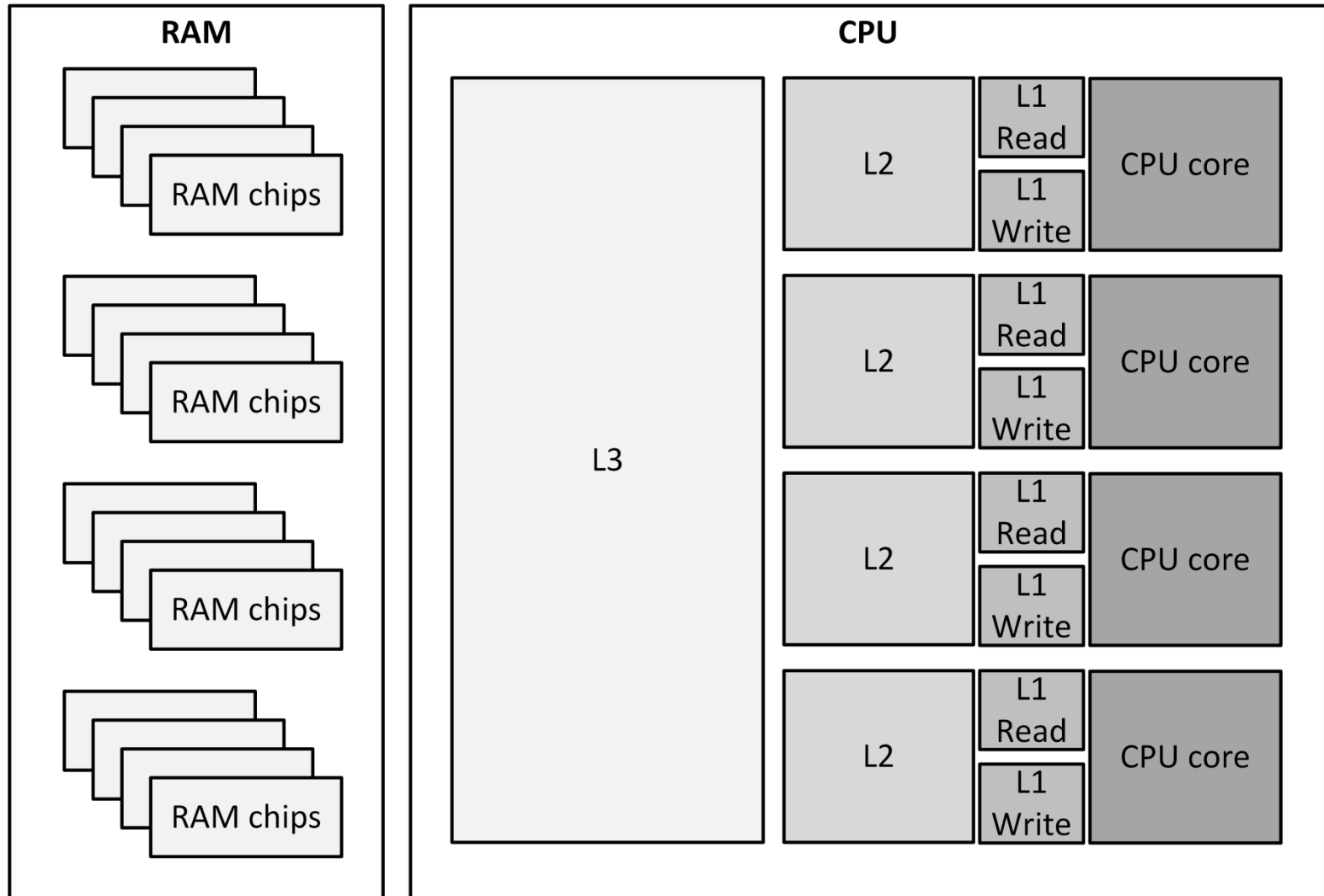
- Today's CPUs use clock speeds as high as 3 GHz (3,000,000,000 clock ticks per second)
- Because of physical limitations, oscillators cannot run at this speed
- An oscillator with a lower frequency is used (for instance 400 MHz) and this clock rate is multiplied on the CPU chip
- The oscillator speed is known as the Front Side Bus (FSB) speed



CPU: Caching

- All CPUs in use today contain on-chip caches
- A cache is a relatively small piece of high speed static RAM on the CPU
 - Temporarily stores data received from slower main memory
 - Most CPUs contain two types of cache: level 1 and level 2 cache
 - Some multi-core CPUs also have a large level 3 cache; a cache shared by the cores
- Cache memory runs at full CPU speed (say 3 GHz), main memory runs at the CPU external clock speed (say 100 MHz, which is 30 times slower)

CPU: Caching



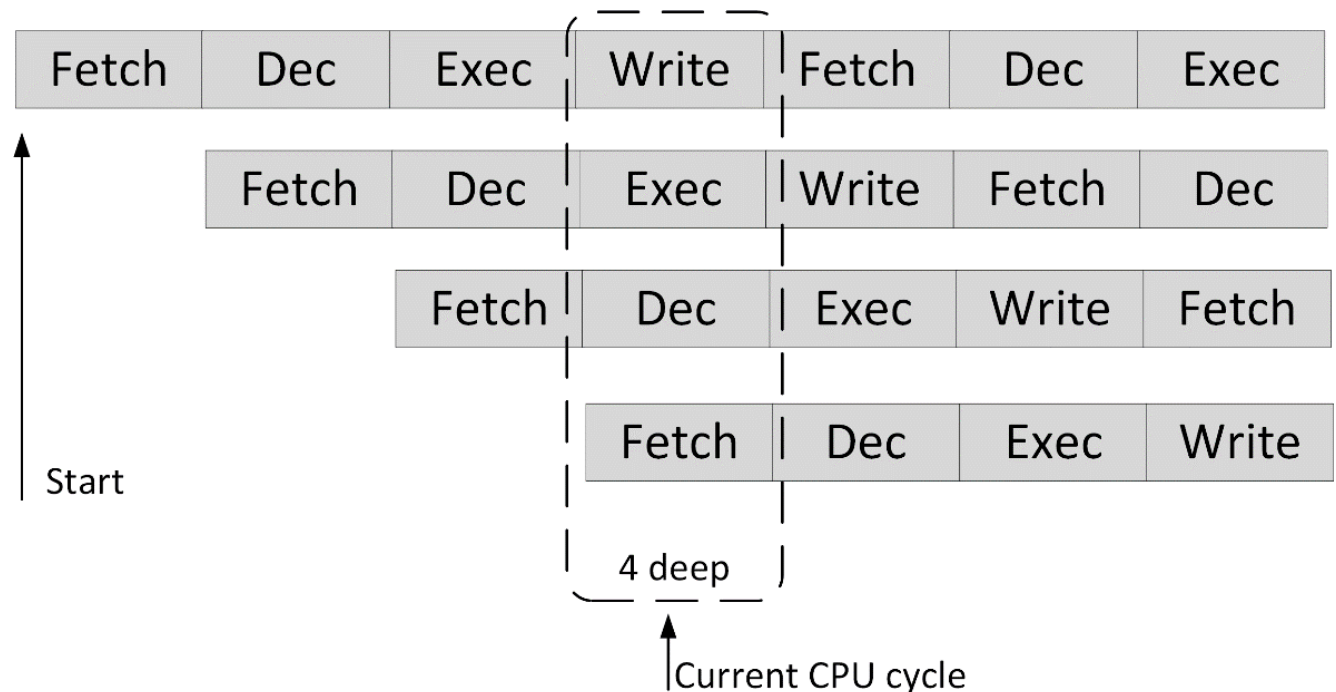
CPU: Pipelines

- Early processors first fetched an instruction, decoded it, then executed the fetched instruction, and wrote the result back before fetching the next instruction and starting the process over again



CPU: Pipelines

- Later CPUs used pipelines
- While the first instruction is being executed, the second instruction can be fetched (since that circuitry is idling anyway), creating instruction overlap



CPU: Prefetching and branch prediction

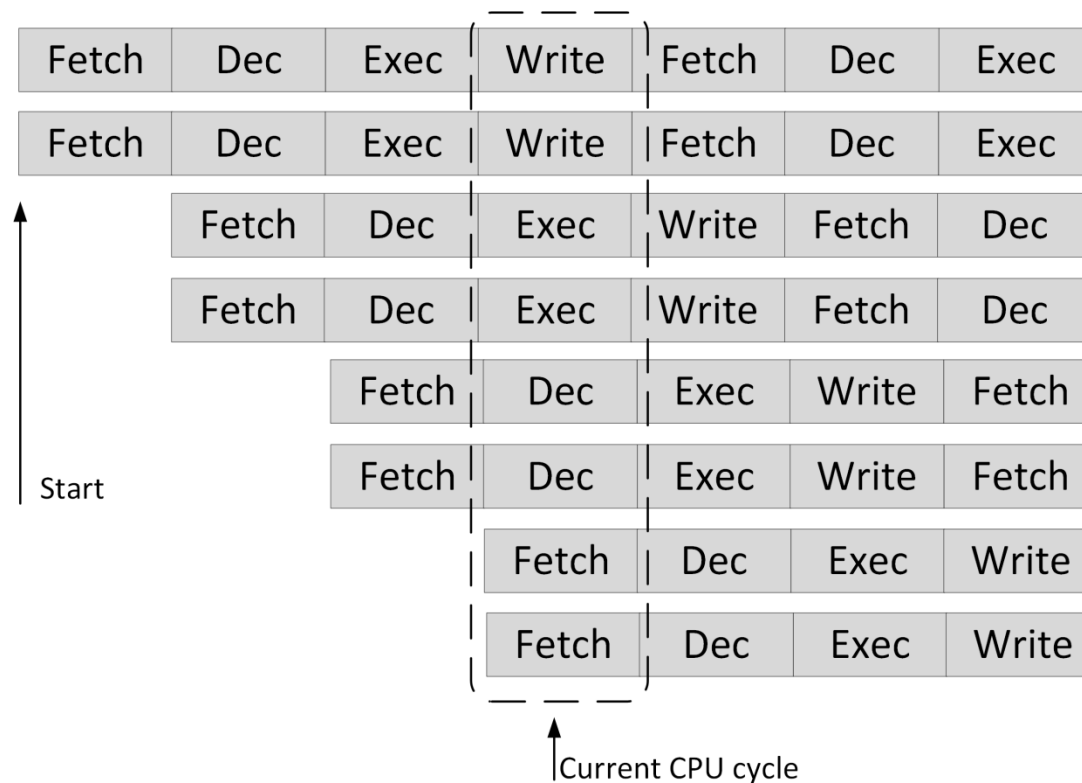
- Prefetching:
 - When an instruction is fetched from main memory, also the next instructions are fetched and stored in cache
 - When the CPU needs the next instruction it is already available in cache
- Unfortunately, most programs contain jumps (also known as branches), resulting in cache misses
 - The next instruction is not the next instruction in memory

CPU: Prefetching and branch prediction

- The cache system tries to predict the outcome of branch instructions before they are executed by the CPU (called branch prediction)
 - In practice more than 80% of the processor instructions are delivered to the CPU from cache memory using prefetching and branch prediction

CPU: Superscalar CPUs

- A superscalar CPU can process more than one instruction per clock tick
- This is done by simultaneously dispatching multiple instructions to redundant functional units on the processor



CPU: Multi-core CPUs

- The fastest commercial CPUs have been between running between 3 GHz and 4 GHz for a number of years now
- Reasons:
 - High clock speeds make connections on the circuit board work as a radio antenna
 - A frequency of 3 GHz means a wavelength of 10 cm. When signals travel for more than a few cm on a circuit board, the signal gets out of phase with the clock
 - The CPU can heat up tremendously at certain spots, which could lead to a meltdown

CPU: Multi-core CPUs

- A multi-core processor is a CPU with multiple separate cores
 - The equivalent of getting multiple processors in one package
- The cores in a multi-core CPU run at a lower frequency
 - Reduce power consumption
 - Reduce heat (no hot spots)
- Trend is to have CPUs with tens or even hundreds of cores

CPU: Hyper-threading

- Certain Intel CPUs contain a propriety technology called hyper-threading
 - For example the Core i3/i5/i7 and Xeon CPUs
- Hyper-threading makes a single processor core virtually work as a multi-core processor
- Hyper-threading can provide some increase in system performance by keeping the processor pipelines busier

Virtualization performance

- Consolidating multiple virtual machines on one physical machine increases CPU usage and reduces CPU idle time
 - This is a primary driver for the use of virtualization
- The physical machine needs to handle the disk and network I/O of *all* running virtual machines
 - This can easily lead to an I/O performance bottleneck

Virtualization performance

- When choosing a physical machine to host virtual machines, consider getting a machine with:
 - Much CPU and memory capacity
 - Capability of very high I/O throughput
- Virtualization introduces performance penalties:
 - Resources required to run the hypervisor
 - Operation transformations
- This is usually less than 10% of the total performance

Virtualization performance

- Databases generally require a lot of network bandwidth and high disk I/O performance
 - This makes databases less suitable for a virtualized environment
 - Raw Device Mapping allows a virtual machine exclusive access to a physical storage medium
 - This diminishes the performance hit of the hypervisor on storage to almost zero

Virtualization performance

- Often one physical server is needed per database
 - The physical server runs just one virtual machine
 - Many benefits of virtualization remain
 - Database servers can easily be migrated to other physical machines without downtime
 - Management of the servers is unified when all servers run hypervisors

Compute security

Physical security

- Disable external USB ports in the BIOS
- BIOS settings in an x86 server should be protected using a password,
 - Via the BIOS, external USB ports can be enabled, and other parameters can be set
- Some servers allow the detection of the physical opening of the server housing
 - Such an event can be sent to a central management console using for instance SNMP traps
 - If possible, enable this to detect unusual activities

Virtualization security

- Virtual machines must be protected the same way as physical machines
- The use of virtualization introduces new security vulnerabilities of its own:
 - If possible, firewalls and Intrusion Detection Systems (IDSs) in the hypervisor should be deployed
 - The virtualization platform itself needs patching too
 - The size and complexity of the hypervisor should be kept to a minimum
- DMZ
 - Consider using separate physical machines that run all the virtual machines needed in the DMZ

Virtualization security

- Systems management console
 - The systems management console connects to all hypervisors and virtual machines
 - When the security of the systems management console is breached, security is breached on all virtual machines
 - Not all systems managers should have access to all virtual machines
 - Special user accounts and passwords should be configured for high risk operations like shutting down physical machines or virtualized clusters
 - All user activity in the systems management console should be logged