

Computer Architecture
Lecture Week 3 (Part I)

Von Neumann and Harvard Architectures

Performance Metrics of Computers

Artem Burmyakov

September 19, 2025

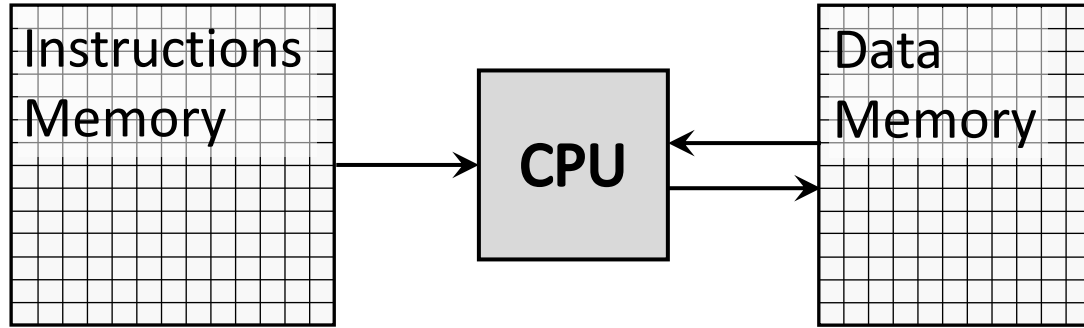


**INNOPOLIS
UNIVERSITY**

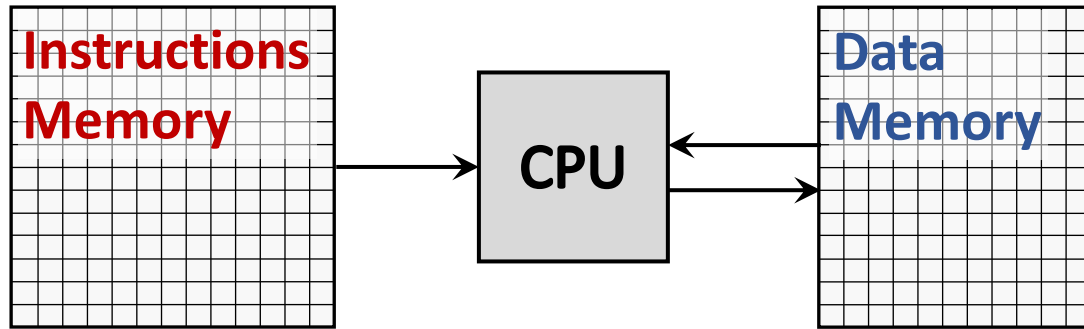
Recap: CA Fundamental Ideas

1. Memory hierarchy
2. Design abstraction
3. Moore's law
4. Parallelism
5. Pipelining
6. Speculation (prediction)
7. Dependability via redundancy
8. A faster common case
9. Finite State Machines

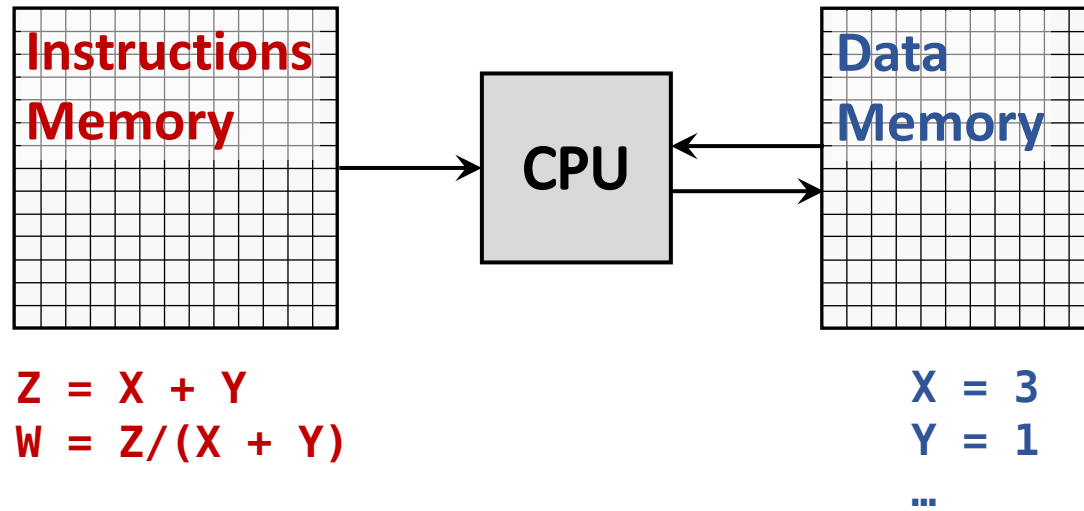
Harvard Computer Architecture



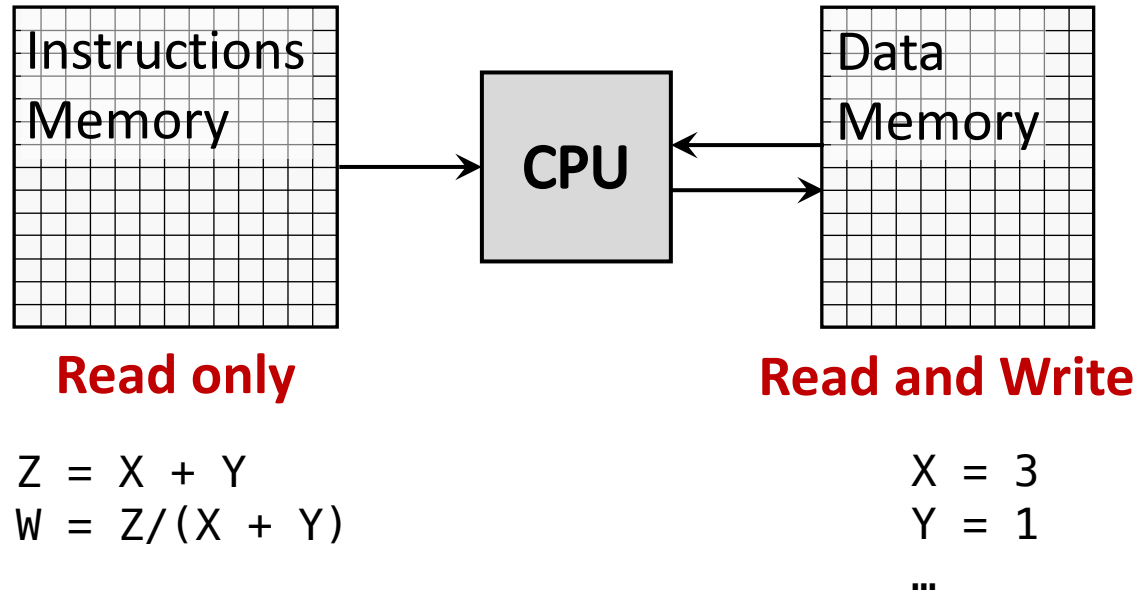
Harvard Architecture



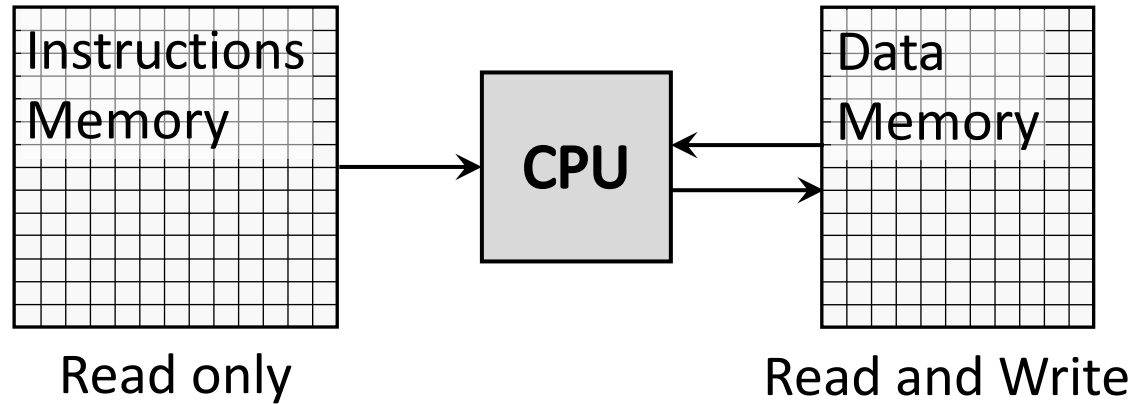
Harvard Architecture



Harvard Architecture

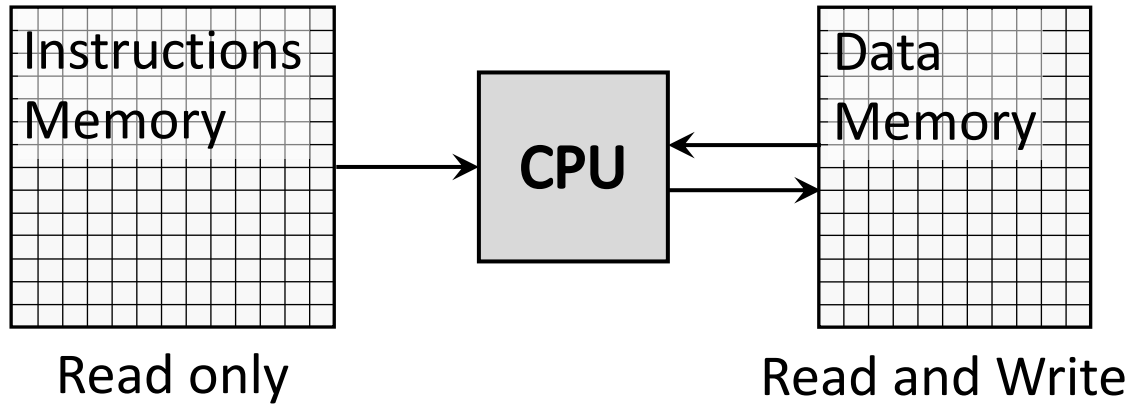


Harvard Architecture



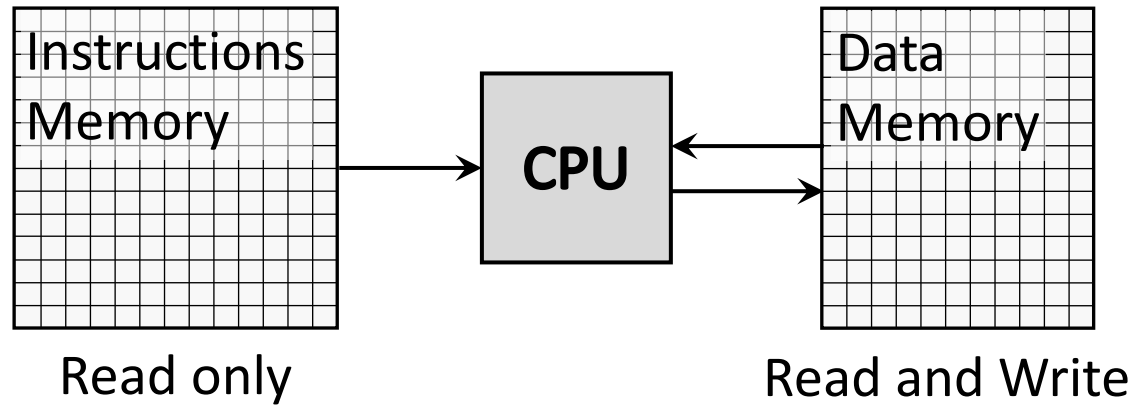
Principle	Instructions and data are in different physical memory units (or address spaces with dedicated buses)
-----------	---

Harvard Architecture



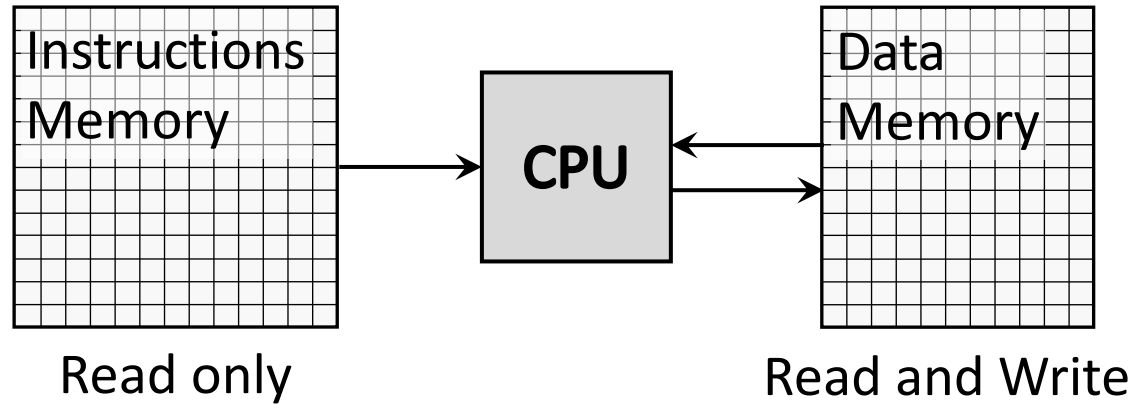
Principle	Instructions and data are in different physical memory units
Pros	+ Lower performance overheads: Instructions and data are fetched concurrently

Harvard Architecture



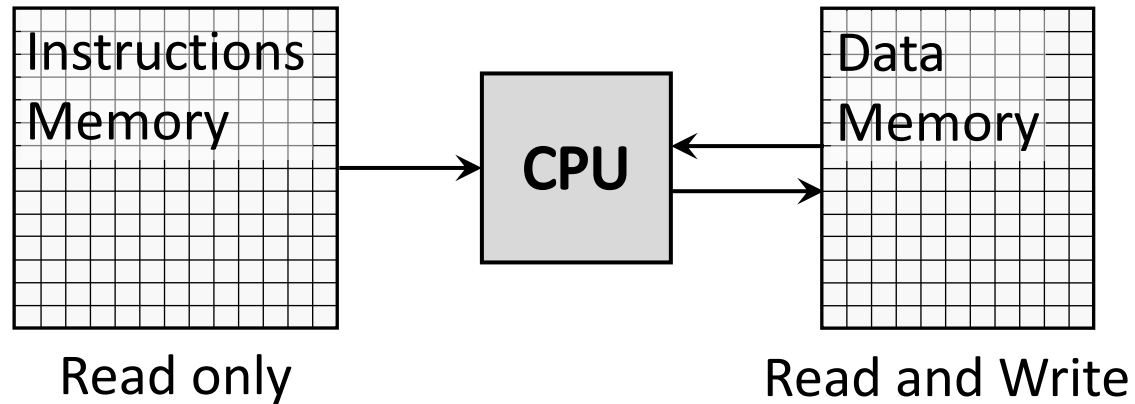
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation

Harvard Architecture



Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

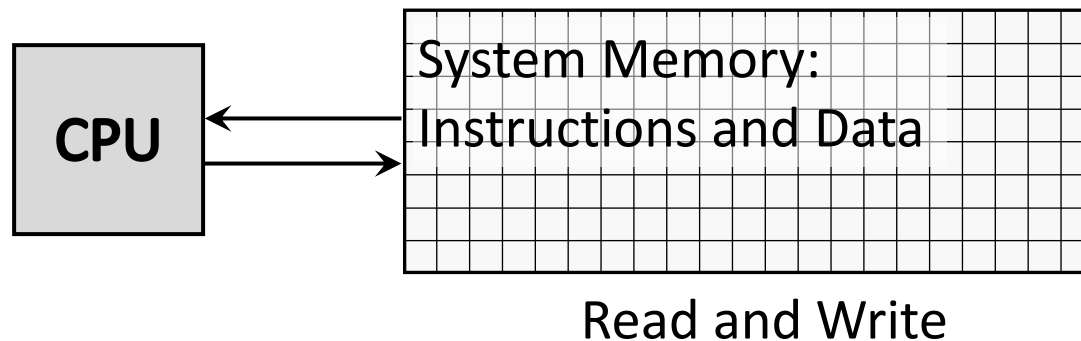
Harvard Architecture



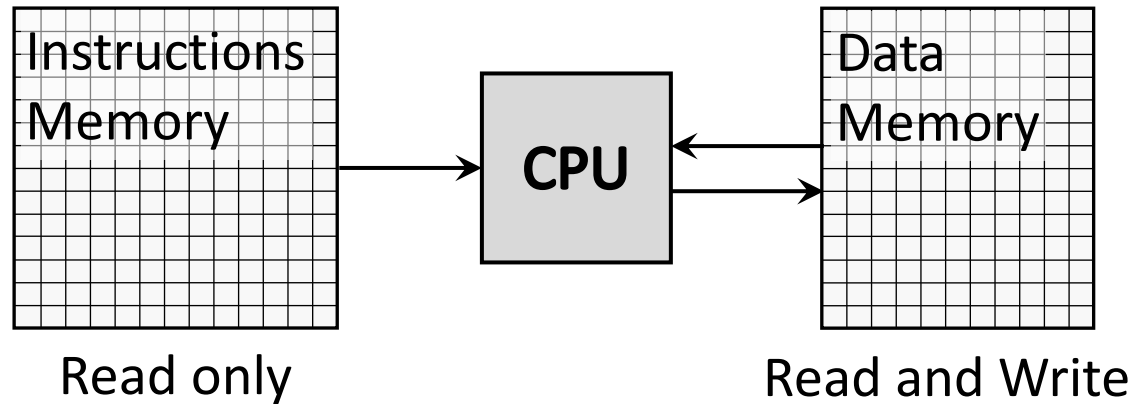
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Von Neumman Architecture

(or Princeton/Stored-Program)

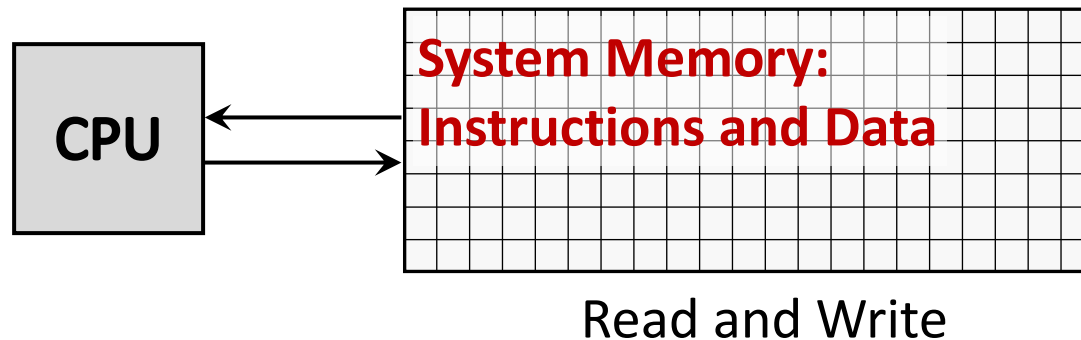


Harvard Architecture



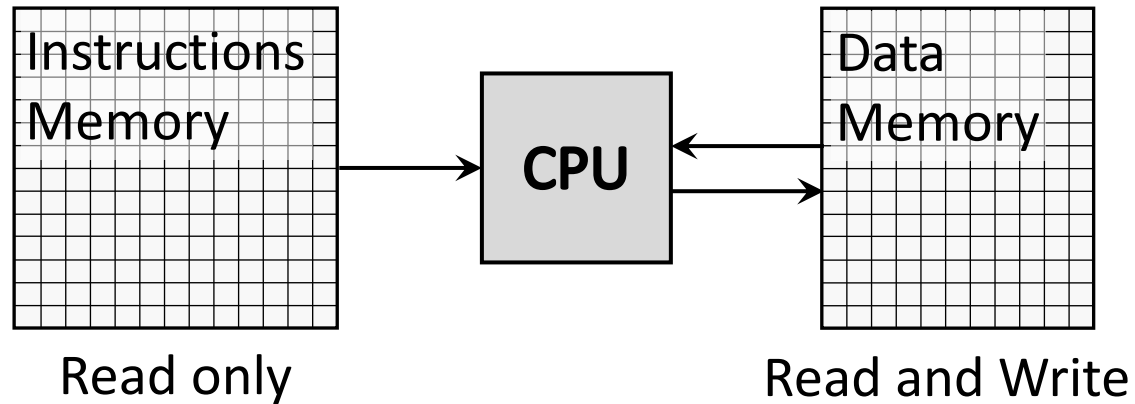
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Von Neumman Architecture (or Princeton/Stored-Program)



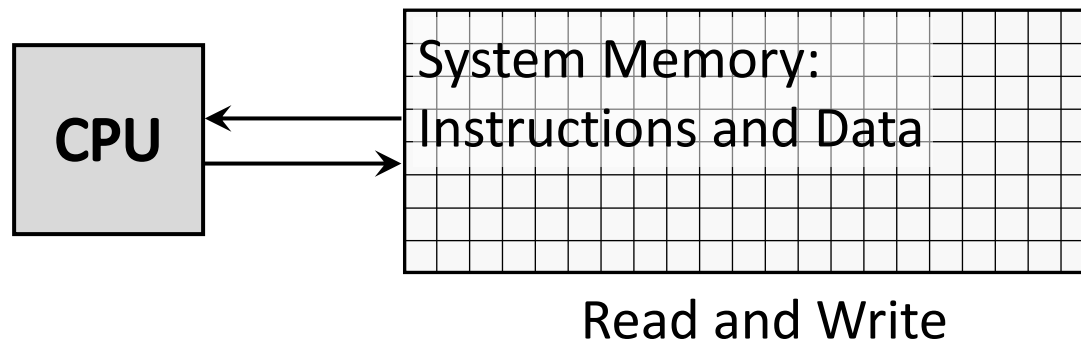
Principle	1) Instructions and data are stored in System Memory
-----------	---

Harvard Architecture



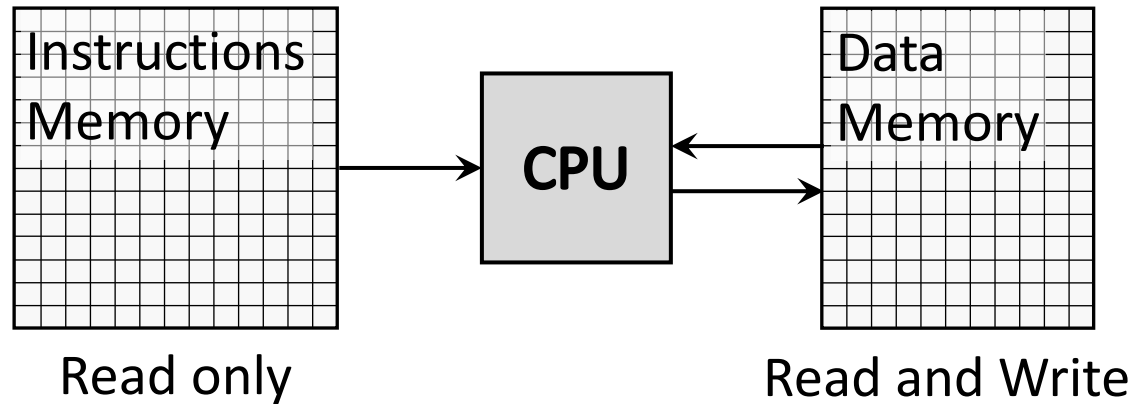
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Von Neumman Architecture (or Princeton/Stored-Program)



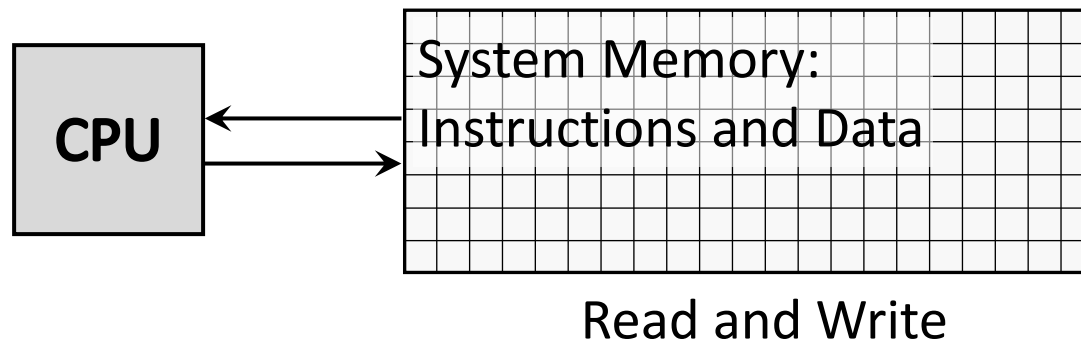
Principle	1) Instructions and data are stored in System Memory 2) Instructions are executed sequentially
-----------	--

Harvard Architecture



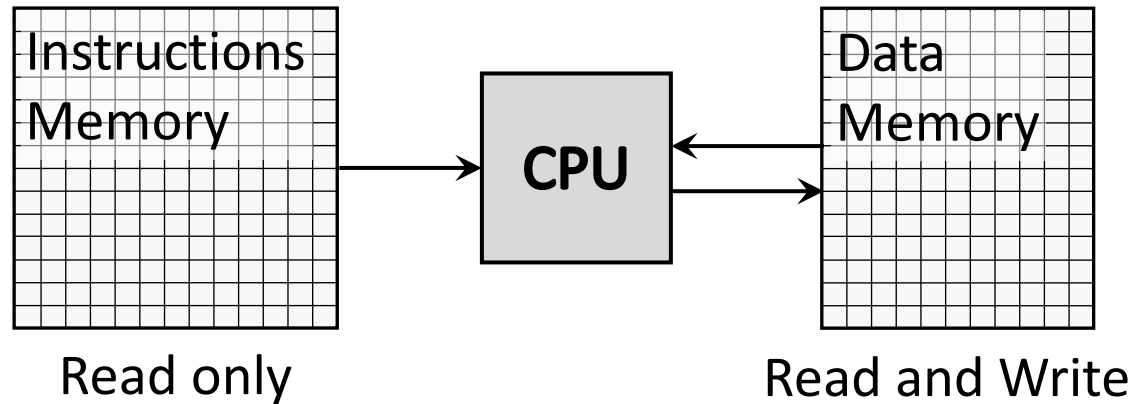
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Von Neumman Architecture (or Princeton/Stored-Program)



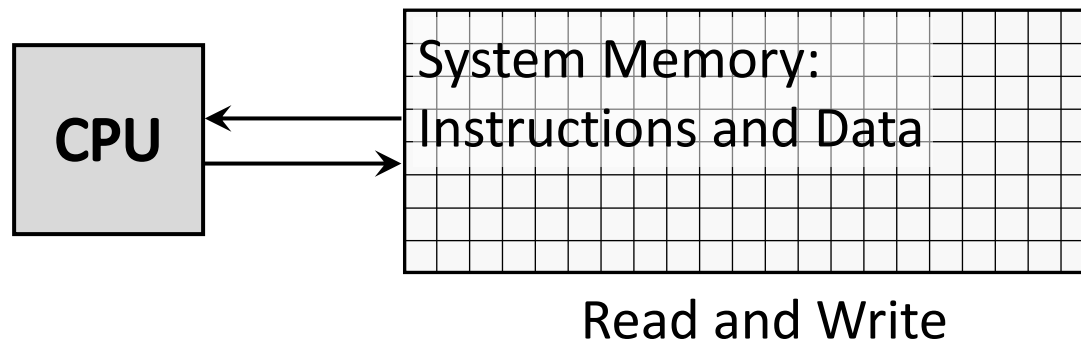
Principle	1) Instructions and data are stored in System Memory 2) Instructions are executed sequentially
Pros	+ Simpler memory management + Simpler hardware implementation

Harvard Architecture



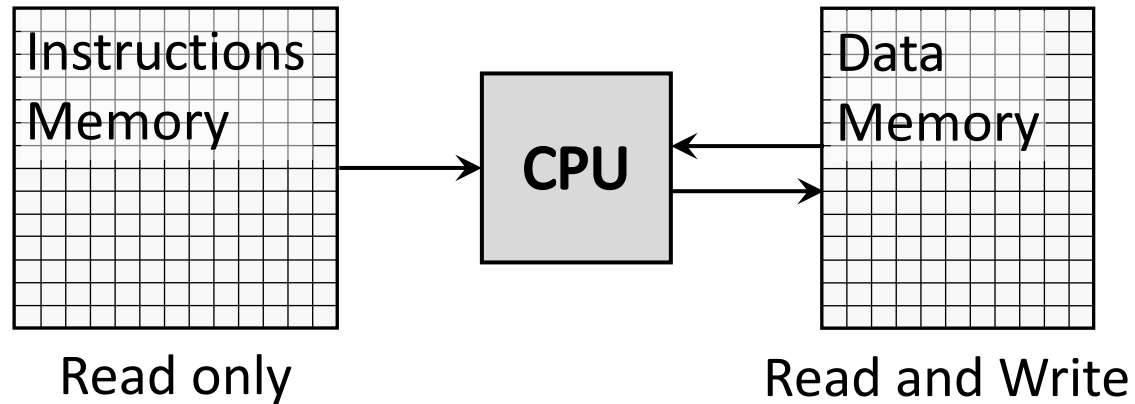
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Von Neumman Architecture (or Princeton/Stored-Program)



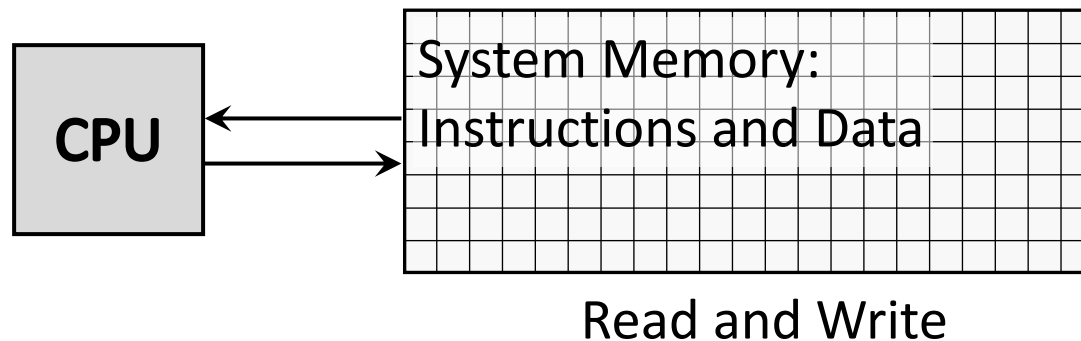
Principle	1) Instructions and data are stored in System Memory 2) Instructions are executed sequentially
Pros/Cons	+ Simpler memory management + Simpler hardware implementation - Higher performance overhead

Harvard Architecture



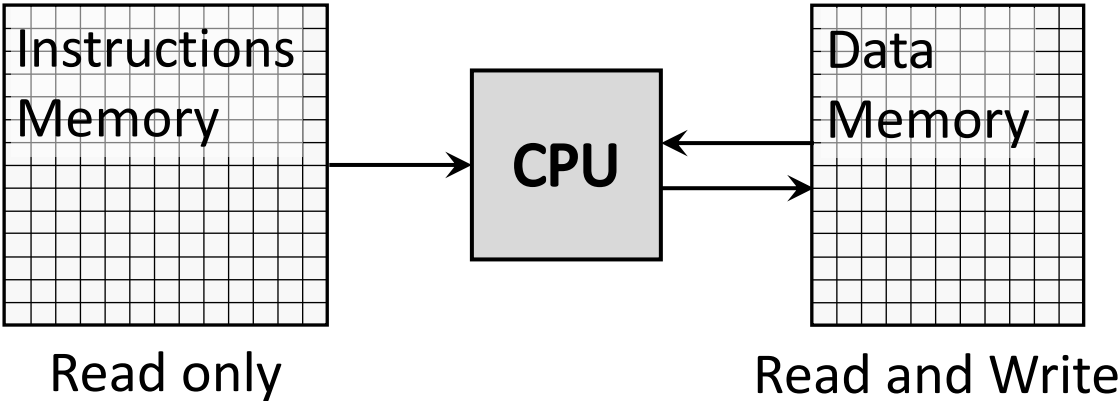
Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Von Neumman Architecture (or Princeton/Stored-Program)



Principle	1) Instructions and data are stored in System Memory 2) Instructions are executed sequentially
Pros/Cons	+ Simpler memory management + Simpler hardware implementation - Higher performance overhead
Usage	All general-purpose computers

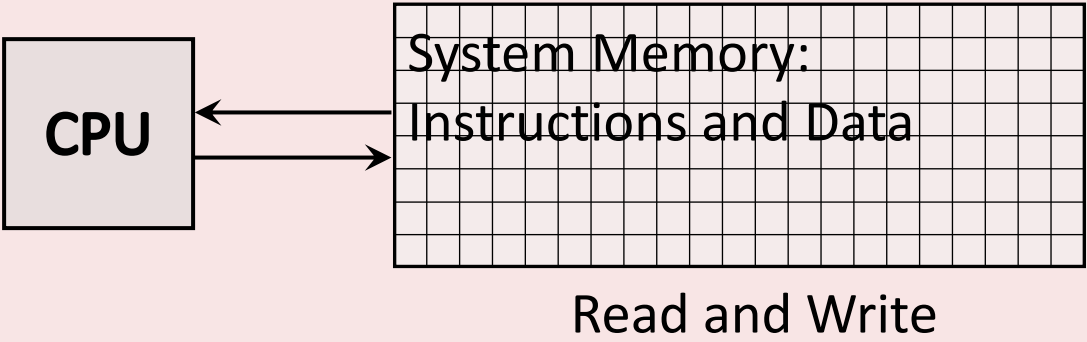
Harvard Architecture



Principle	Instructions and data are in different physical memory units
Pros/Cons	+ Lower performance overheads - Complex hardware implementation
Usage	Embedded systems

Dominant

Von Neumman Architecture (or Princeton/Stored-Program)



Principle	1) Instructions and data are stored in System Memory 2) Instructions are executed sequentially
Pros/Cons	+ Simpler memory management + Simpler hardware implementation - Higher performance overhead
Usage	All general-purpose computers

Sample Configurations

	Config. 1	Config. 2	Config. 3	Config. 4
Cores Number	1	2	4	4
Cores Speed, GHz	3.5	2.4	3.5	2.8
Memory Capacity, GB	32	16	32	32
Memory Speed, GB/sec	32	16	32	16

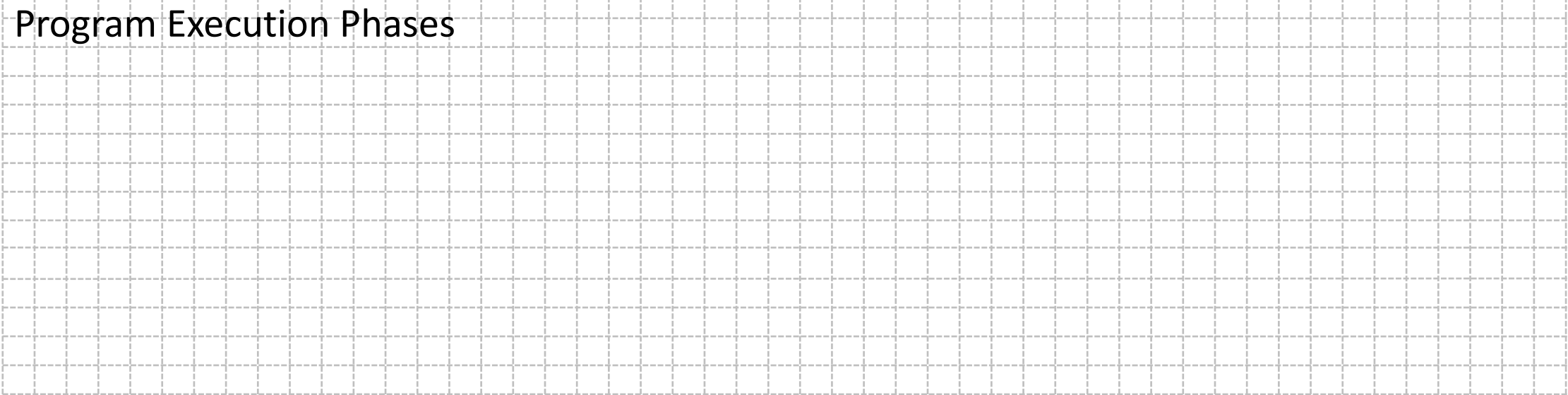
Performance Evaluation: **The Best Configuration?**

	Config. 1	Config. 2	Config. 3	Config. 4
Cores Number	1	2	4	4
Cores Speed, GHz	3.5	2.4	3.5	2.8
Memory Capacity, GB	32	16	32	32
Memory Speed, GB/sec	32	16	32	16

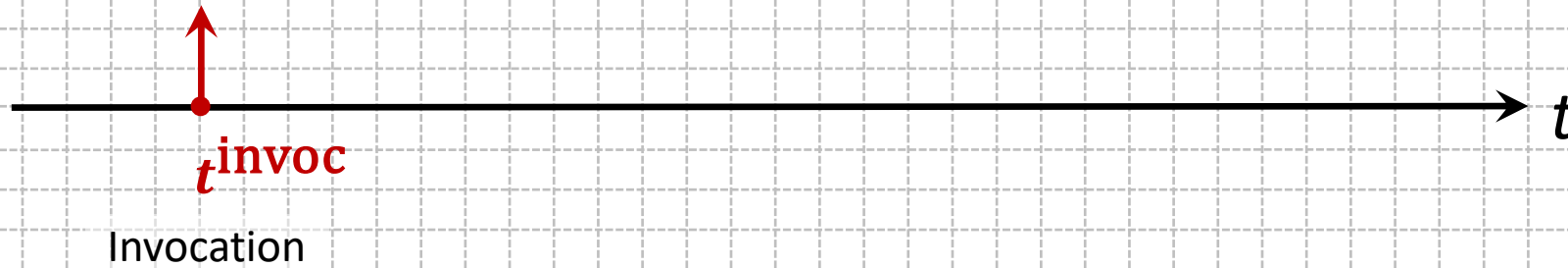
Performance Evaluation: The Best Configuration?

	Config. 1	Config. 2	Config. 3	Config. 4
Cores Number	1	2	4	4
Cores Speed, GHz	3.5	2.4	3.5	2.8
Memory Capacity, GB	32	16	32	32
Memory Speed, GB/sec	32	16	32	16

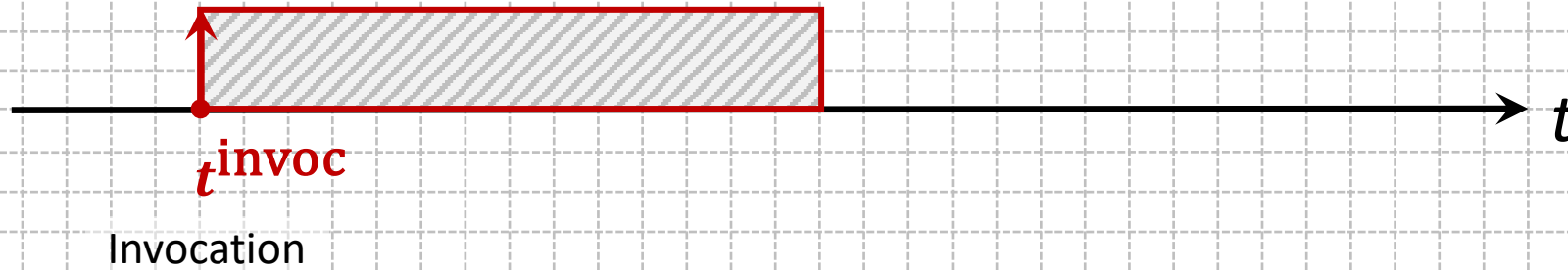
Program Execution Phases



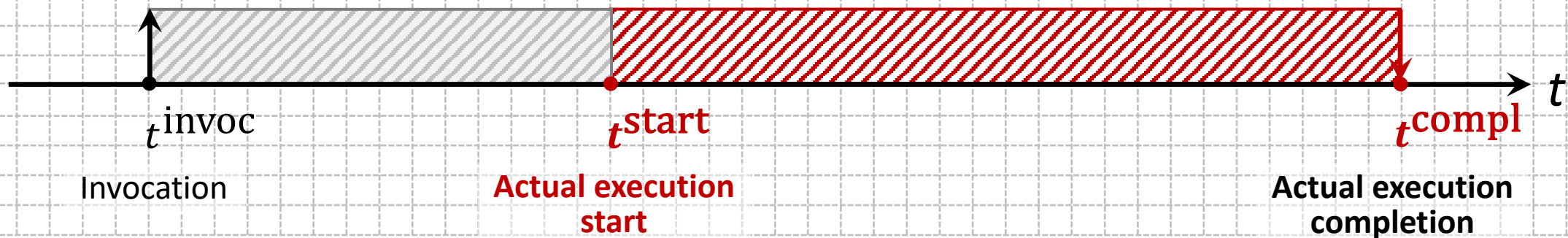
Program Execution Phases



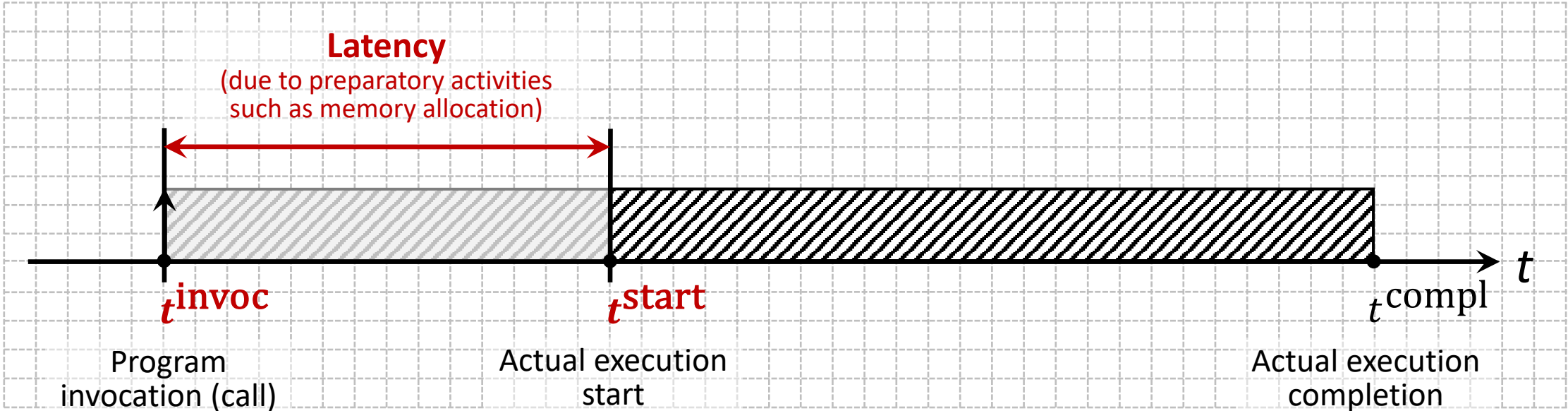
Program Execution Phases



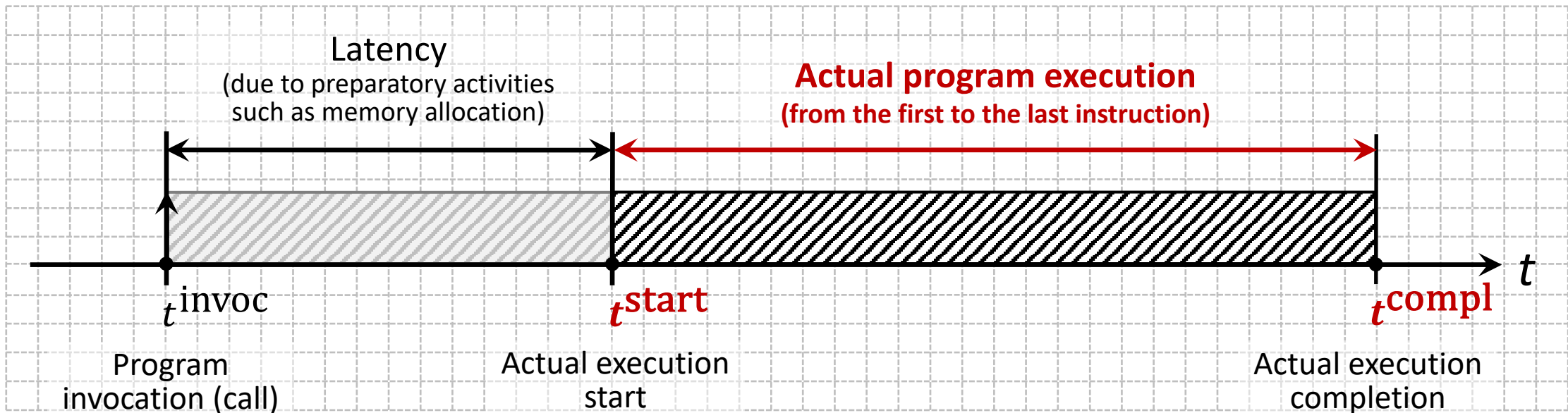
Program Execution Phases



Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time

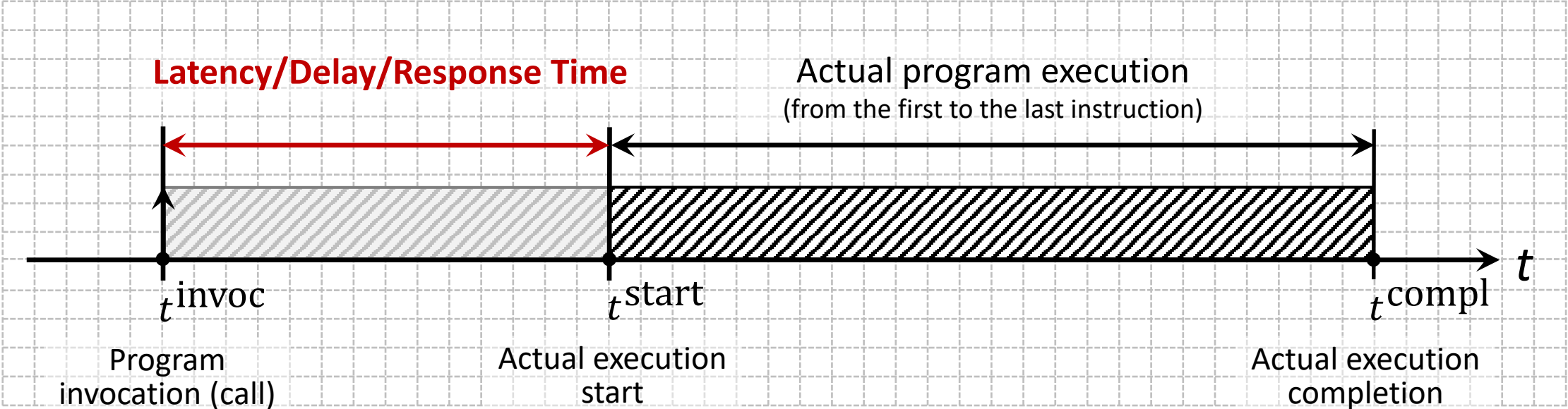


Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time

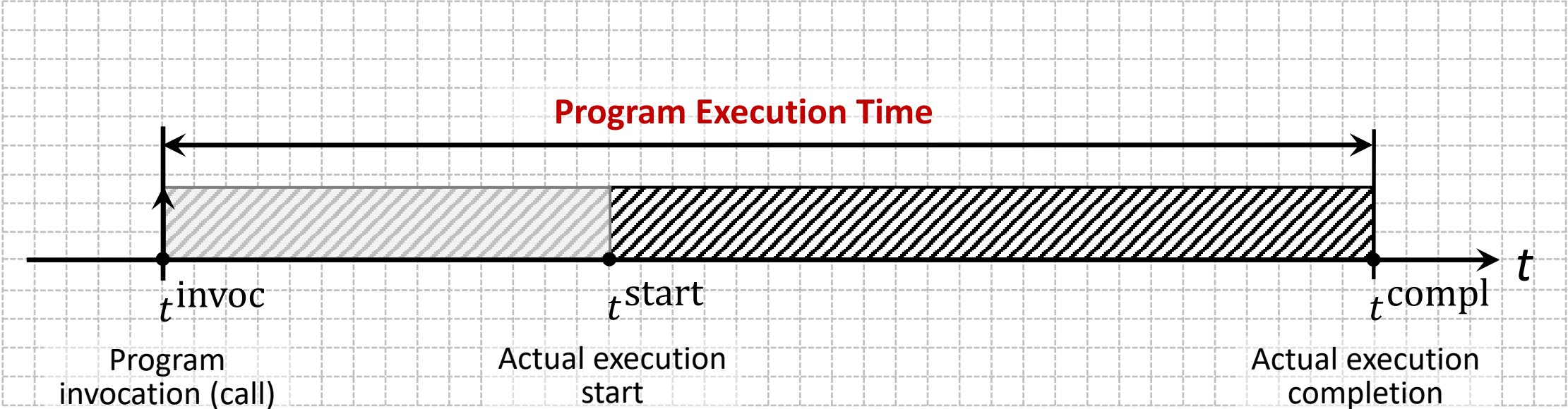


Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	

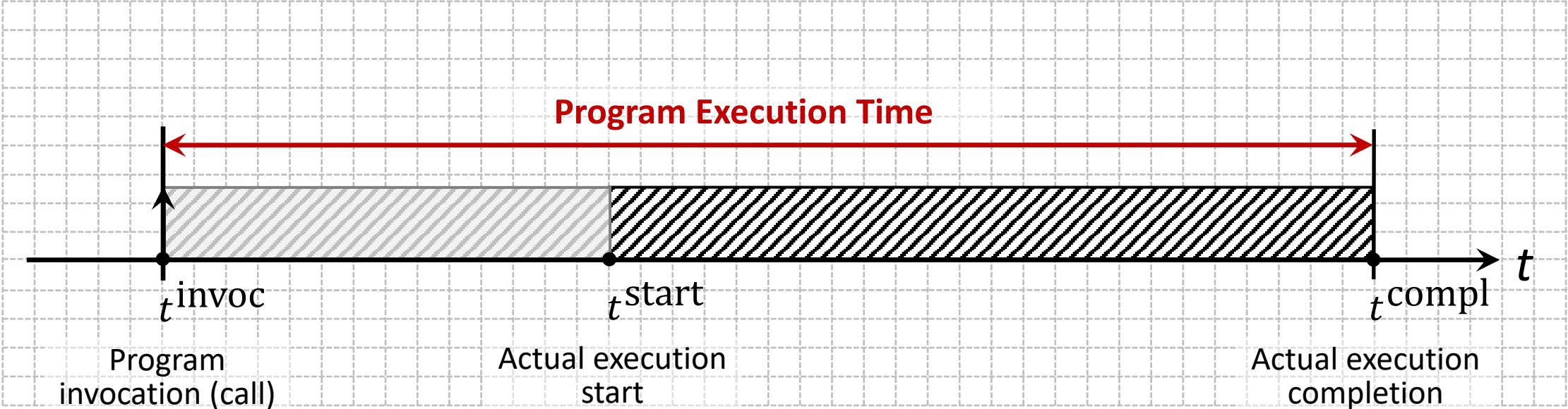
(Synonyms)



Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion



Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion



Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform		

Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently

Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently

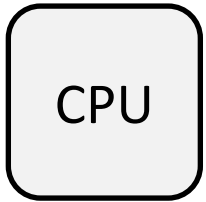
Uniprocessor with bandwidth 1:

CPU

**Not more than 1 job
to execute at a time**

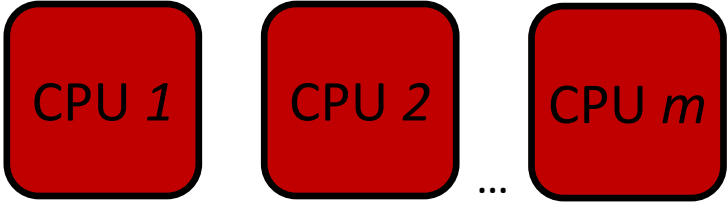
Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently

Uniprocessor



Not more than 1 job at a time
(bandwidth is 1 job)

Multiprocessor with bandwidth m :



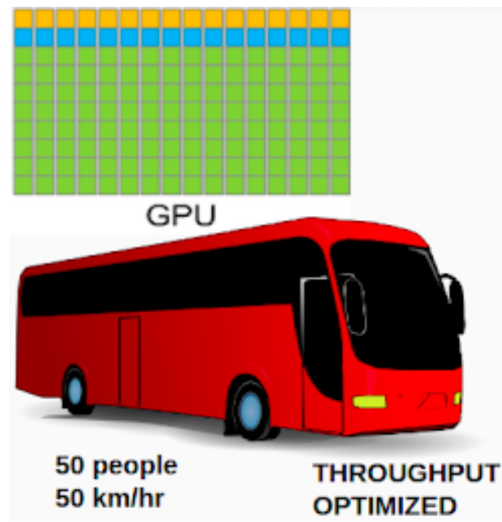
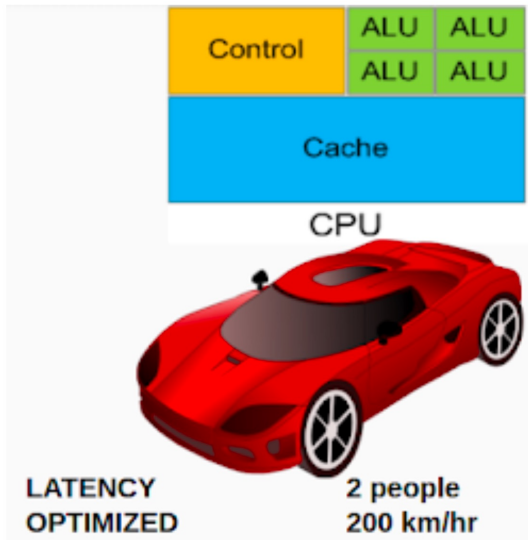
Not more than m jobs to execute concurrently

Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently
	Throughput	Average number of concurrently executing jobs (instructions) over some time

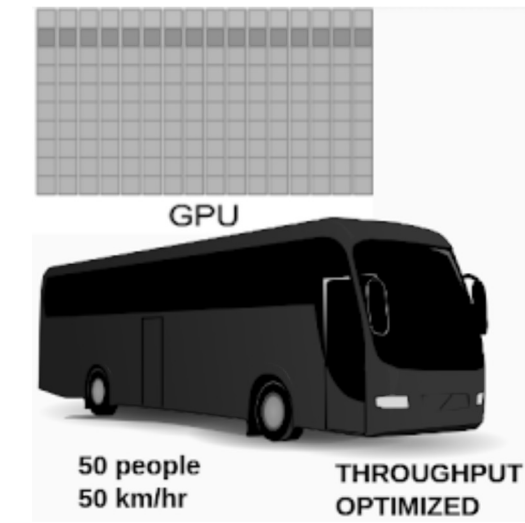
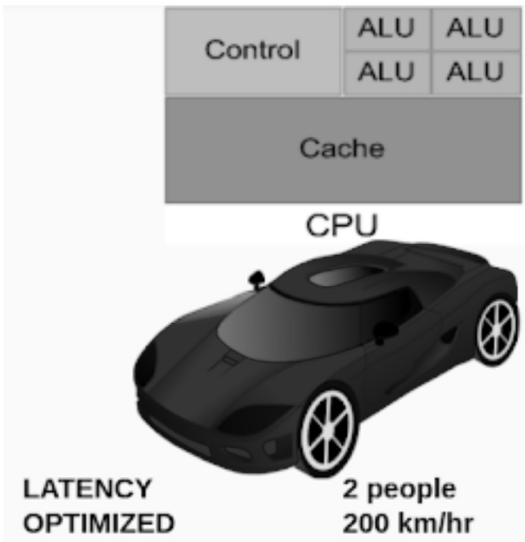
Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently
	Throughput	Average number of concurrently executing jobs (instructions) over some time
	Utilization	Time percentage of an effective component use

Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently
	Throughput	Average number of concurrently executing jobs (instructions) over some time
	Utilization	Time percentage of an effective component use (e.g. CPU utilization is 57%)

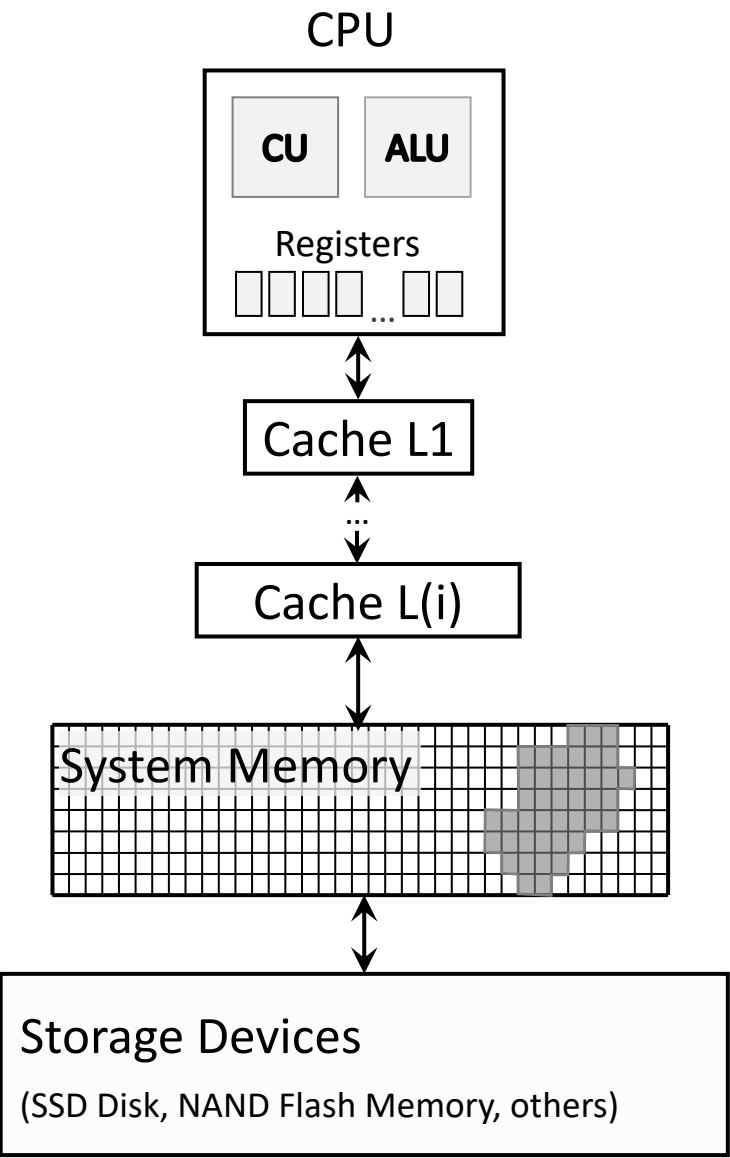
Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently
	Throughput	Average number of concurrently executing jobs (instructions) over some time
	Utilization	Time percentage of an effective component use (e.g. CPU utilization is 57%)



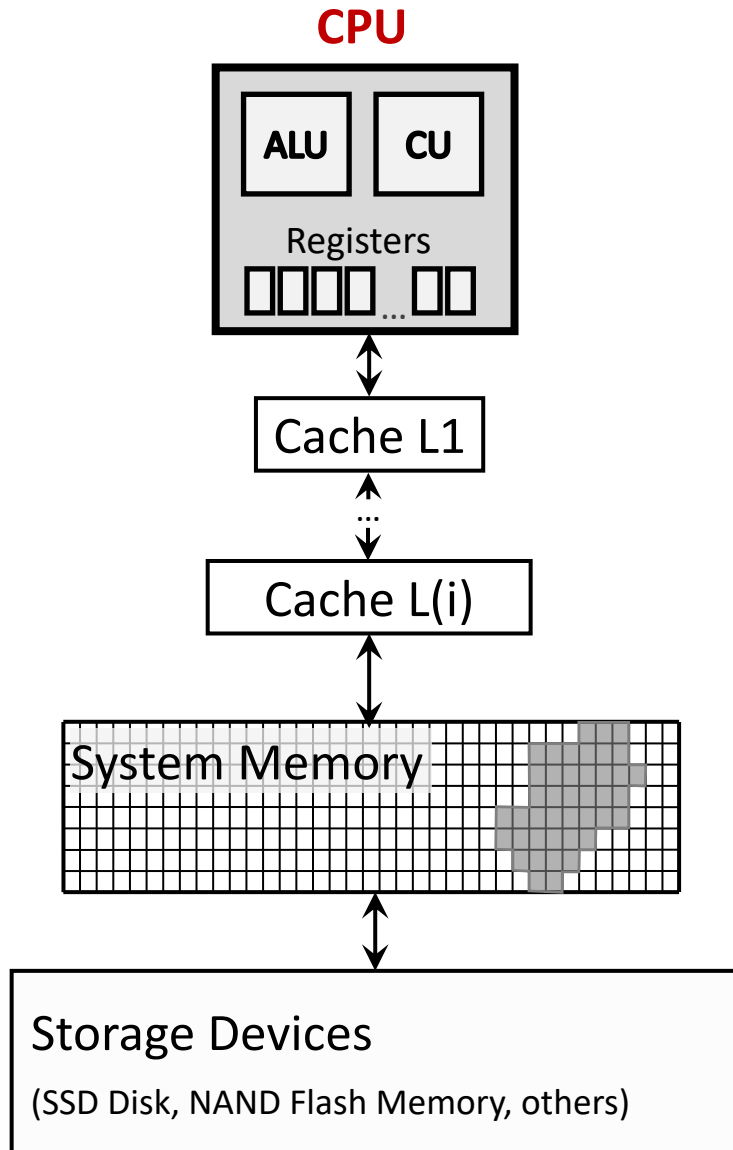
Performance Metric		Meaning
Program	Latency	Delay between a program invocation (call) and an actual execution start time
	Delay	
	Response Time	
	Execution Time	Time between a program invocation and its completion
Hardware Platform	Bandwidth	Theoretical bound on the number of jobs (instructions) to execute concurrently
	Throughput	Average number of concurrently executing jobs (instructions) over some time
	Utilization	Time percentage of an effective component use (e.g. CPU utilization is 57%)



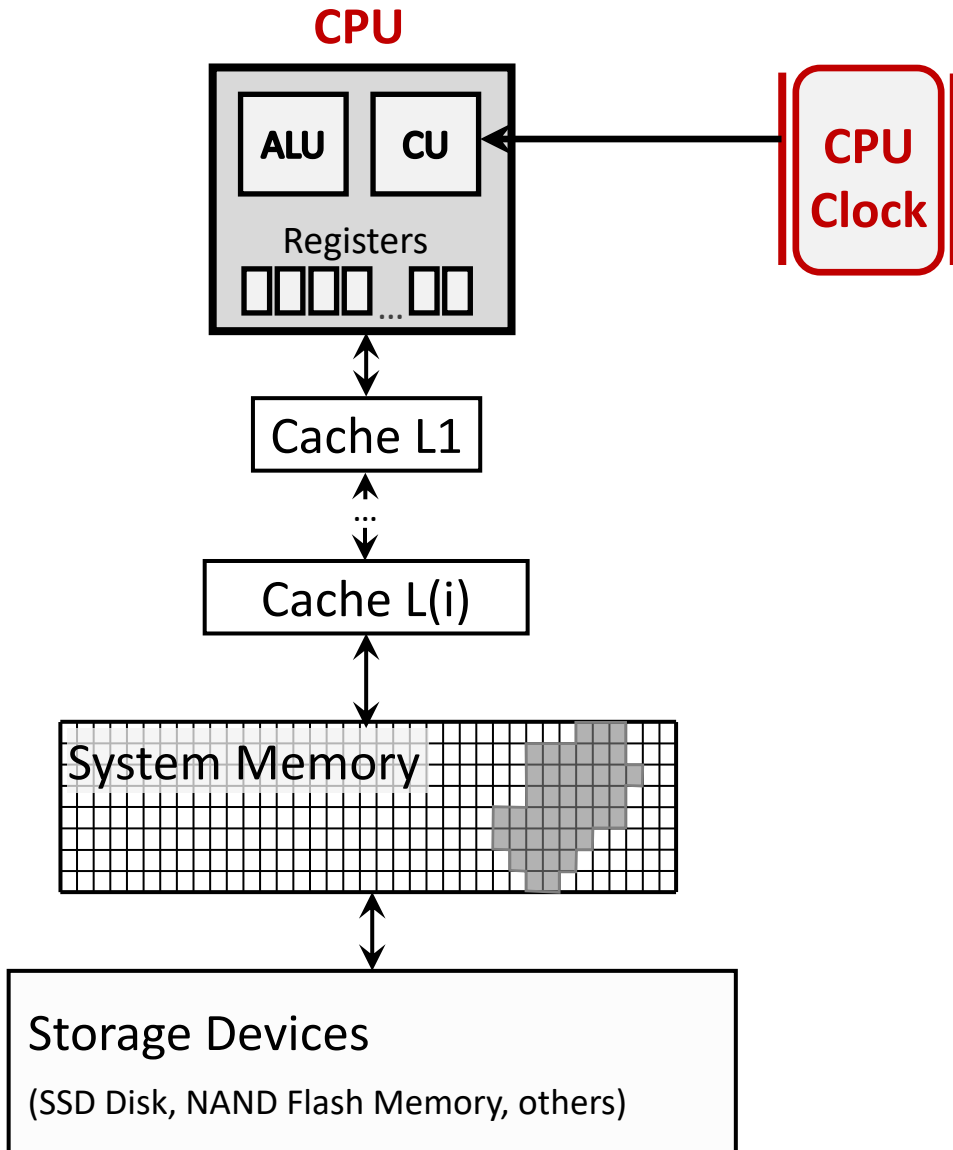
Recap: Computer Layout



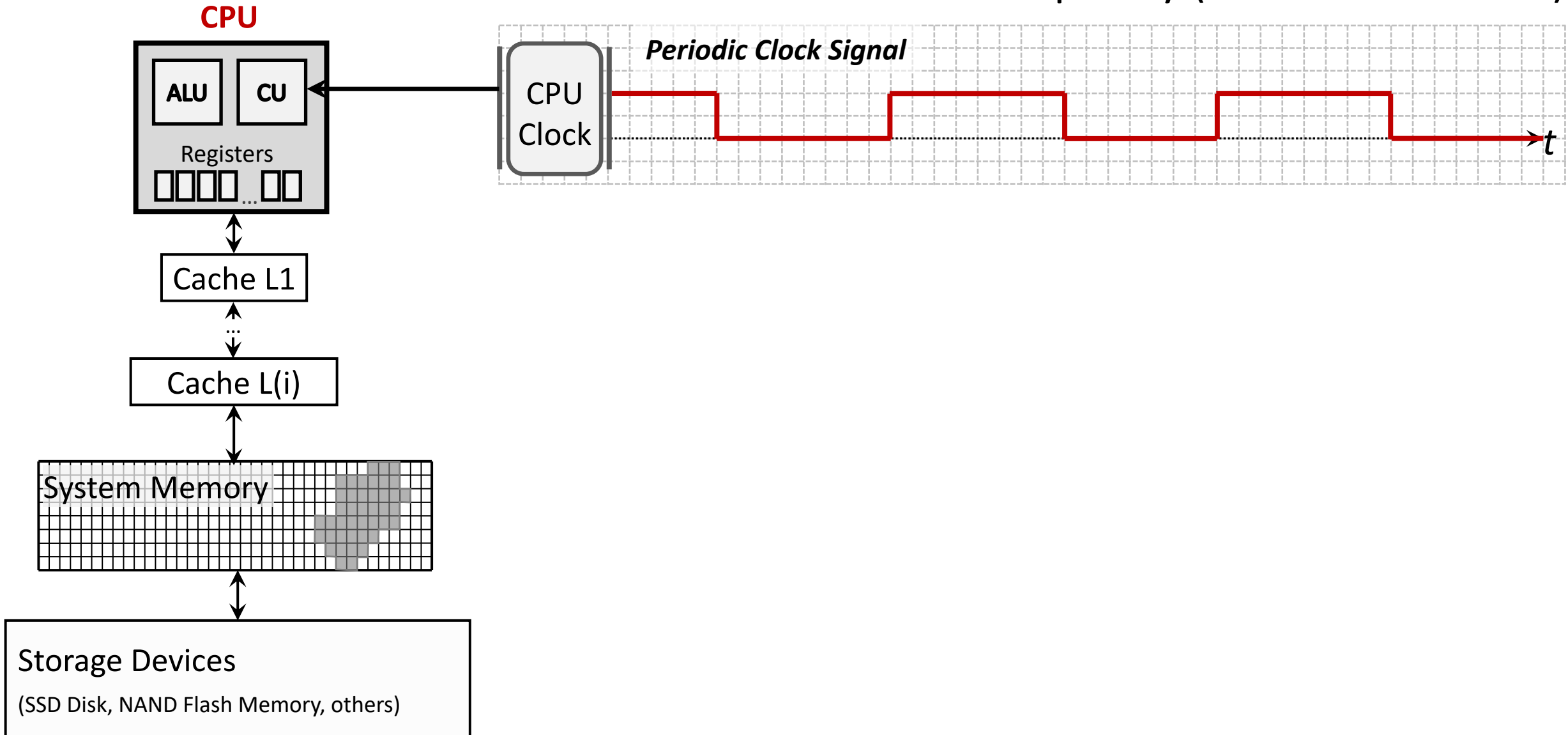
Processor Performance Characteristics



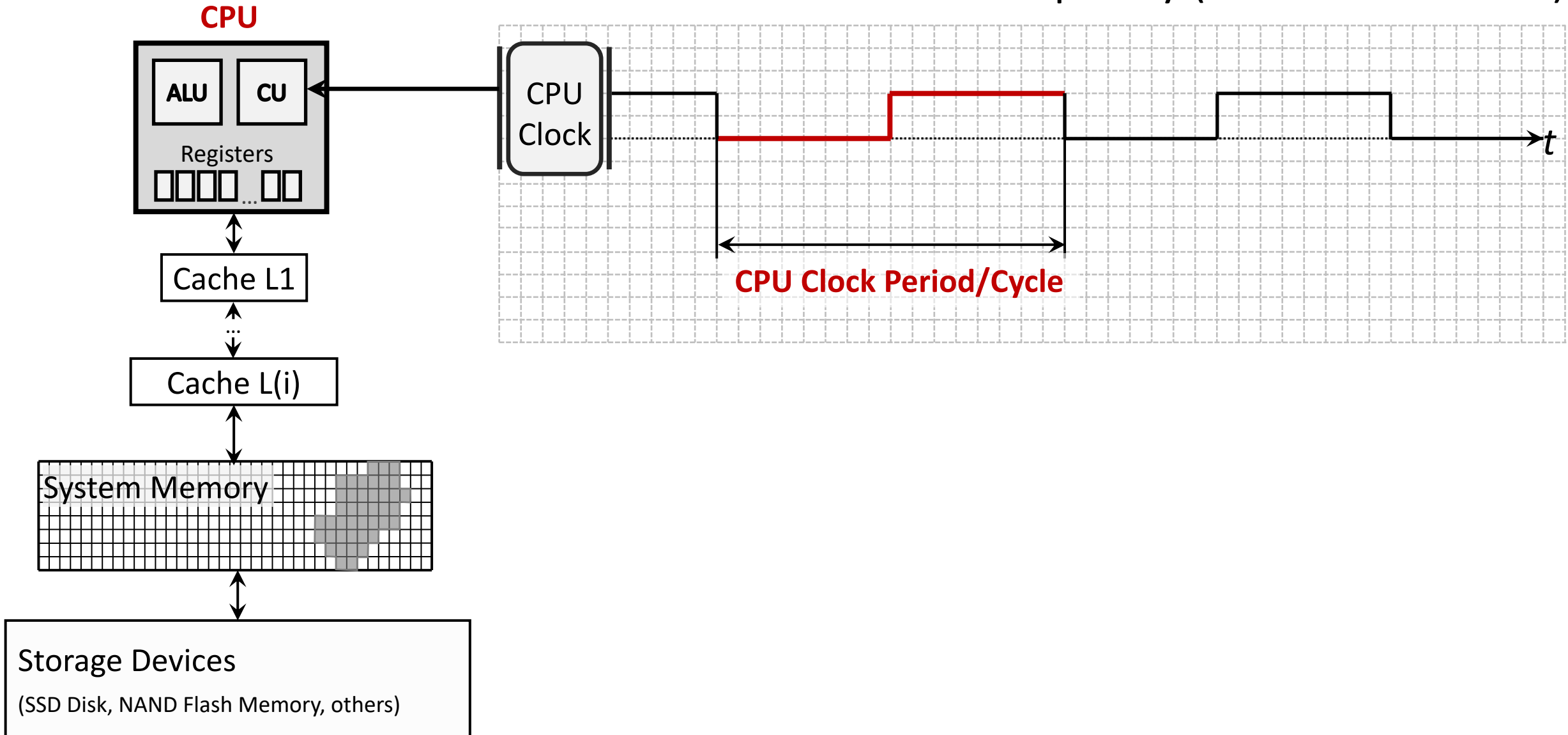
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



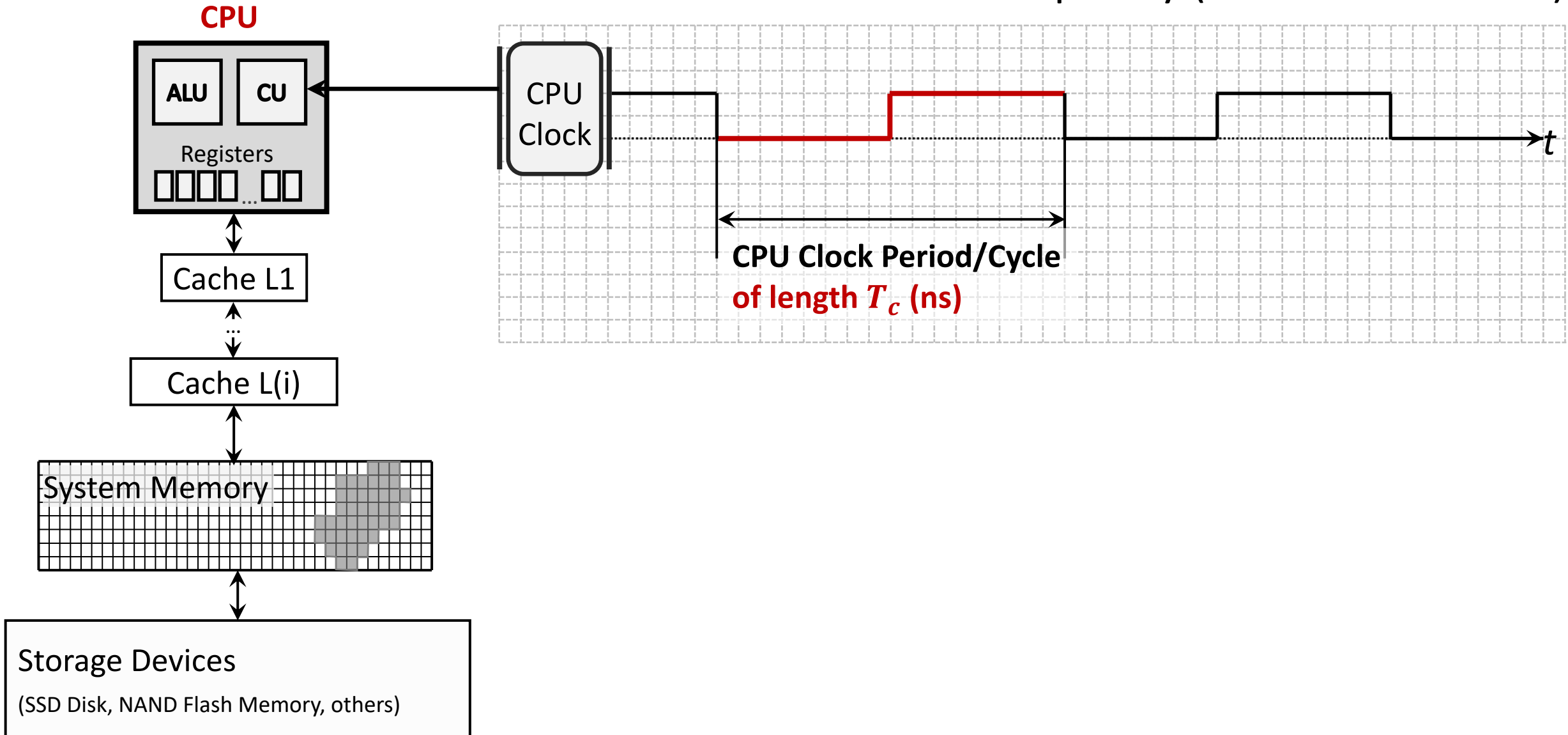
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



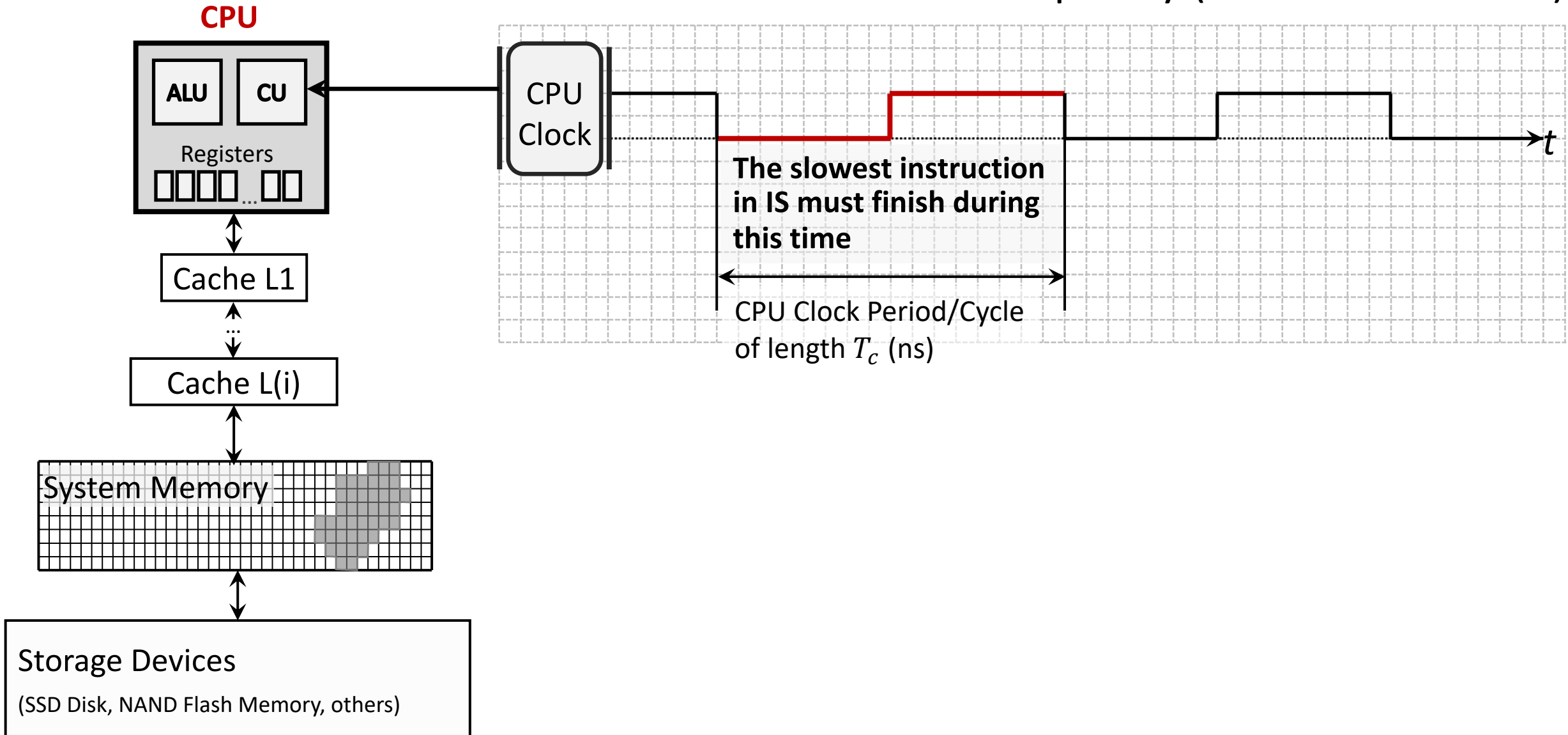
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



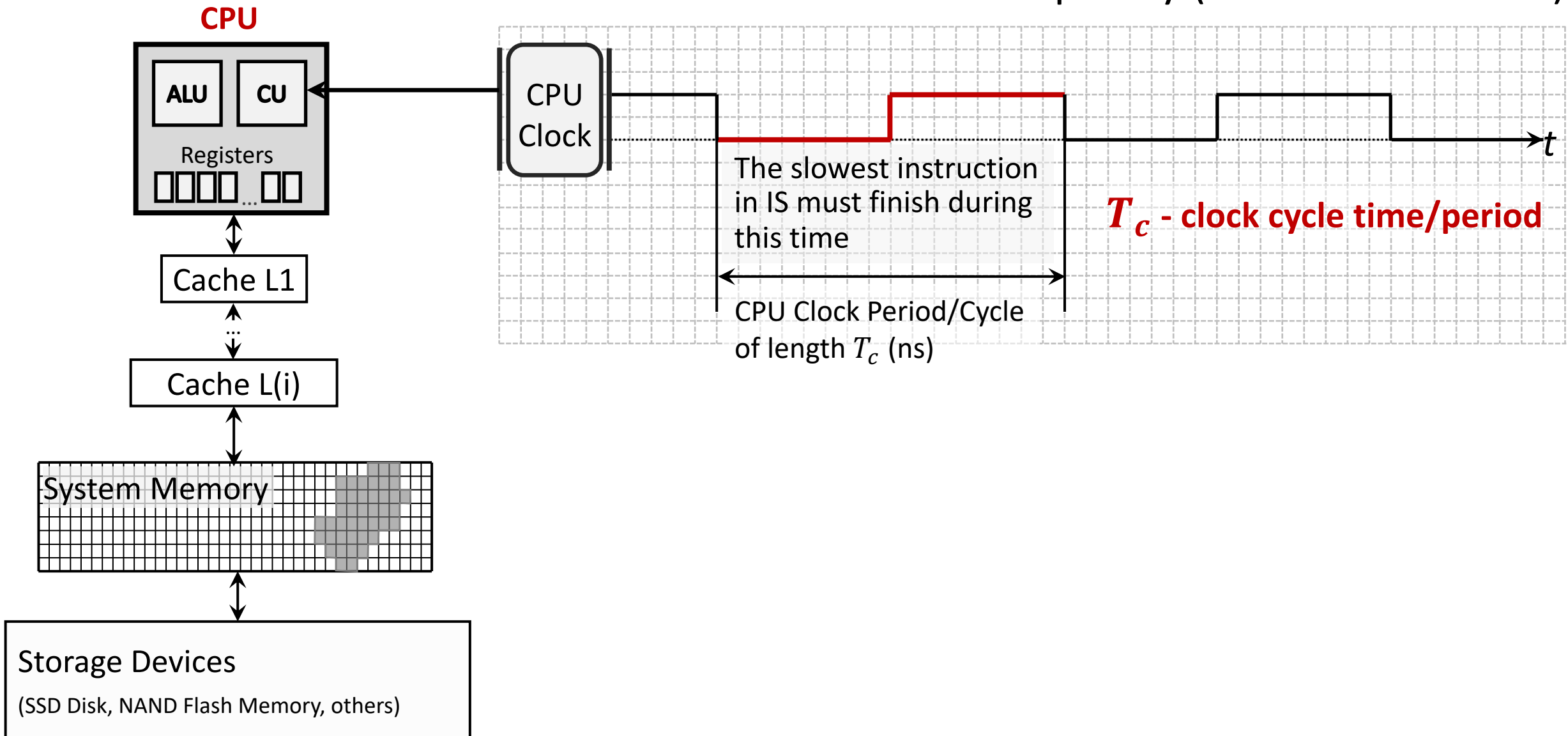
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



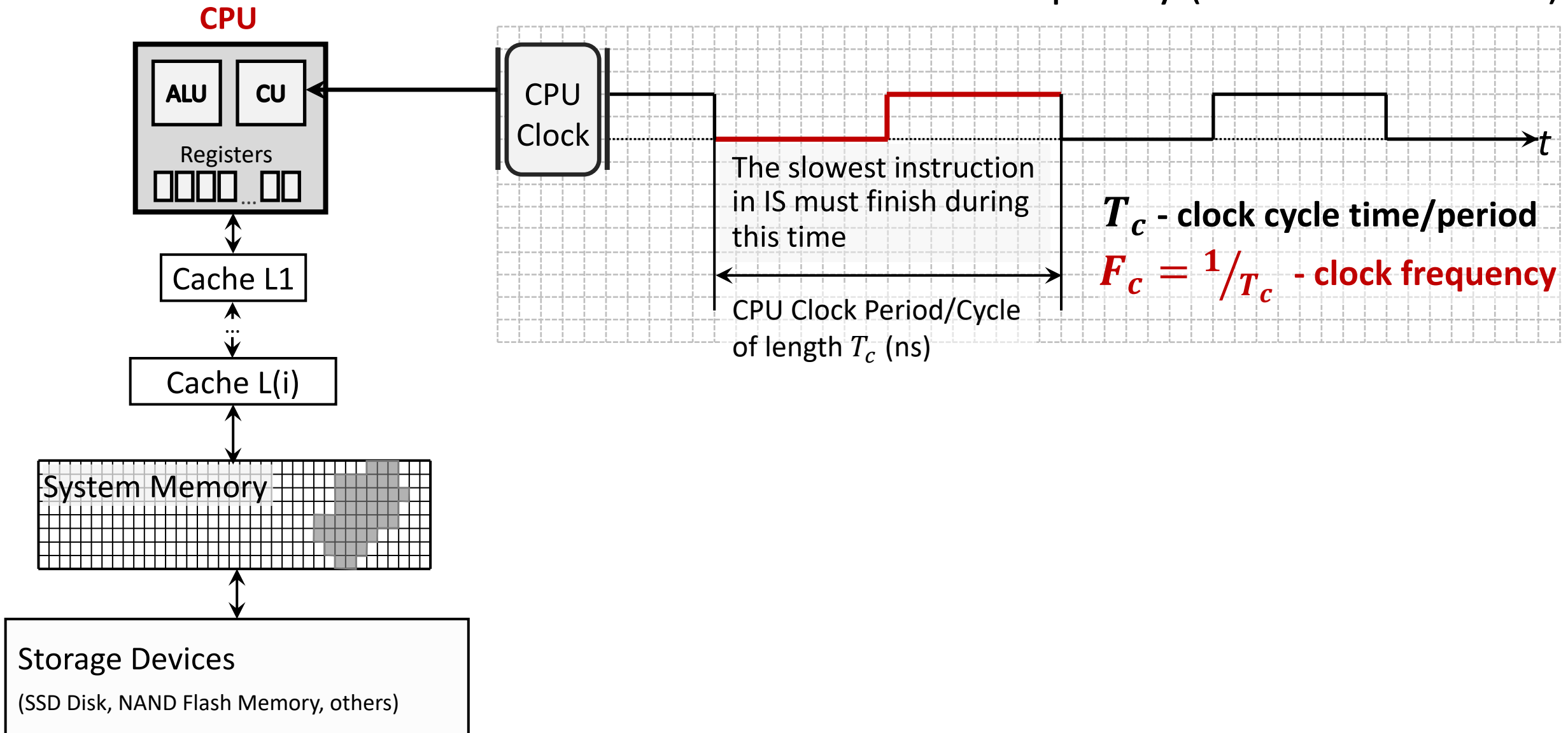
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



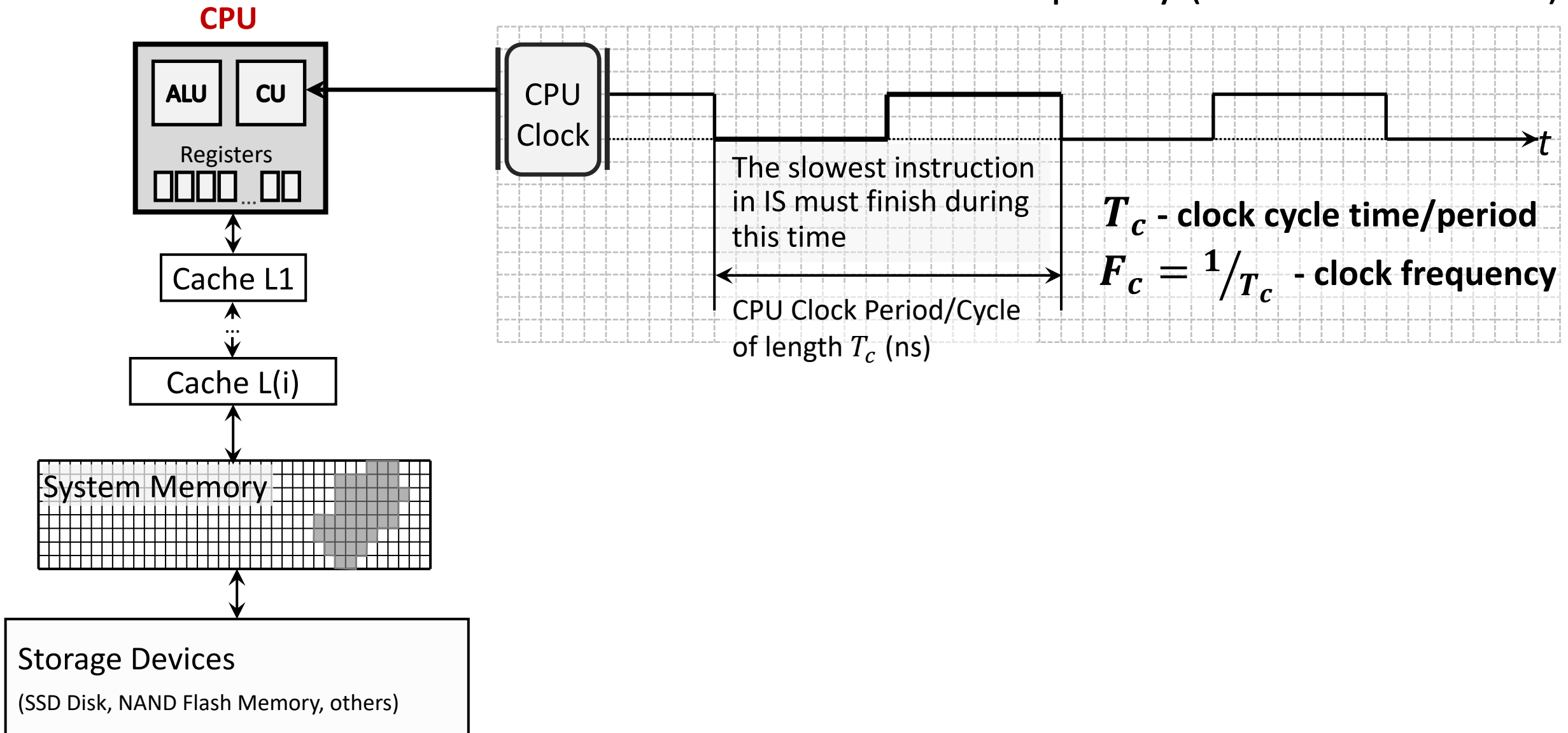
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



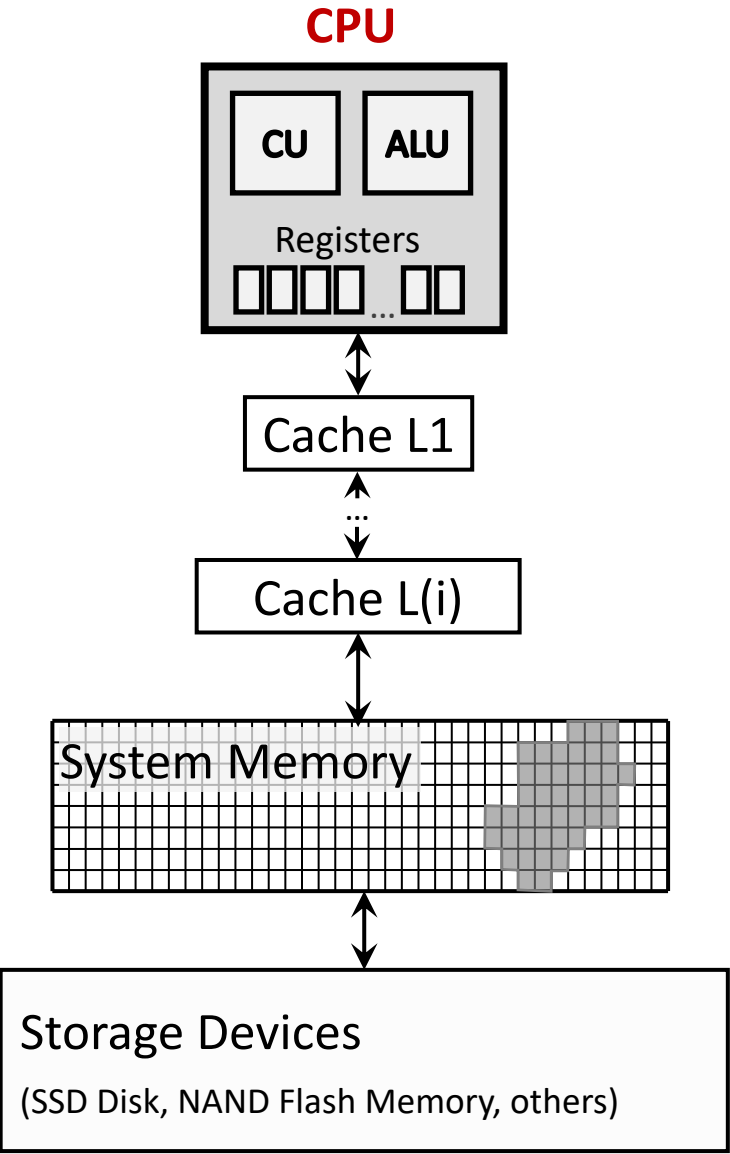
Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)



Processor Performance Characteristic: CPU Frequency (or CPU Clock Rate)

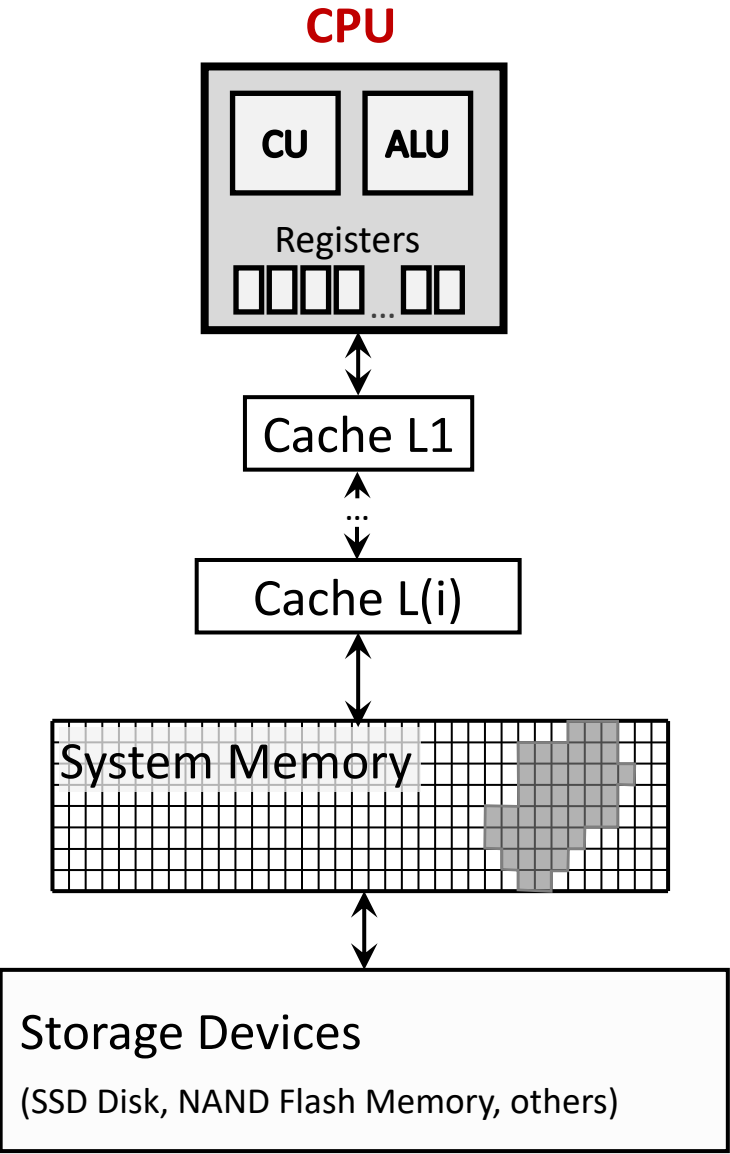


Average Performance Characteristics (Desktop/Laptop)



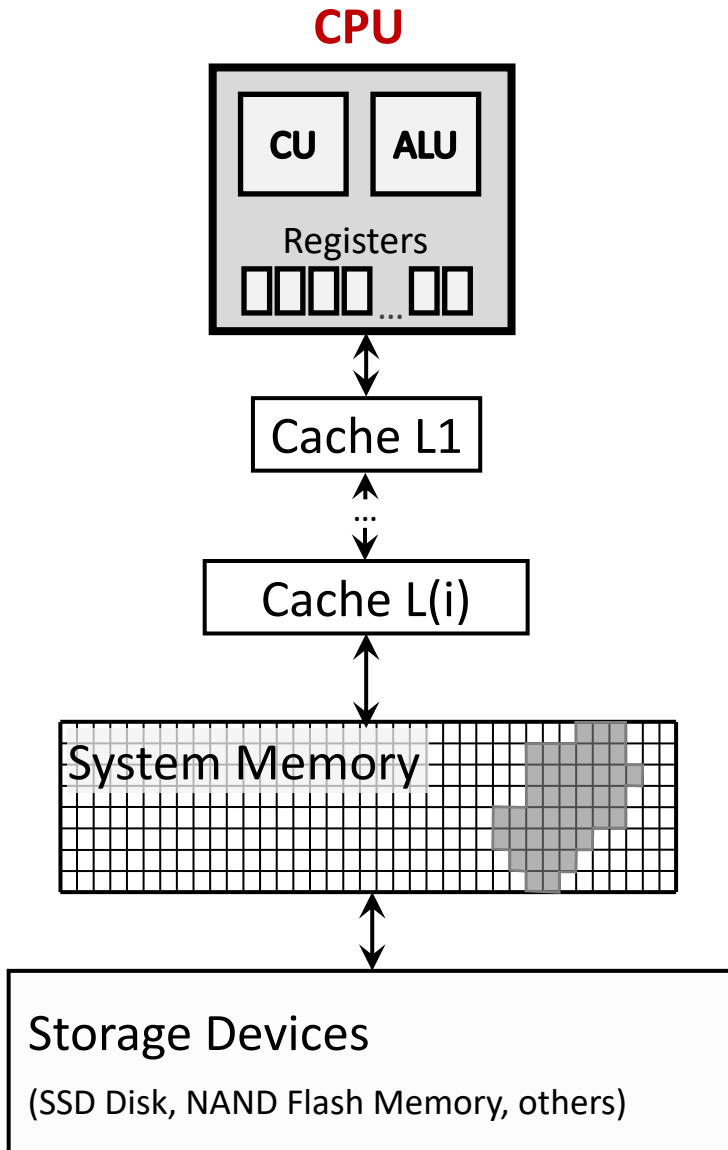
Device	Operation Speed	Storage Capacity
CPU		

Average Performance Characteristics (Desktop/Laptop)



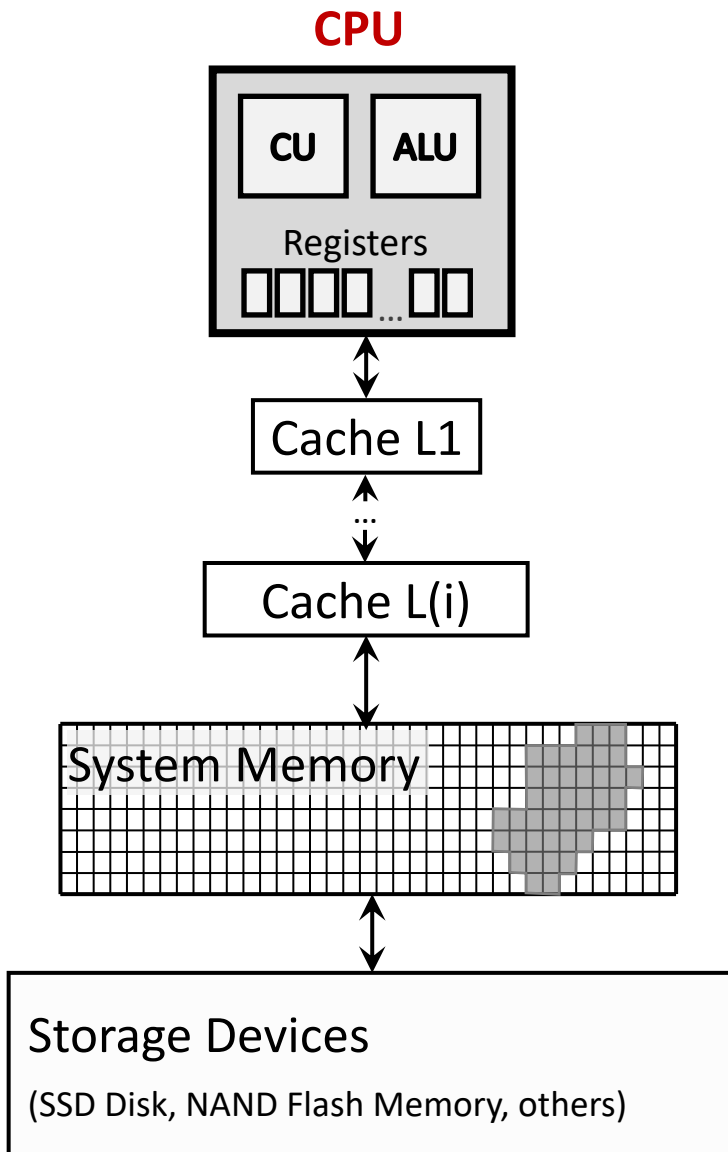
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)		

Average Performance Characteristics (Desktop/Laptop)



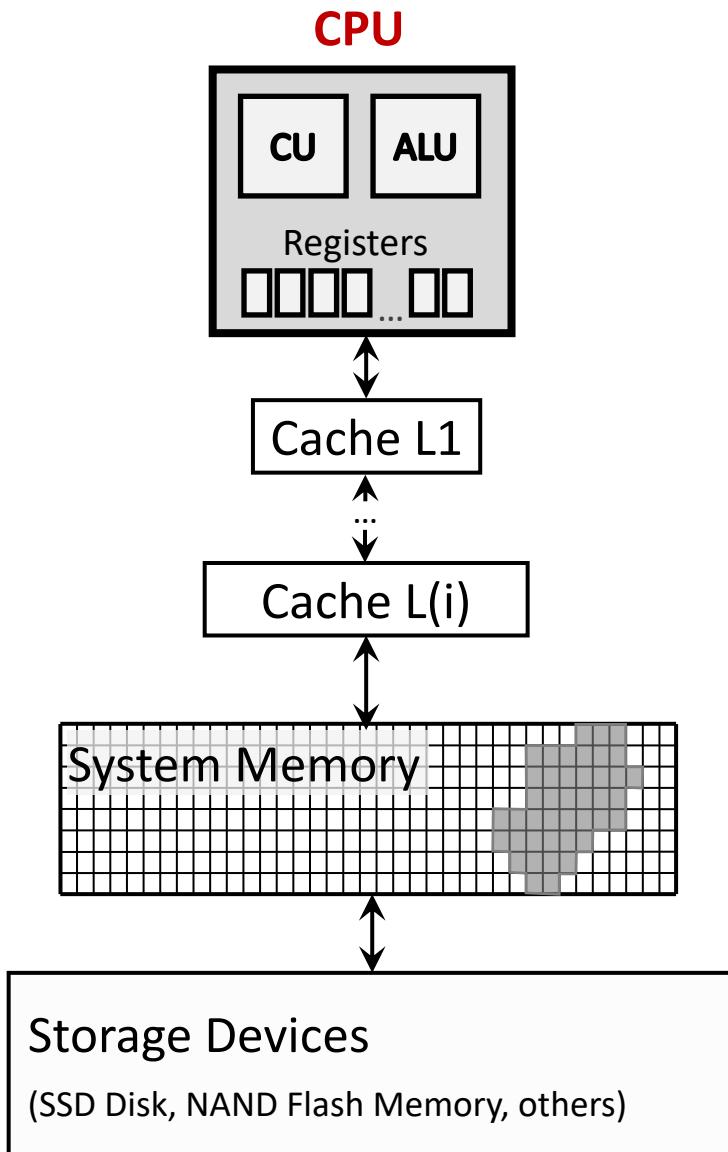
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	

Average Performance Characteristics (Desktop/Laptop)



Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register

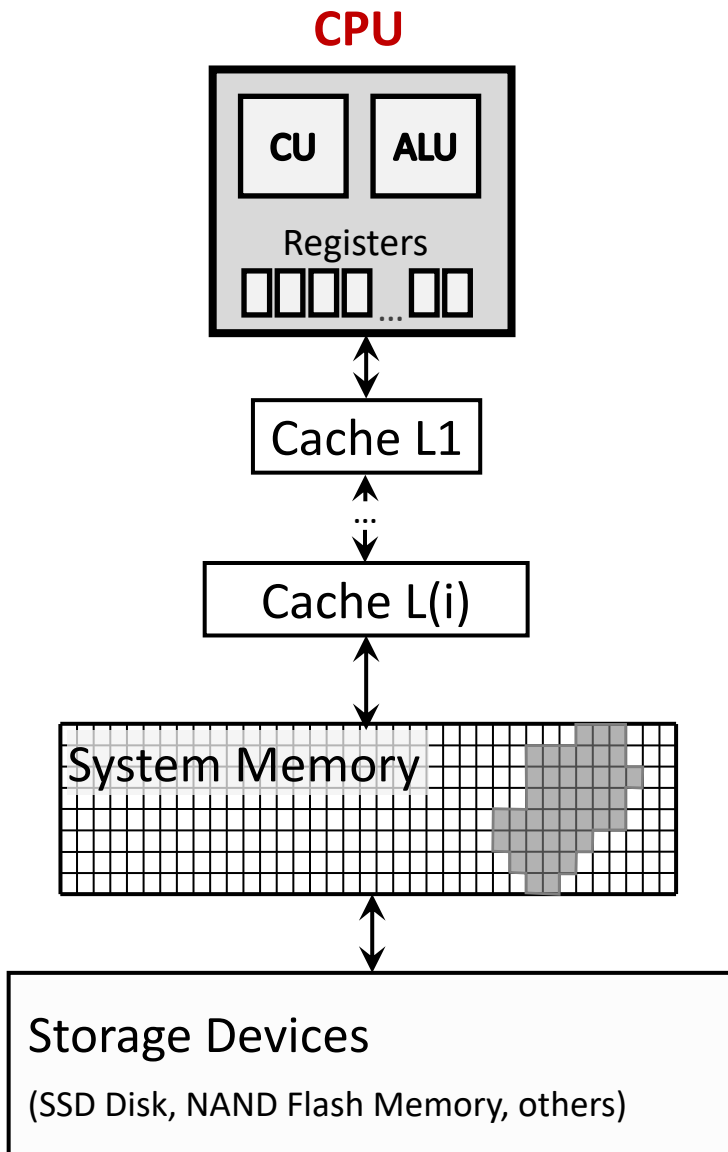
Average Performance Characteristics (Desktop/Laptop)



Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register

MIPS (Millions Instructions Per Second):
Another metric for CPU speed

Average Performance Characteristics (Desktop/Laptop)

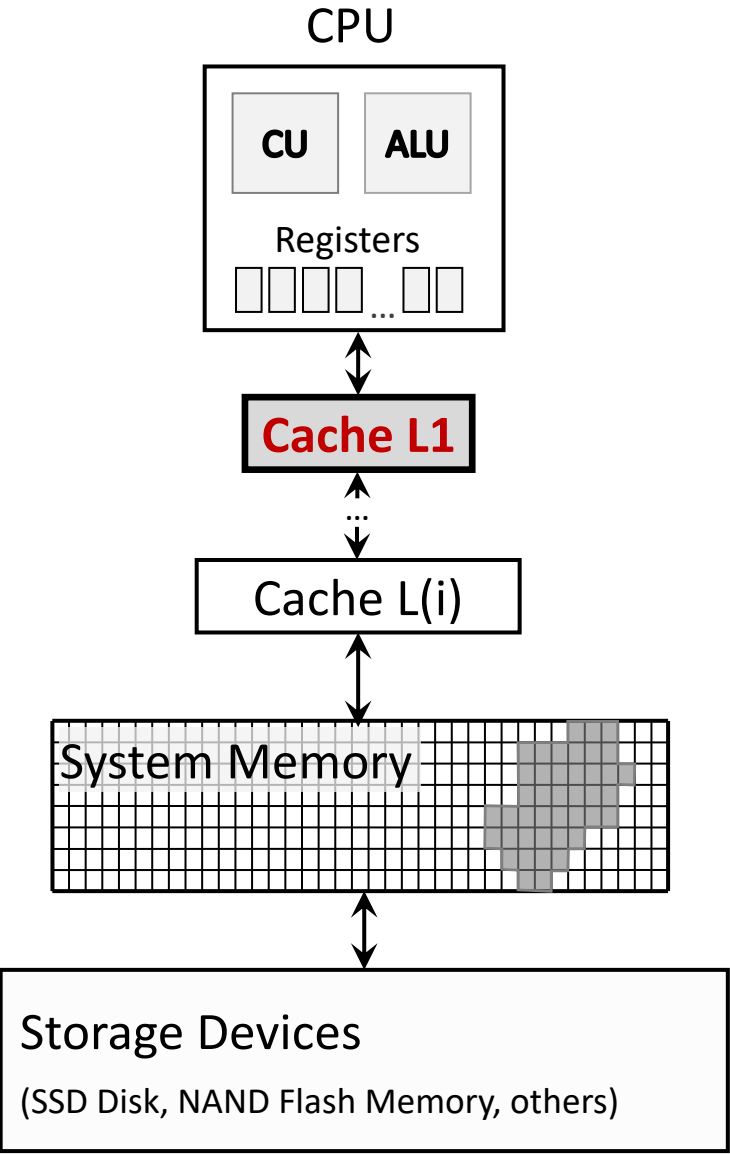


Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register

MIPS (Millions Instructions Per Second):
Another metric for CPU speed

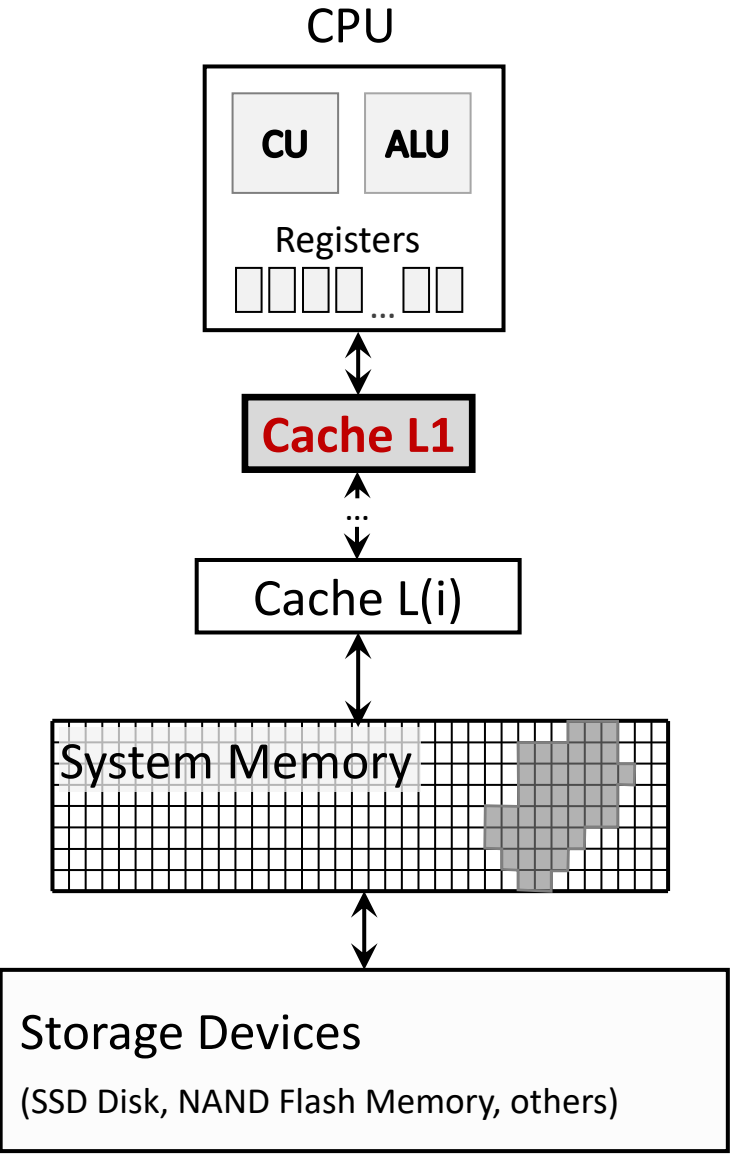
! MIPS also means the family of processors
(Microprocessor without Interlocked Pipelined Stages)

Average Performance Characteristics (Desktop/Laptop)



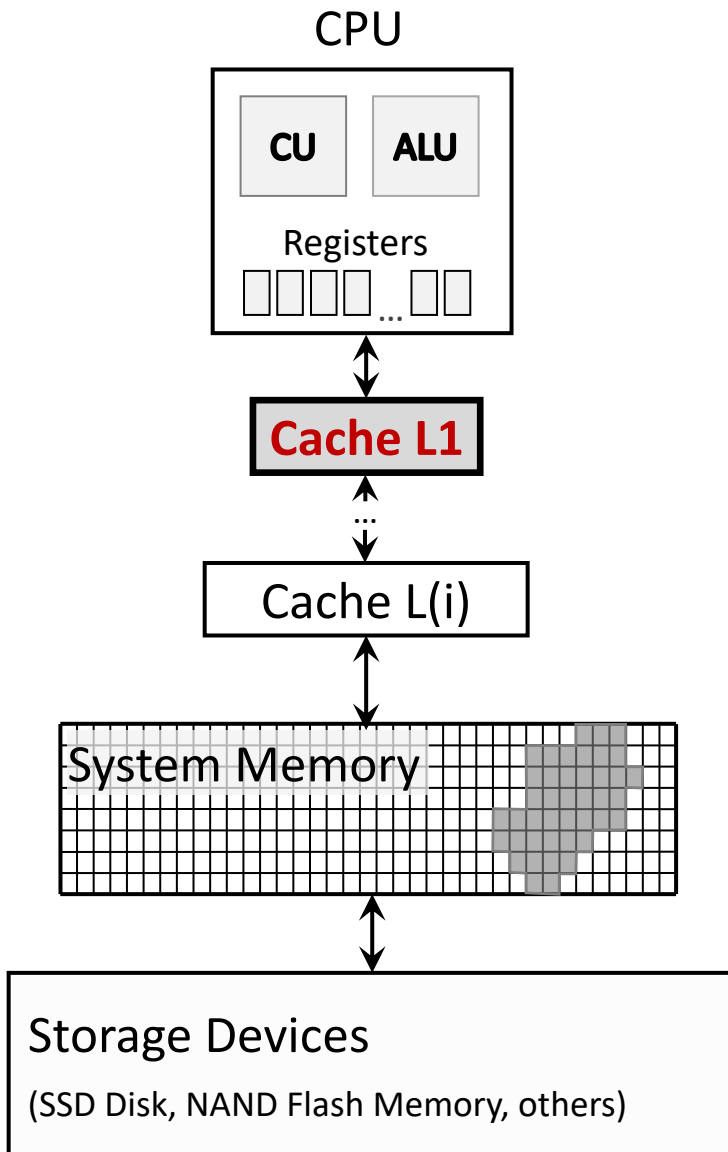
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		

Average Performance Characteristics (Desktop/Laptop)



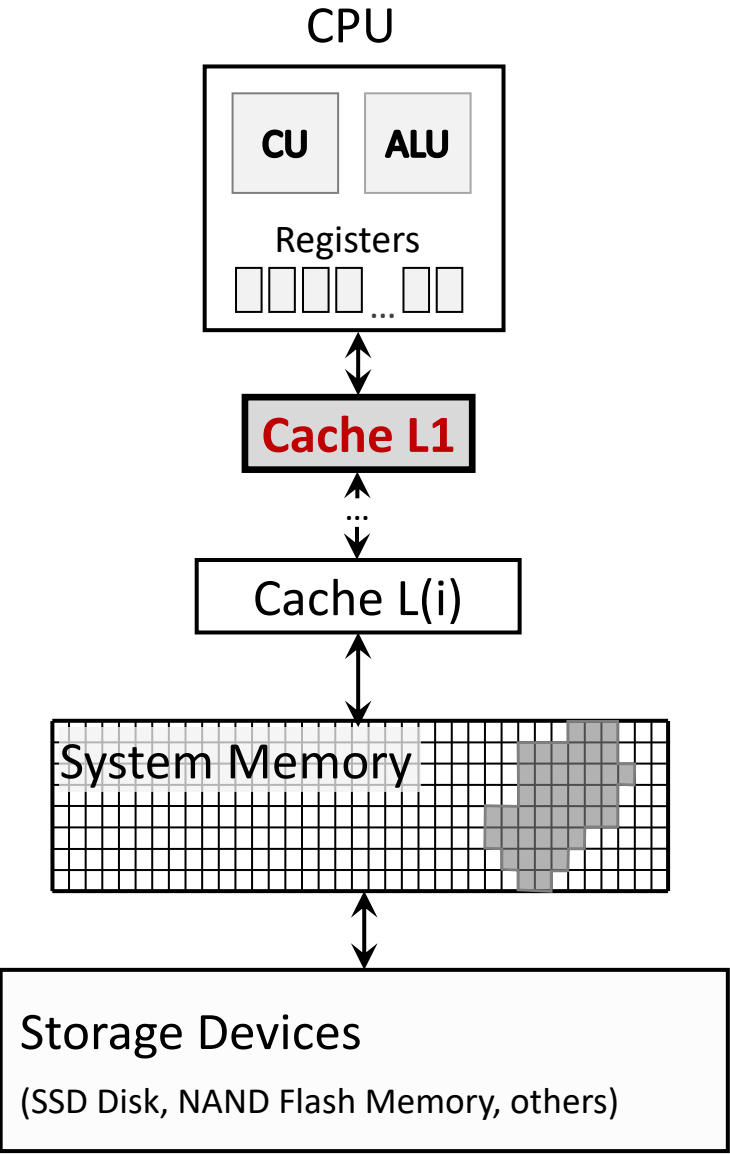
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed;	

Average Performance Characteristics (Desktop/Laptop)



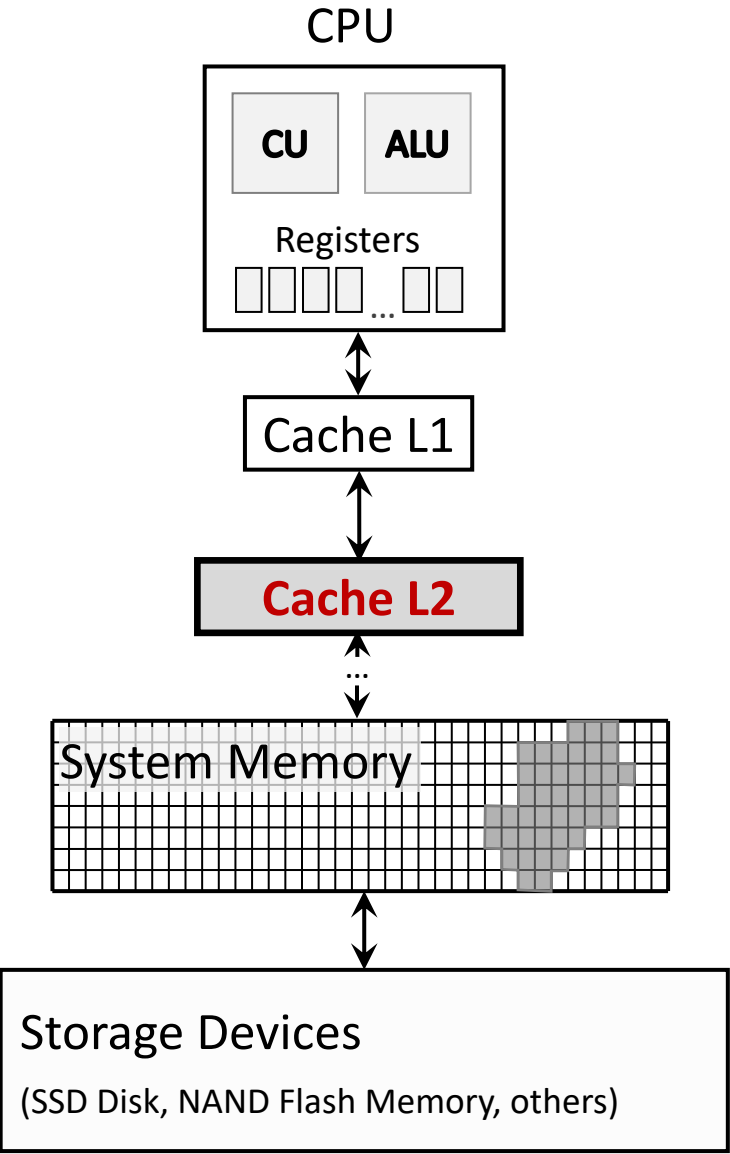
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	

Average Performance Characteristics (Desktop/Laptop)



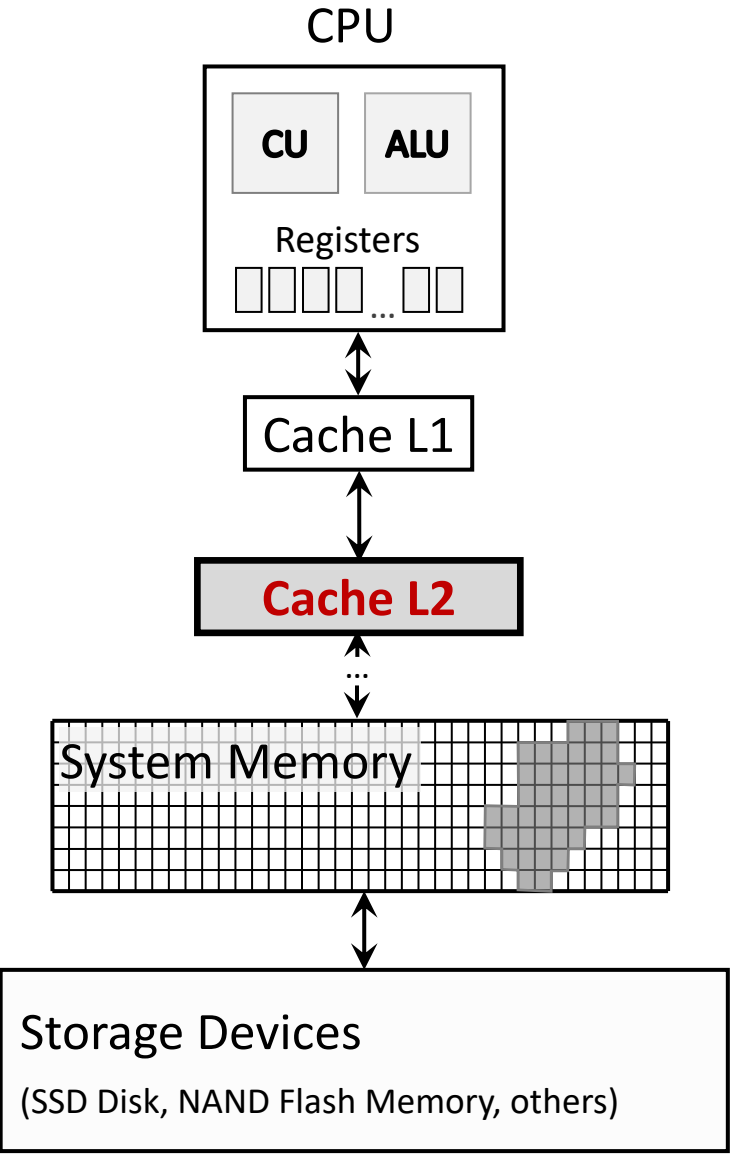
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB

Average Performance Characteristics (Desktop/Laptop)



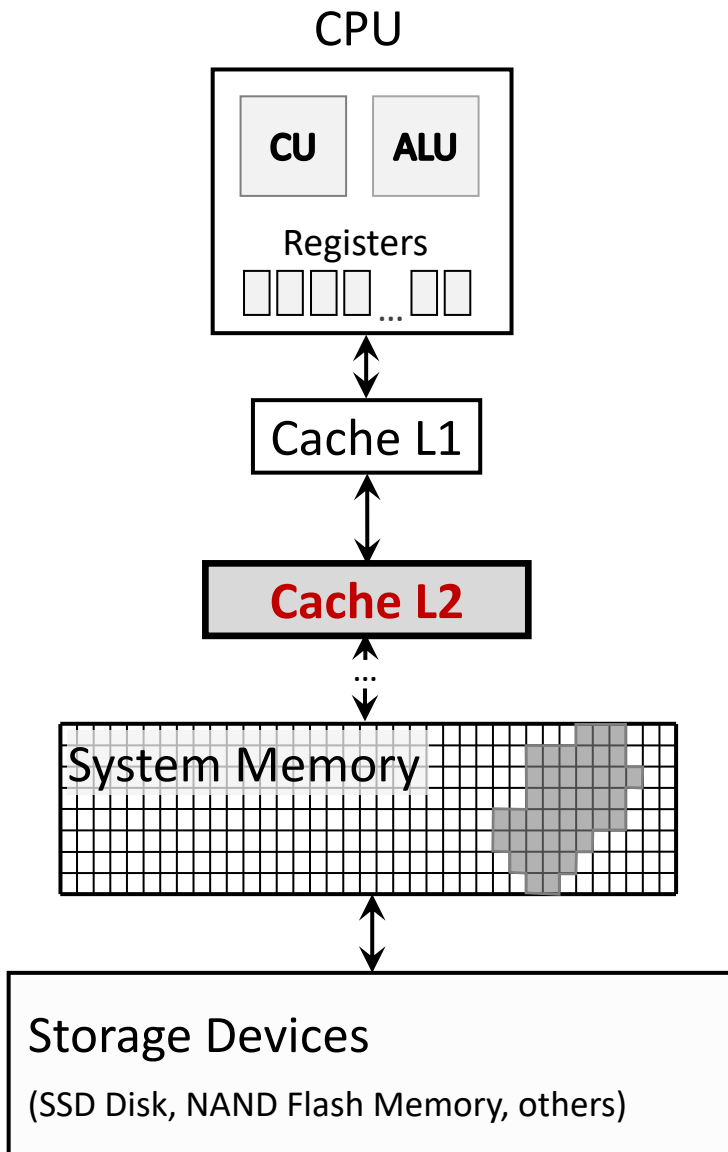
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		

Average Performance Characteristics (Desktop/Laptop)



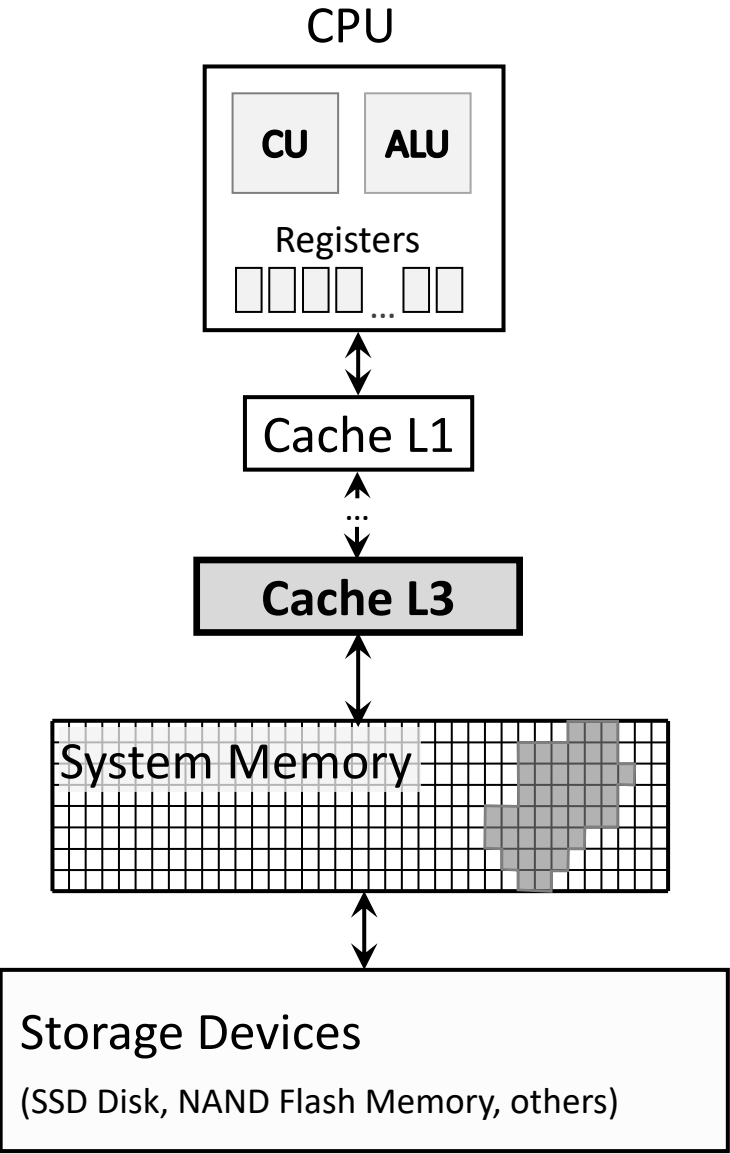
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	

Average Performance Characteristics (Desktop/Laptop)



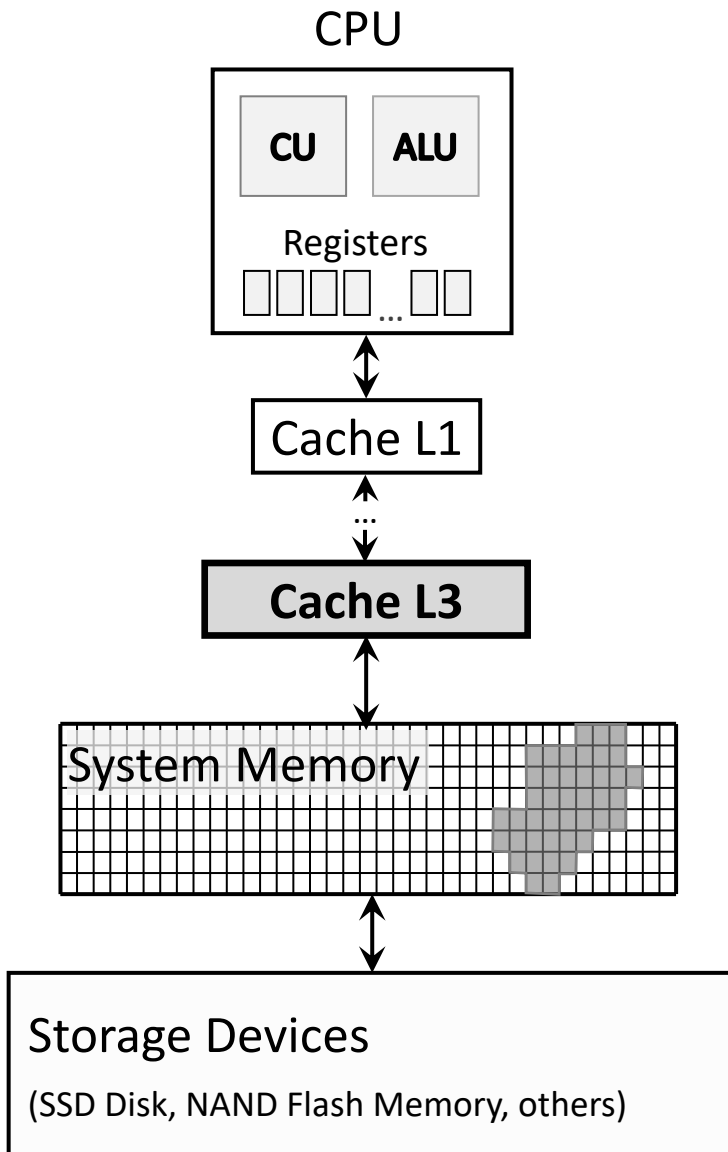
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB

Average Performance Characteristics (Desktop/Laptop)



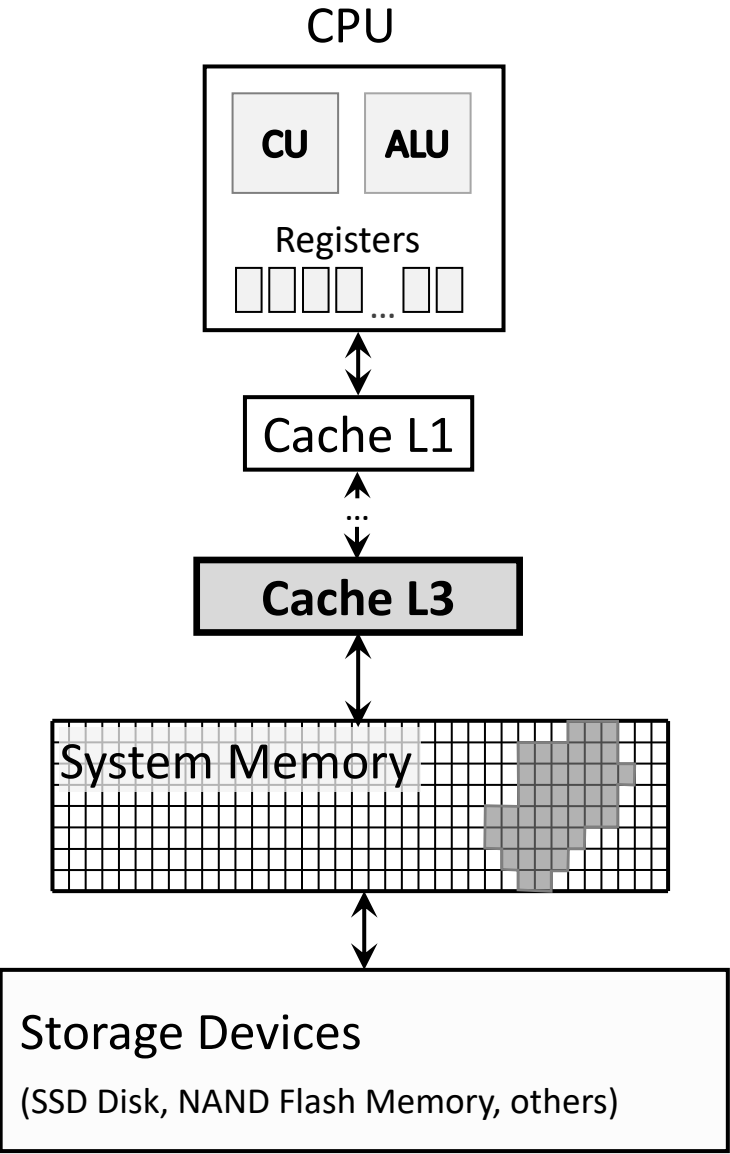
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache		

Average Performance Characteristics (Desktop/Laptop)



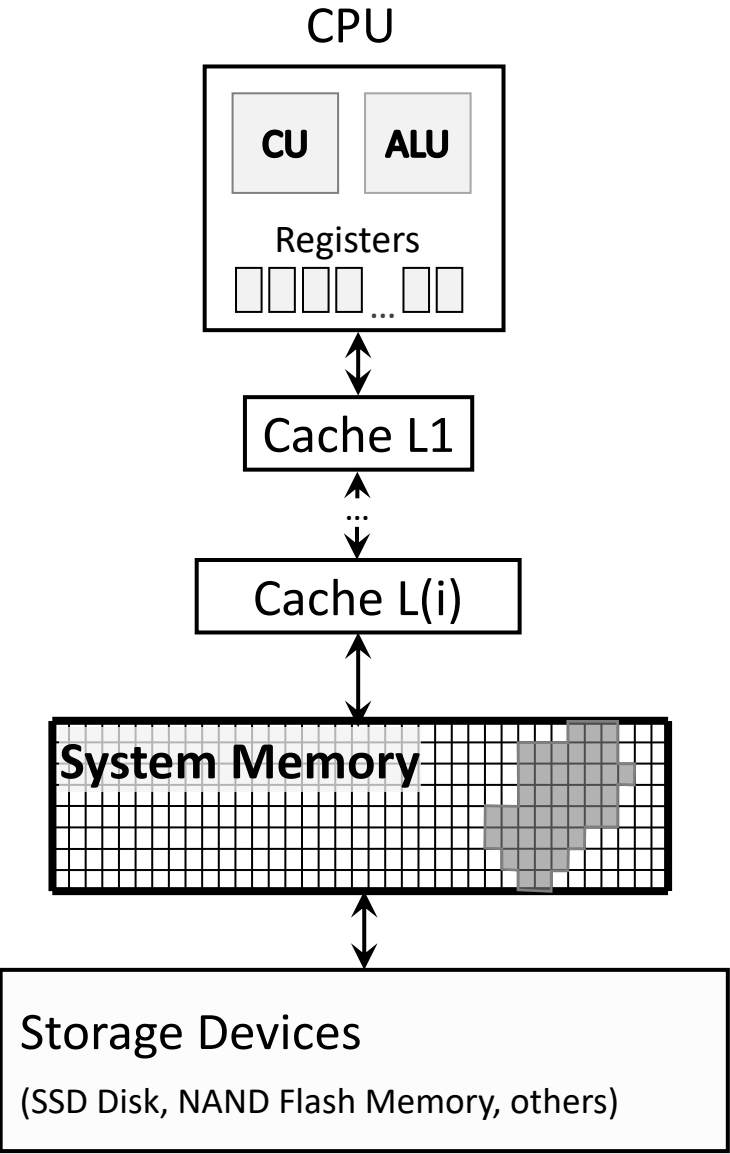
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache	~2 faster than RAM	

Average Performance Characteristics (Desktop/Laptop)



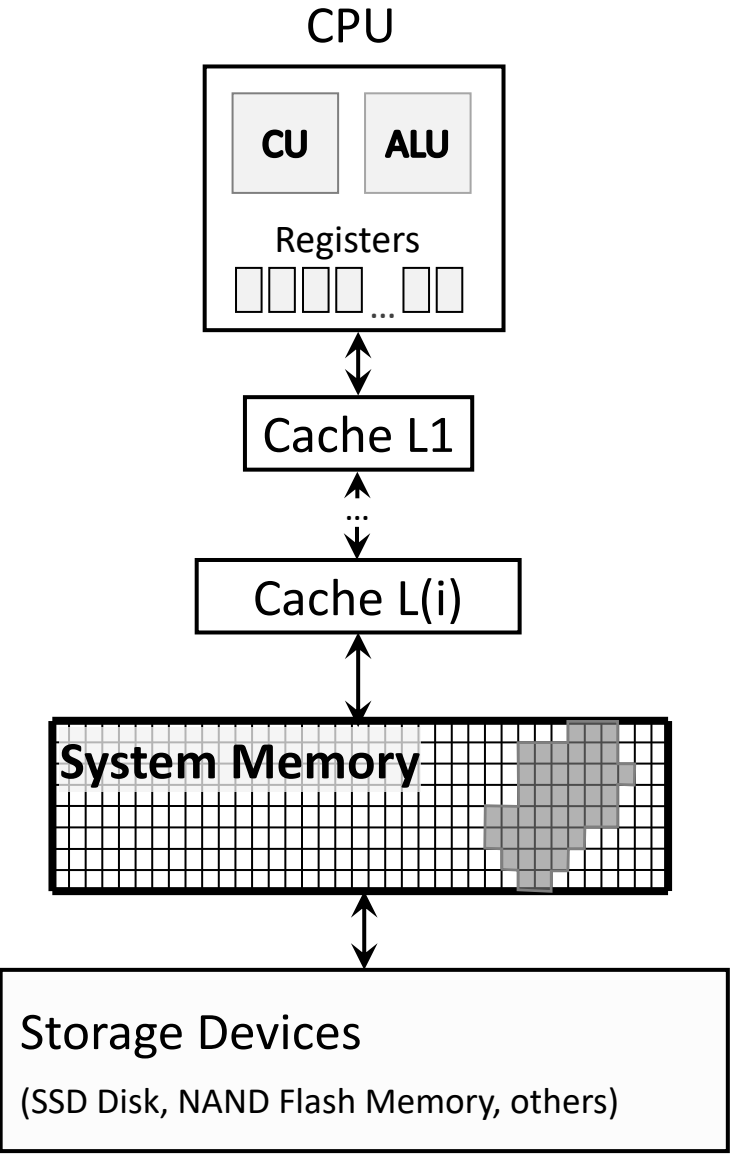
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache	~2 faster than RAM	32-64 MB

Average Performance Characteristics (Desktop/Laptop)



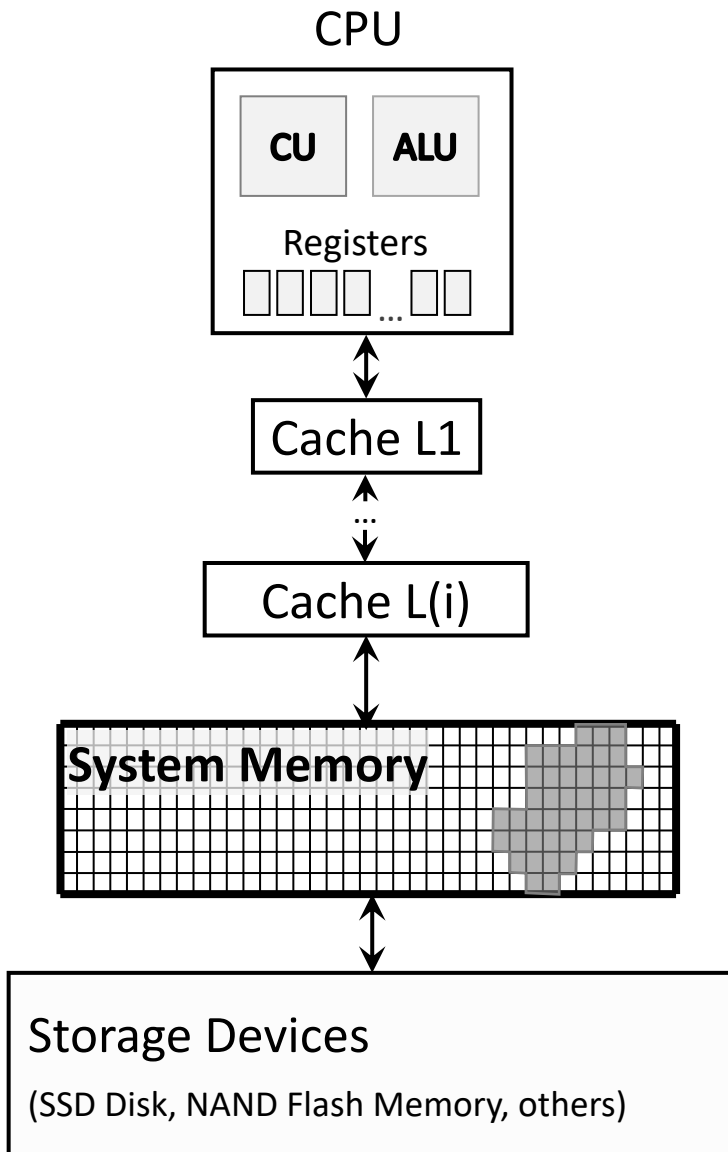
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache	~2 faster than RAM	32-64 MB
System Memory		

Average Performance Characteristics (Desktop/Laptop)



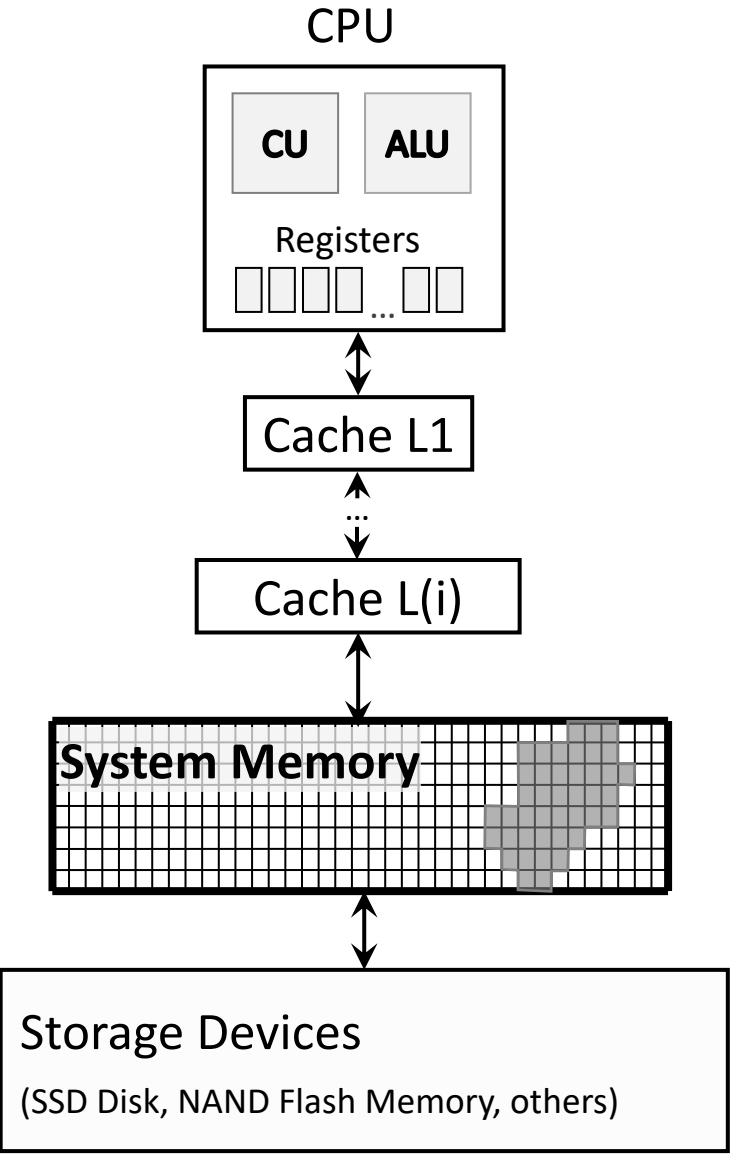
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache	~2 faster than RAM	32-64 MB
System Memory	18-36 GB/sec	

Average Performance Characteristics (Desktop/Laptop)



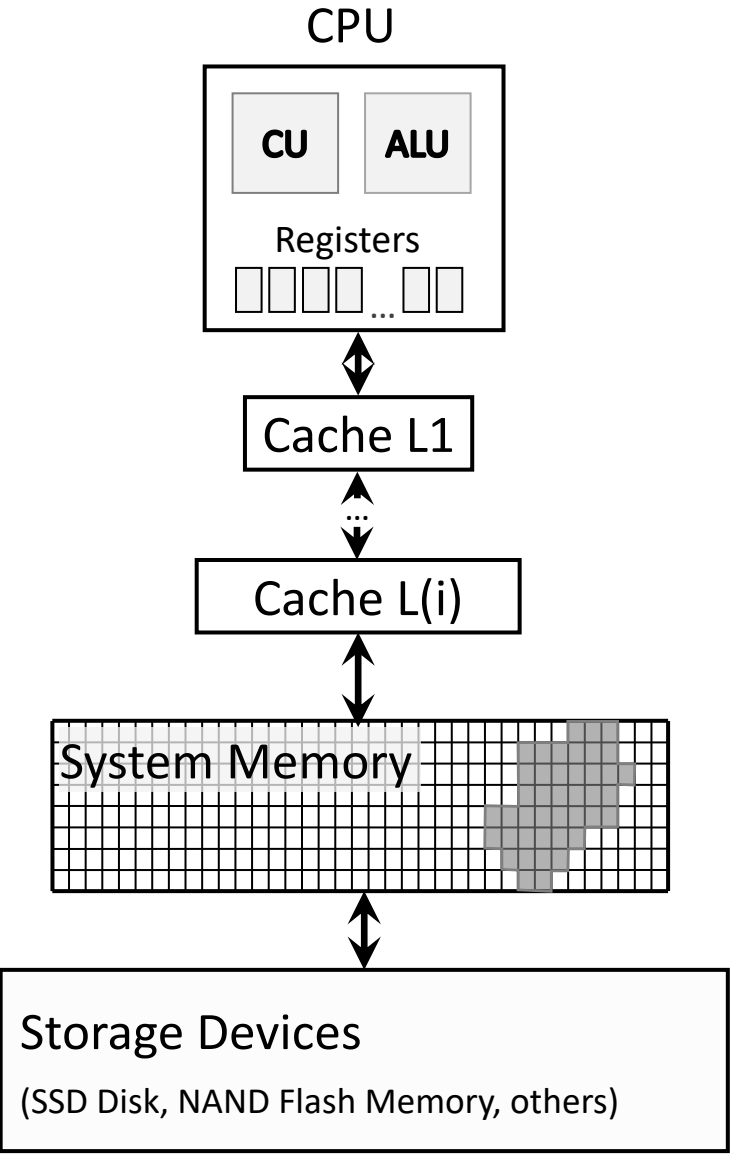
Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache	~2 faster than RAM	32-64 MB
System Memory	18-36 GB/sec	8-32 GB

Average Performance Characteristics (Desktop/Laptop)



Device	Operation Speed	Storage Capacity
CPU (4-6 cores)	2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache	The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache	~25 faster than RAM	256-512 KB
L3 Cache	~2 faster than RAM	32-64 MB
System Memory	18-36 GB/sec	8-32 GB
Communication Bus		

Average Performance Characteristics (Desktop/Laptop)



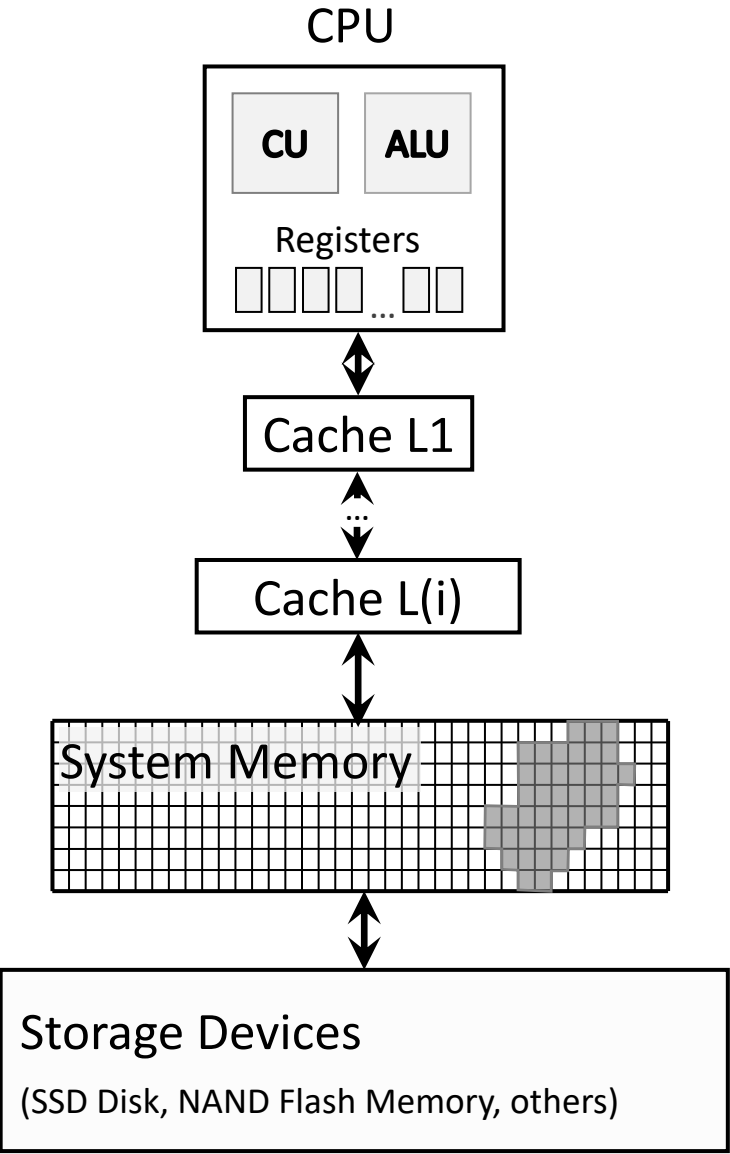
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)		
	PCIe (4.0)		
	USB (3.0)		

RAM (Random Access Memory) – synonym for “System Memory”

PCIe: Peripheral Component Interconnect Express

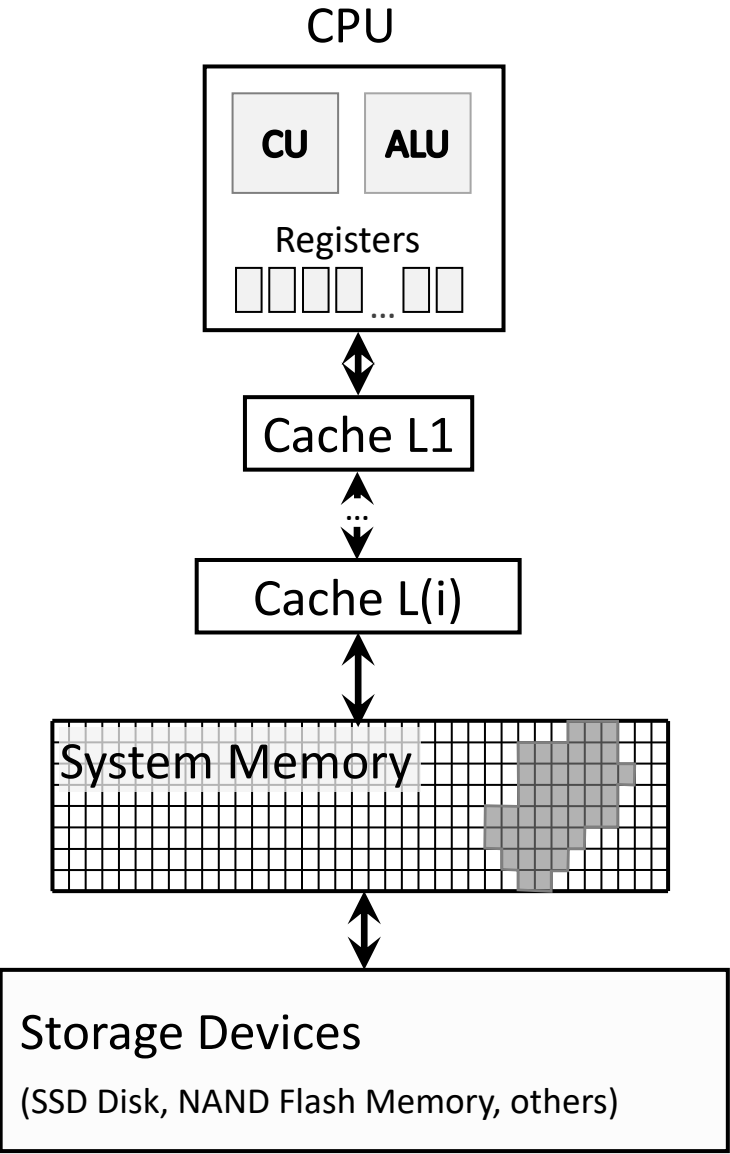
USB: Universal Serial Bus

Average Performance Characteristics (Desktop/Laptop)



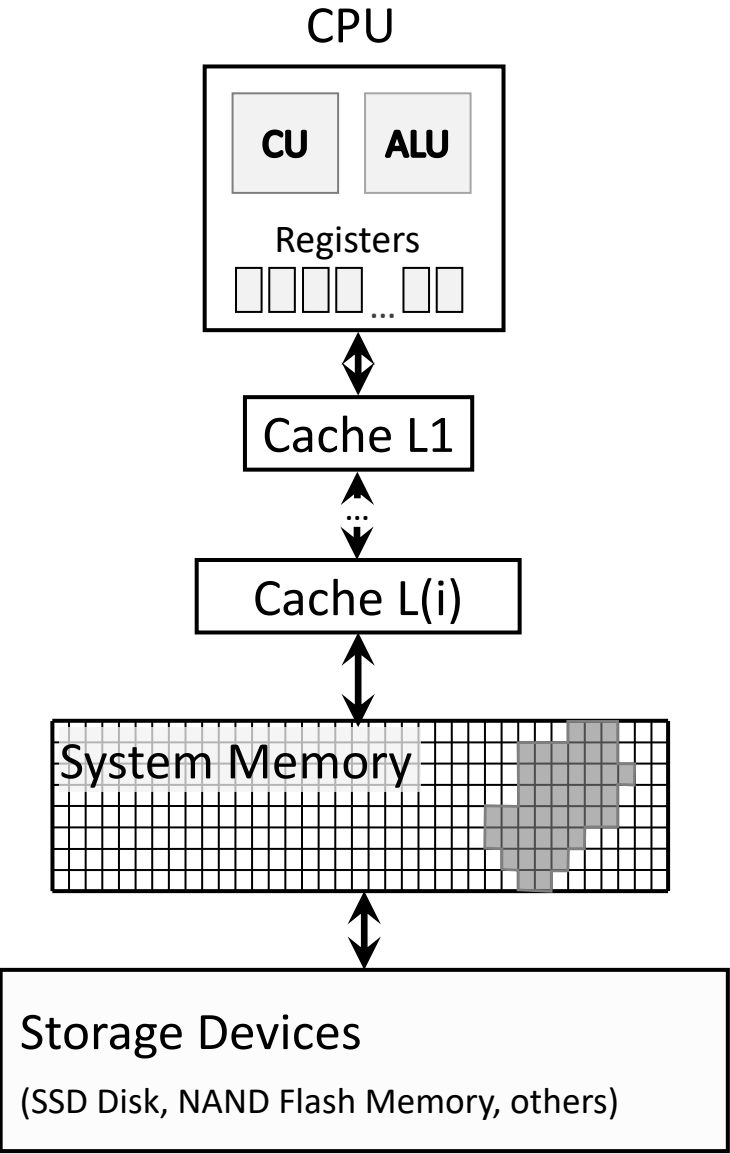
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	
	PCIe (4.0)		
	USB (3.0)		

Average Performance Characteristics (Desktop/Laptop)



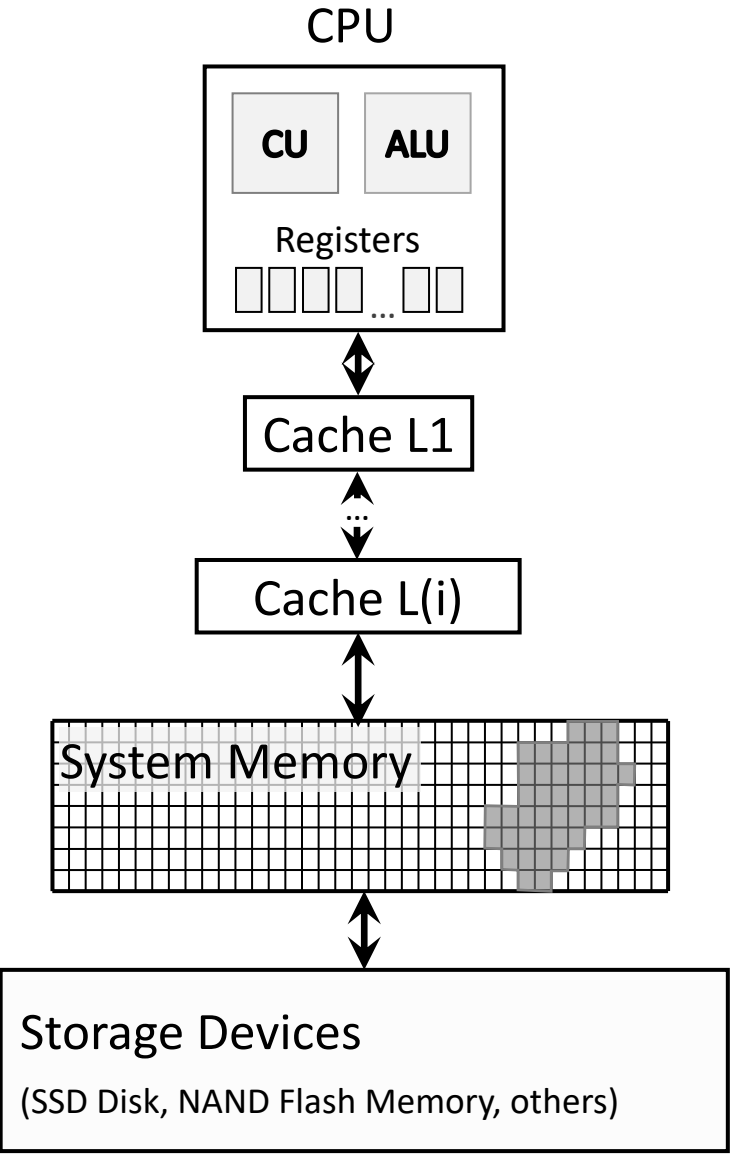
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	The number of lines reflects the number of simultaneously transferred bits
	PCIe (4.0)		
	USB (3.0)		

Average Performance Characteristics (Desktop/Laptop)



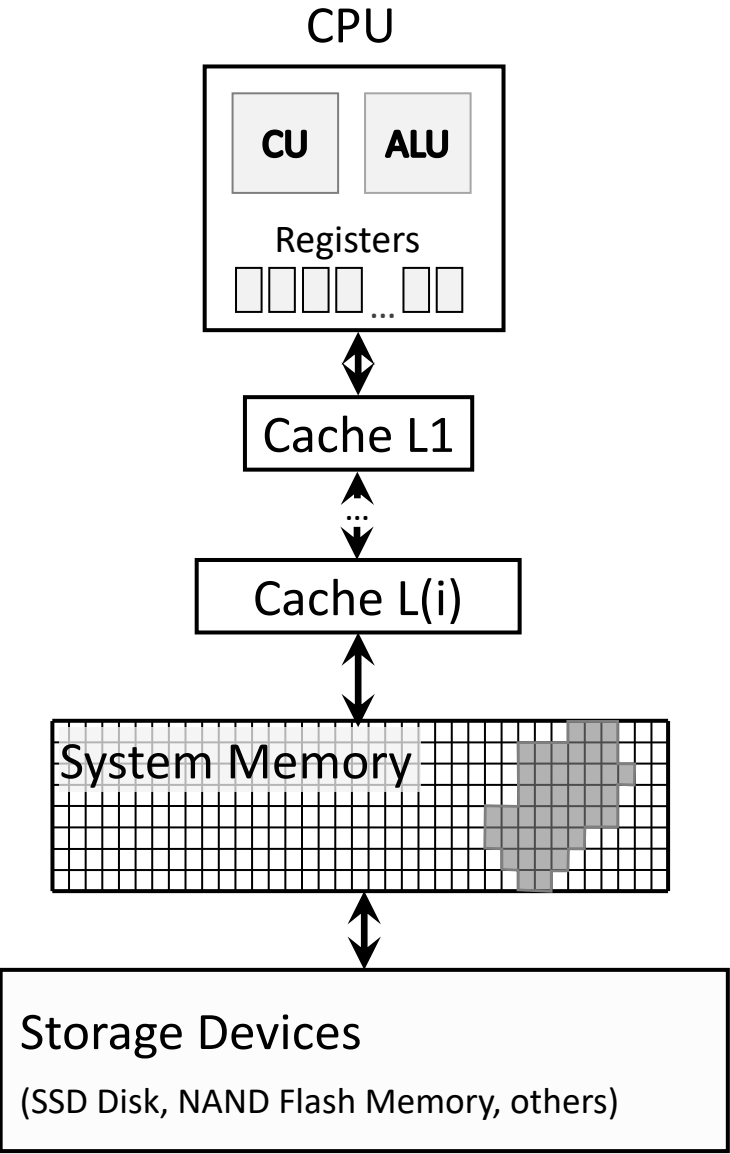
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)		

Average Performance Characteristics (Desktop/Laptop)



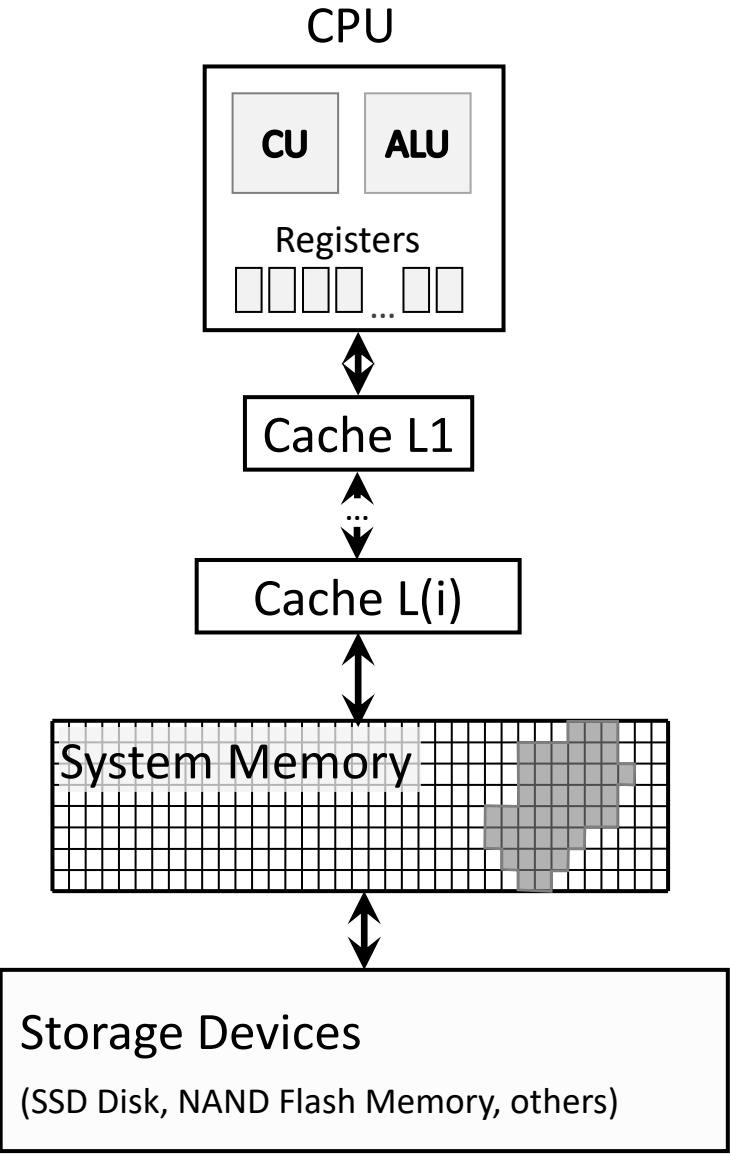
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	

Average Performance Characteristics (Desktop/Laptop)



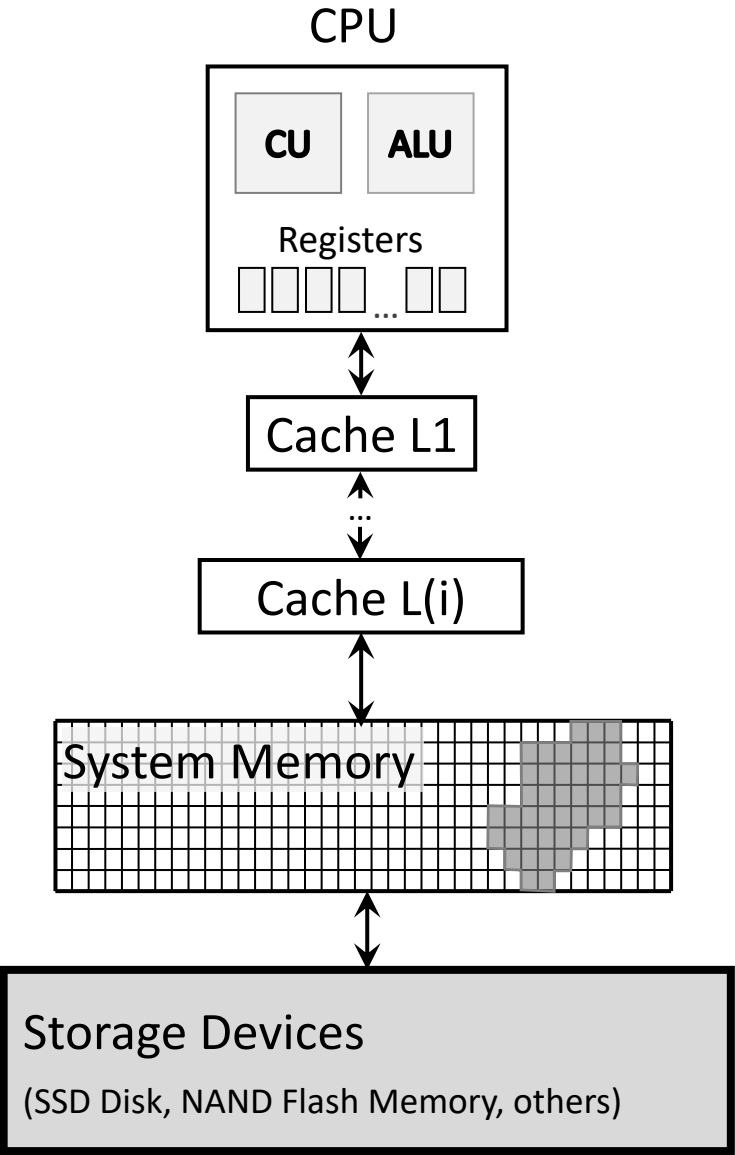
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	

Average Performance Characteristics (Desktop/Laptop)



Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage Device			

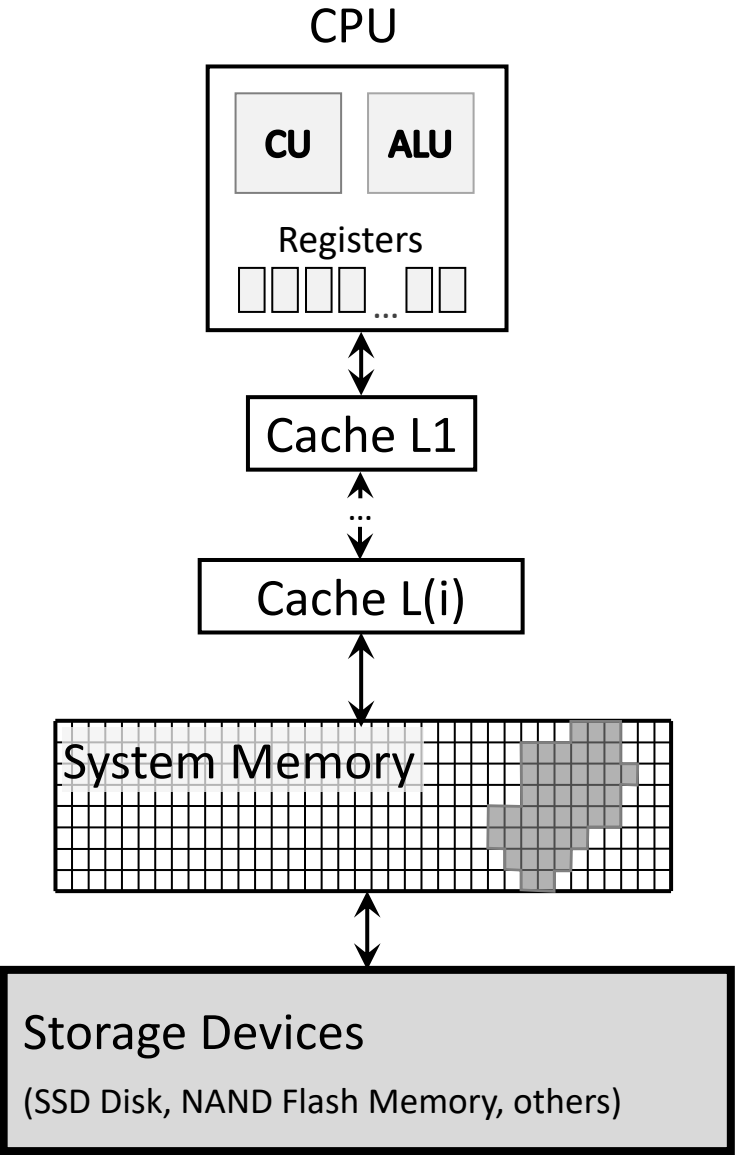
Average Performance Characteristics (Desktop/Laptop)



Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD		
	HDD		

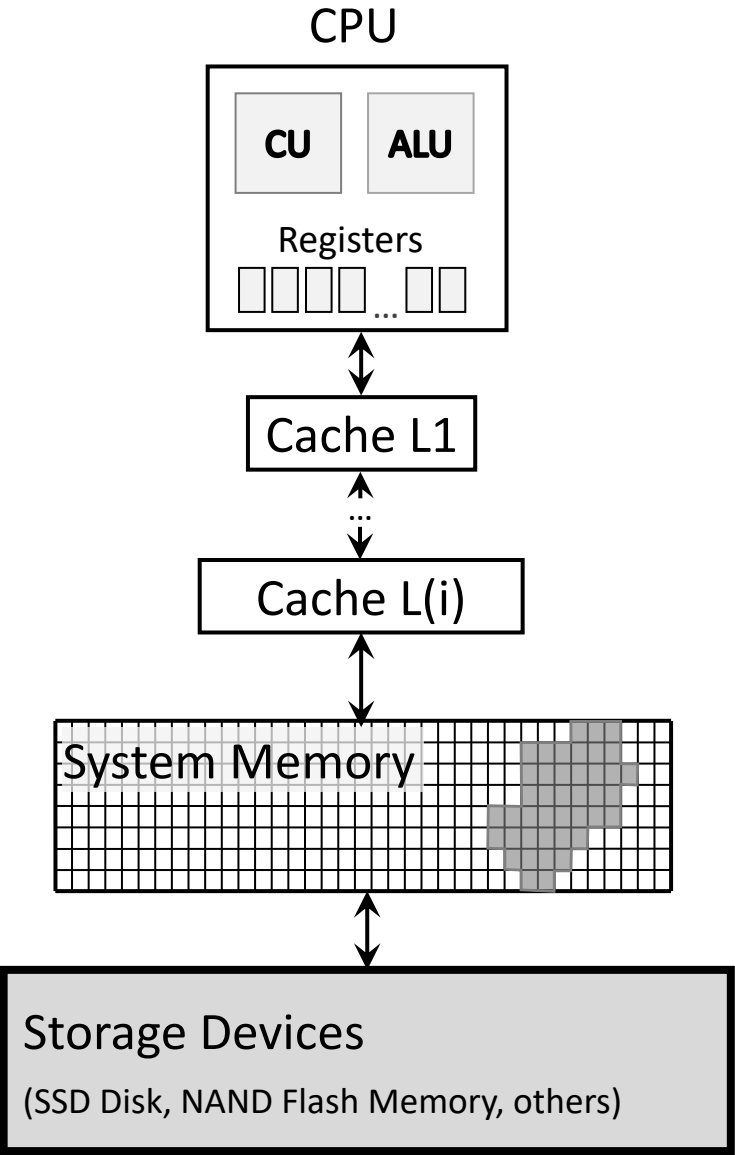
SSD: Solid State Drive; HDD: Hard Disk Drive

Average Performance Characteristics (Desktop/Laptop)



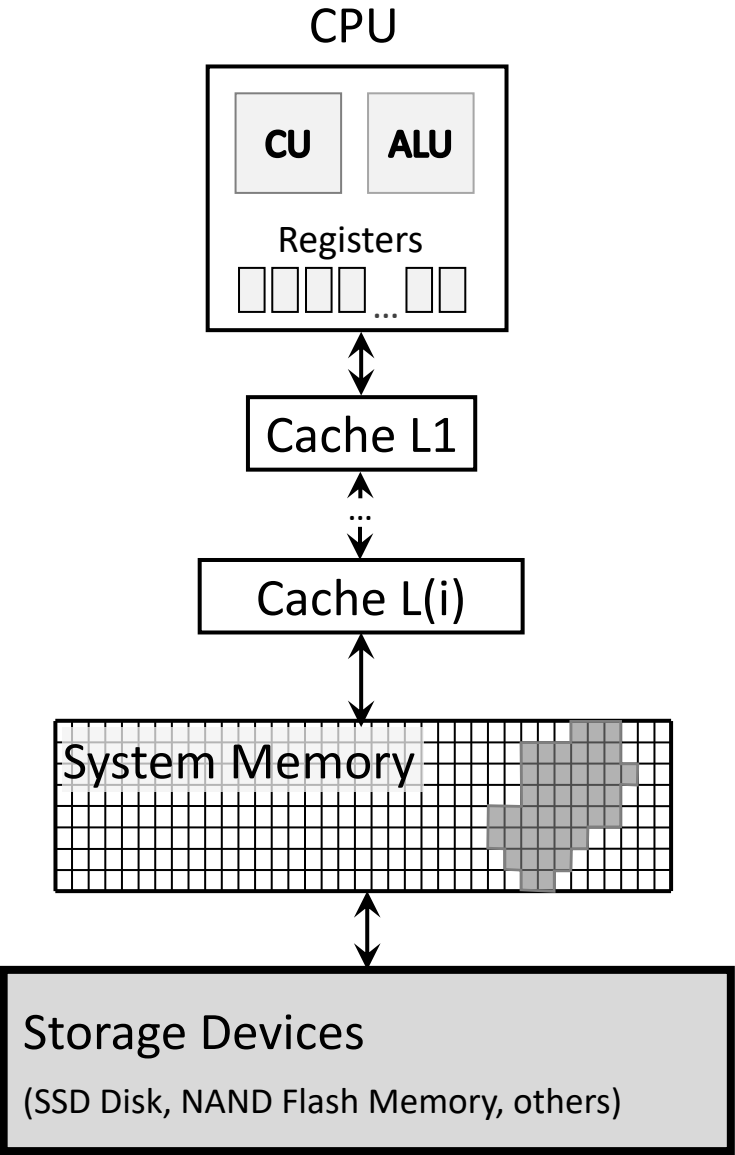
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	
	HDD		

Average Performance Characteristics (Desktop/Laptop)



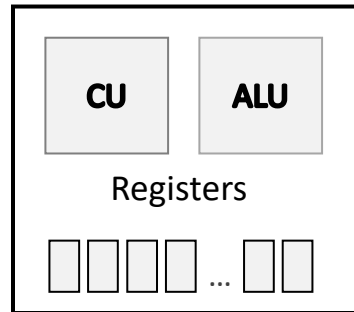
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	
	HDD	0.1-0.15 GB/sec	

Average Performance Characteristics (Desktop/Laptop)



Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

CPU



Cache L1

...

Cache L(i)

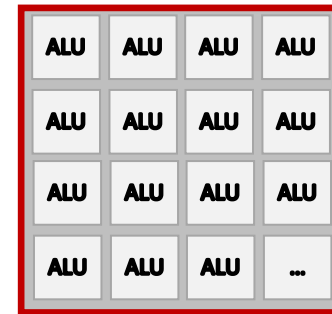
System Memory

Storage Devices

(SSD Disk, NAND Flash Memory, others)

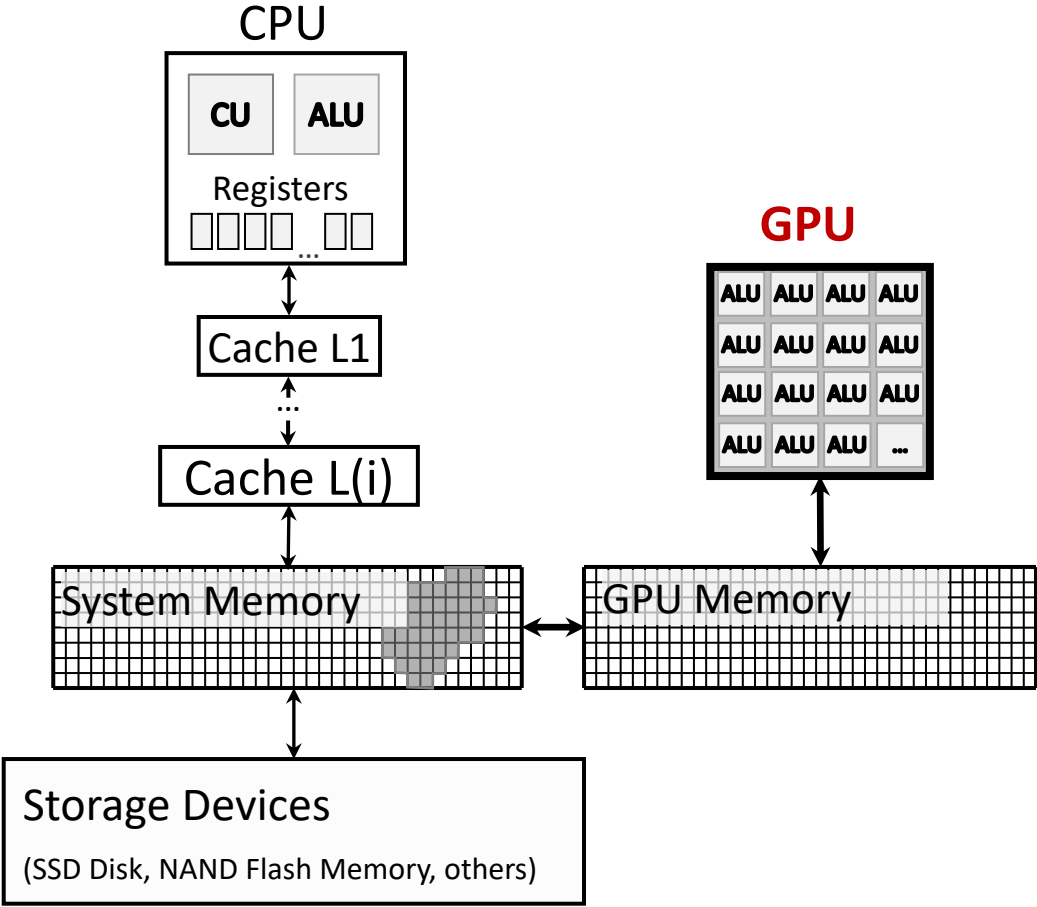
GPU (Graphics Processing Unit) Subsystem

GPU



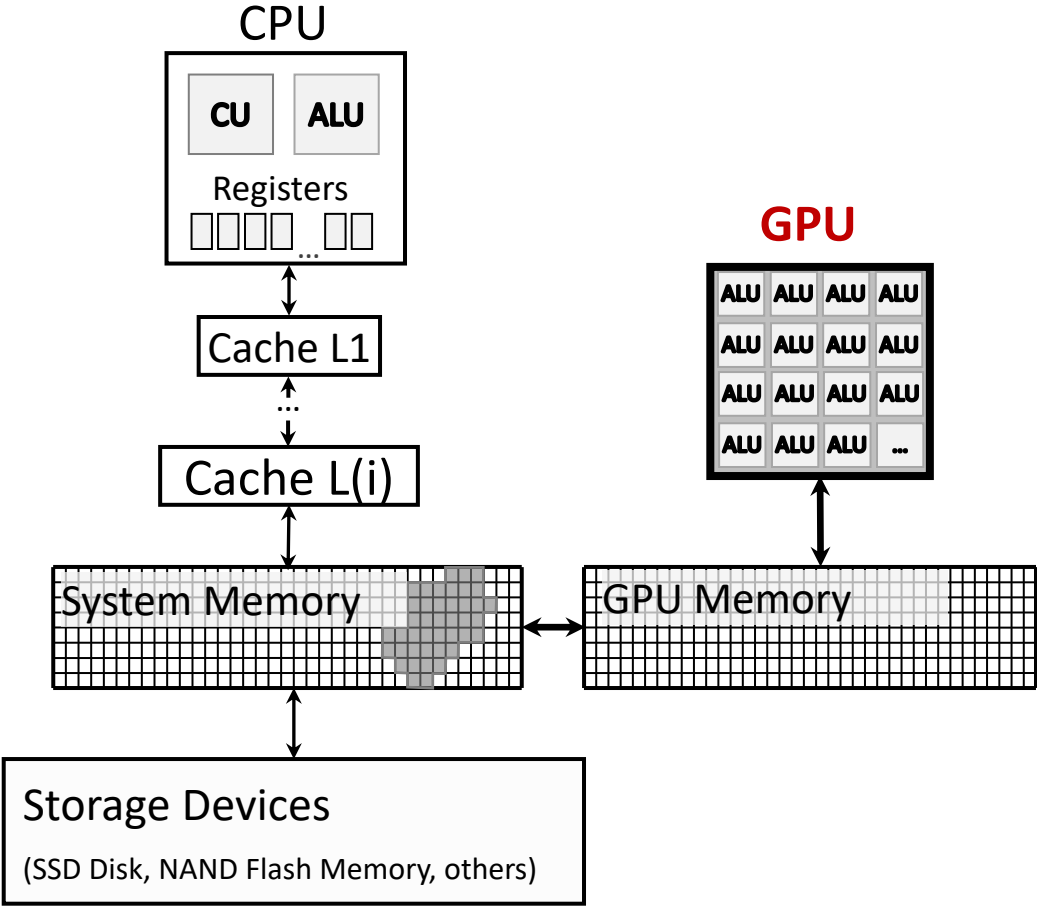
GPU Memory

Average Performance Characteristics (Desktop/Laptop)



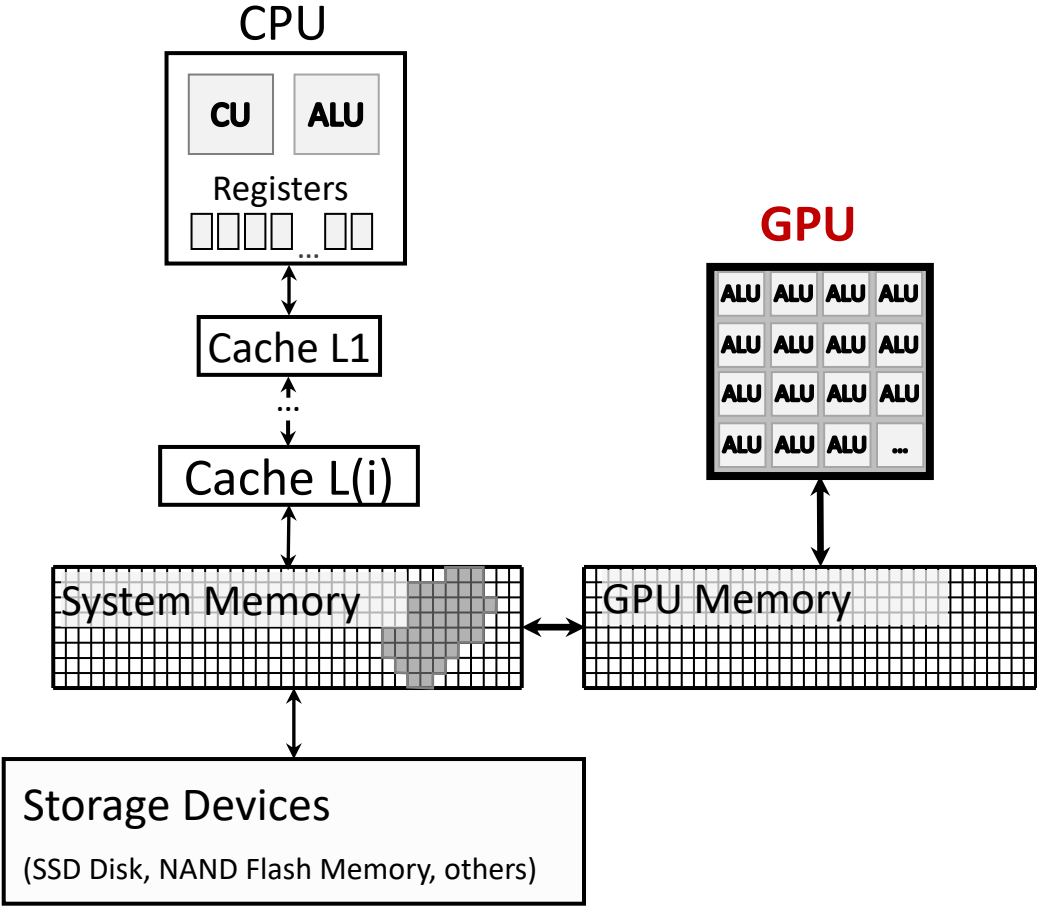
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)			
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



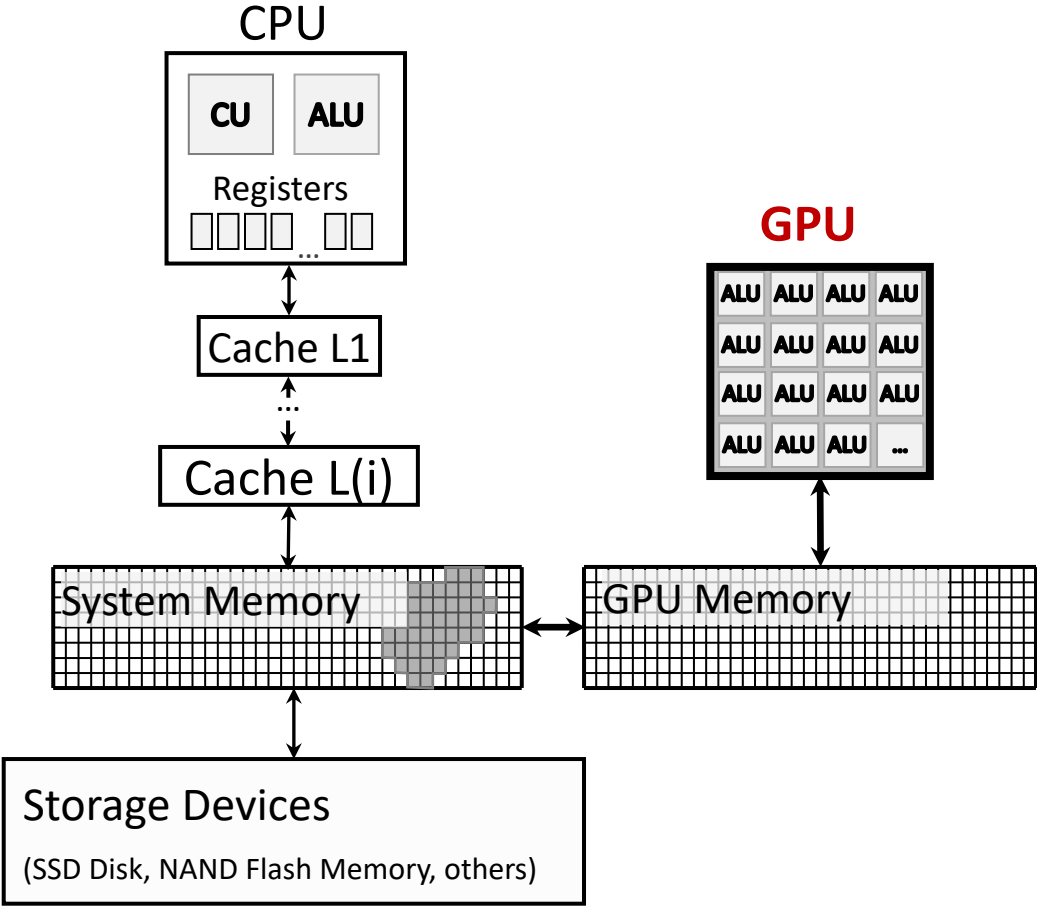
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz	
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



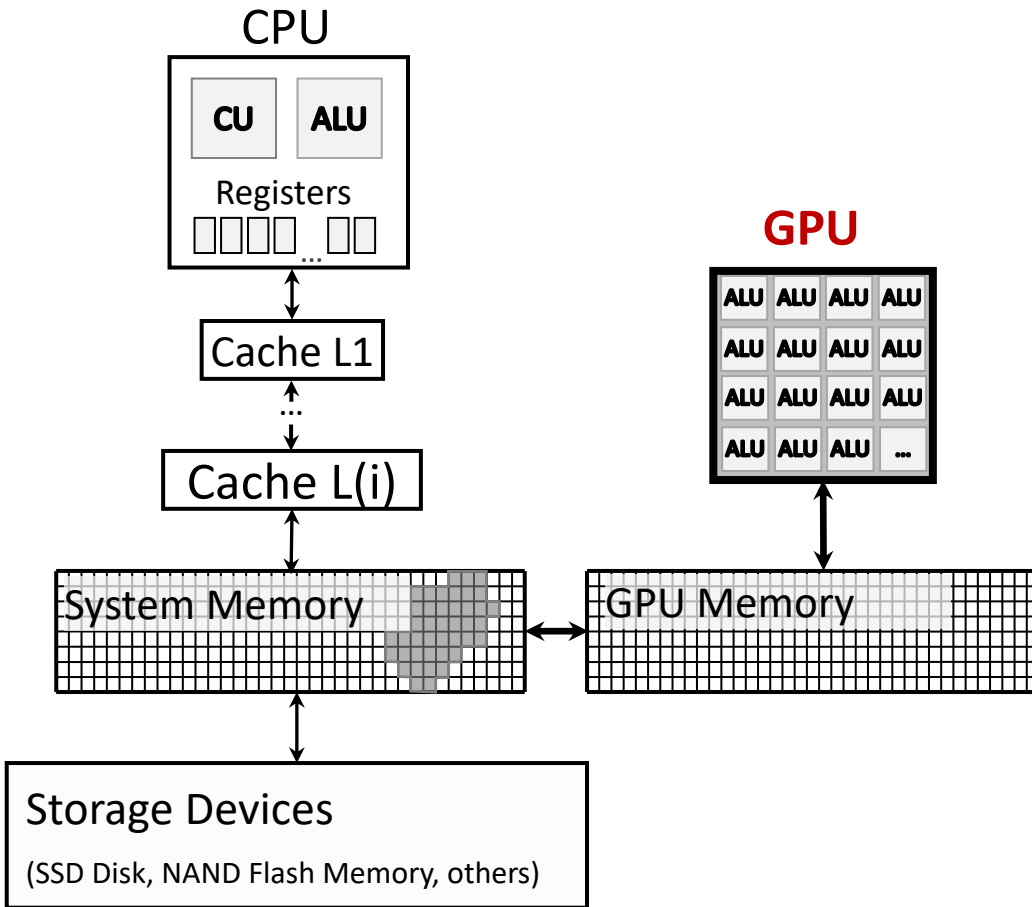
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple Instruction Set (IS) supported	
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



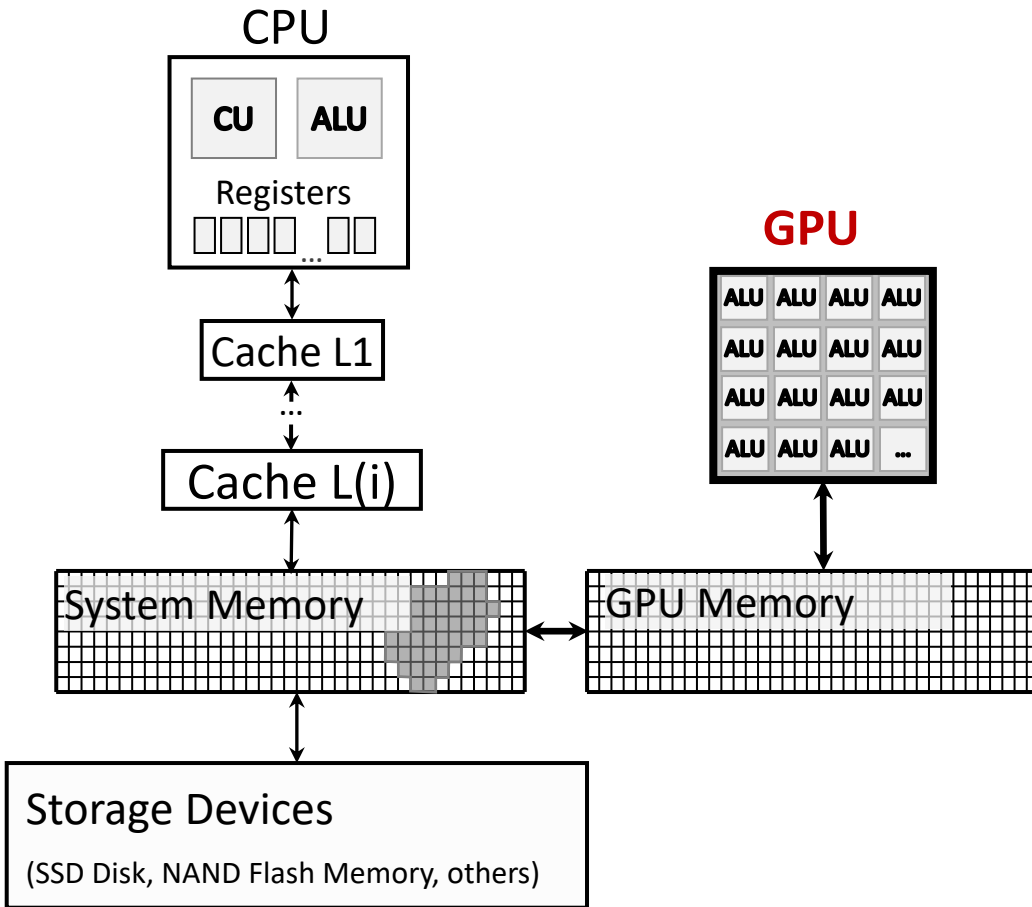
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; Parallel execution (SIMD)	
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



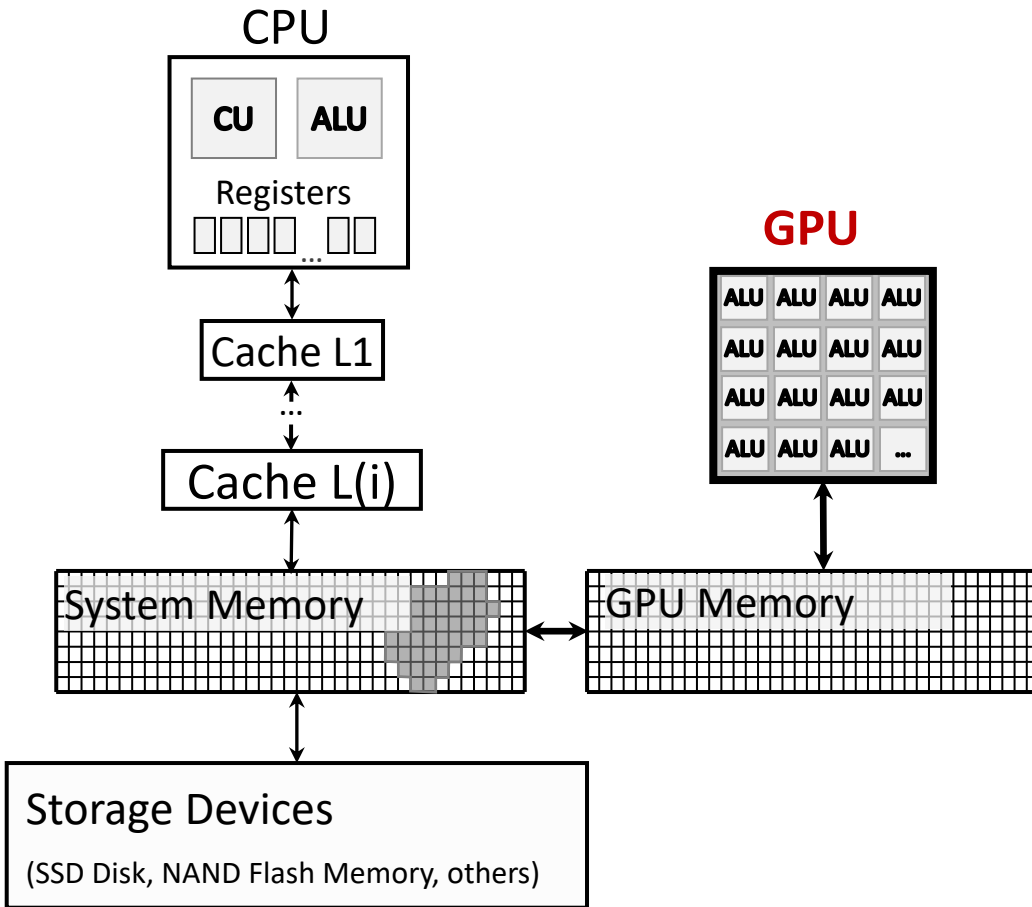
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



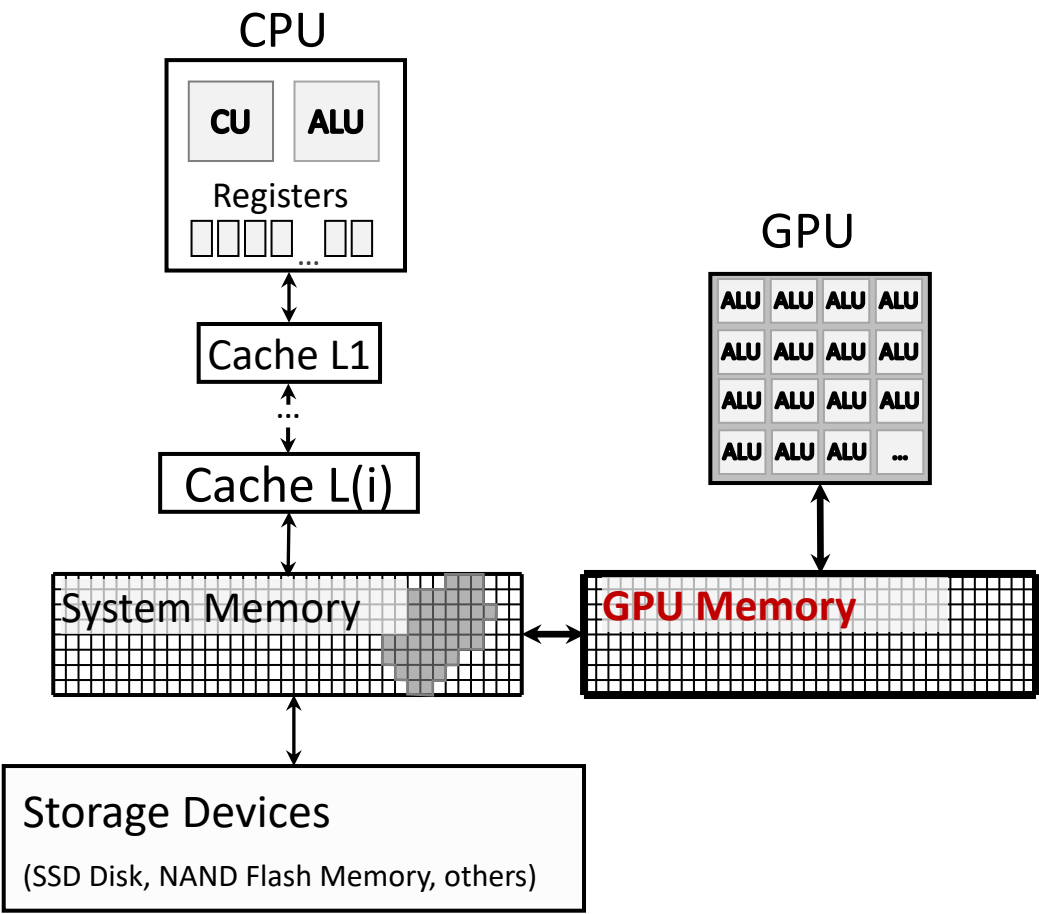
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



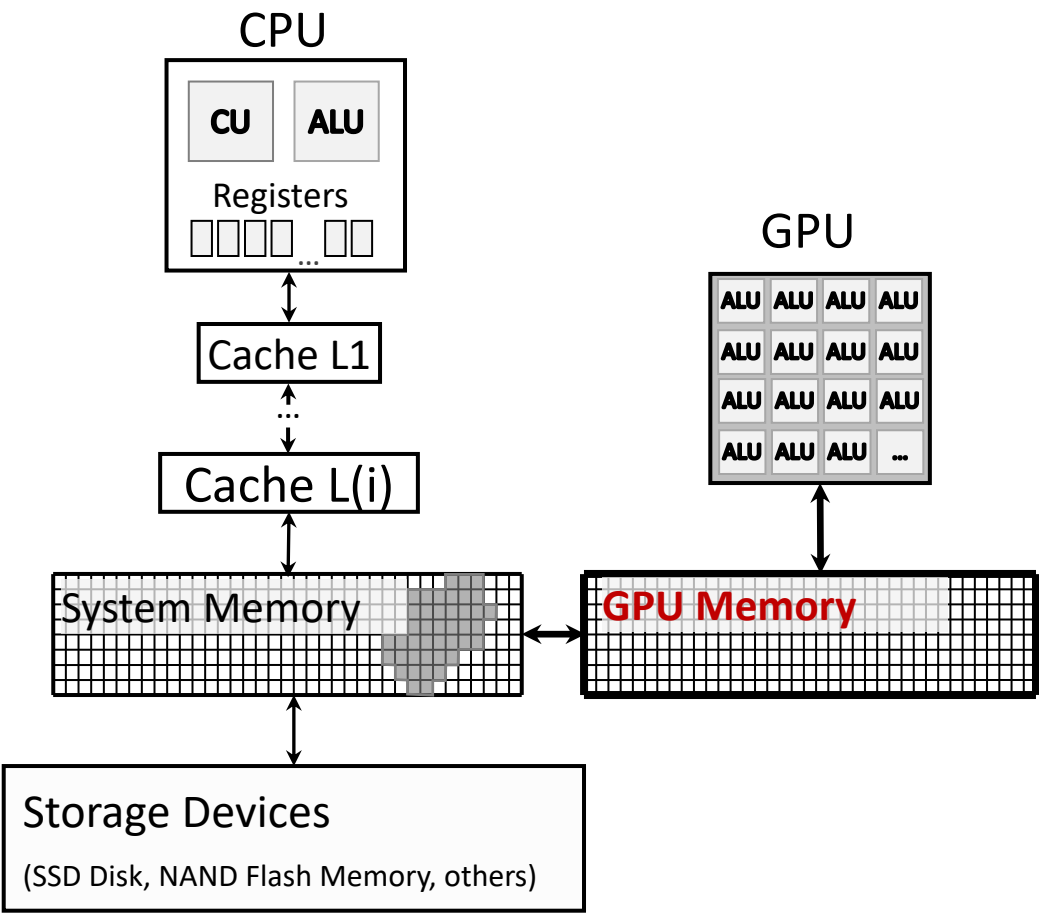
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	32 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



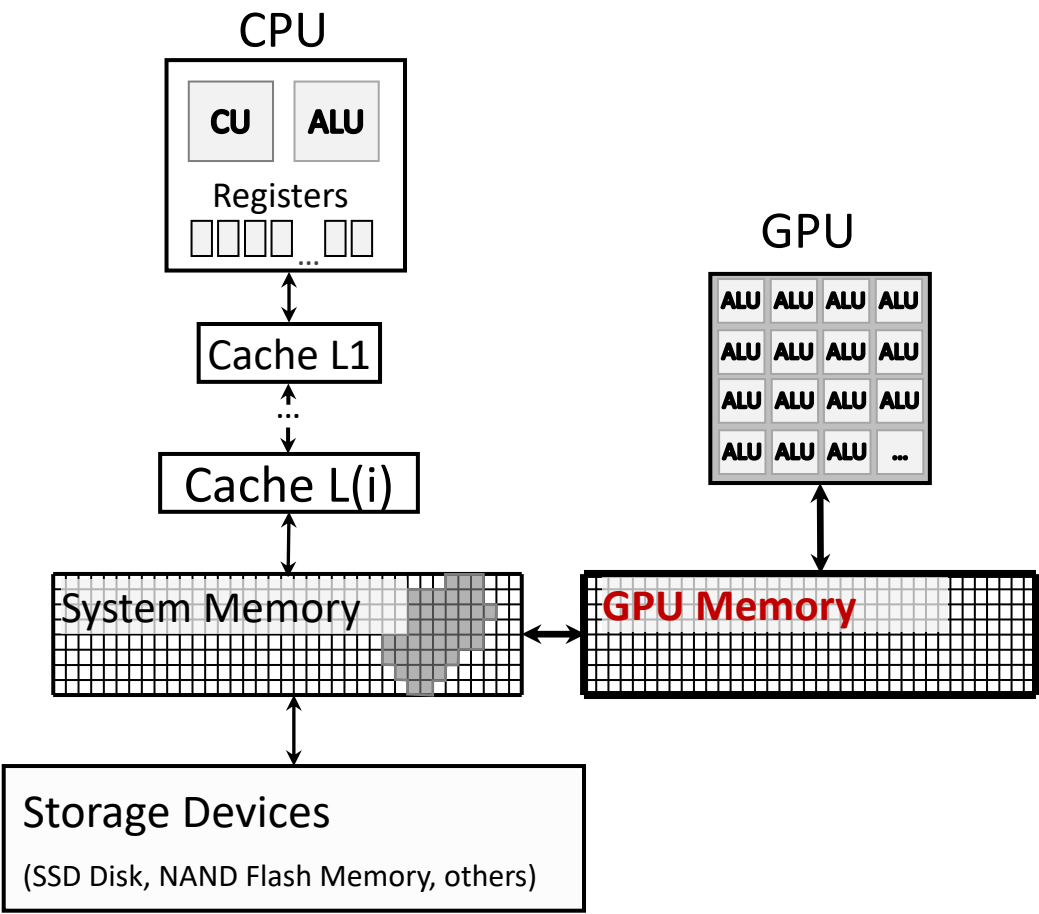
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	32 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
GPU Memory			
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



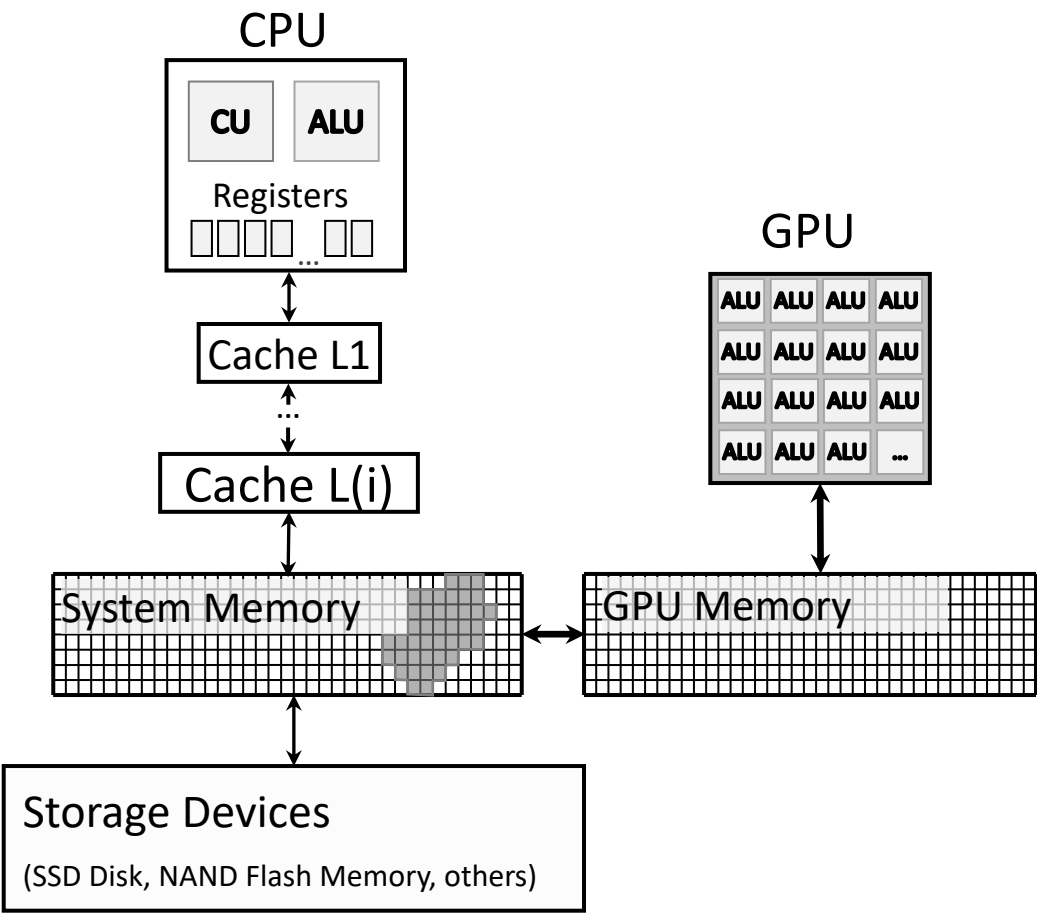
Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	32 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
GPU Memory		300-500 GB/sec (due to high throughput)	
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	32 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
GPU Memory		300-500 GB/sec (due to high throughput)	8-16 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Average Performance Characteristics (Desktop/Laptop)



Device		Operation Speed	Storage Capacity
CPU (4-6 cores)		2.0-3.5 GHz Up to 5.0 GHz (Turbo mode)	32 or 64 Bits per register
GPU (700-2000 cores)		1.5-2.0 GHz Simple IS; parallel execution ~100 times faster than CPU	32 Bits per register
L1 Cache		The same as CPU speed; ~100 times faster than RAM	32-128 KB
L2 Cache		~25 faster than RAM	256-512 KB
L3 Cache		~2 faster than RAM	32-64 MB
System Memory		18-36 GB/sec	8-32 GB
GPU Memory		300-500 GB/sec (due to high throughput)	8-16 GB
Bus	RAM (DDR4)	18-36 GB/sec (64 lines)	N/A
	PCIe (4.0)	8 GB/sec (4 lines)	
	USB (3.0)	0.5 GB/sec (1 line)	
Storage	SSD	0.5-2 GB/sec	250-500 GB
	HDD	0.1-0.15 GB/sec	100-150 GB

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on the execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on the execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Question:

What is a program execution speedup if increasing the number of cores from 1 to m ?

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on the execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Question:

What is a program execution speedup if increasing the number of cores from 1 to m ?

Solution?

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on the execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Question:

What is a program execution speedup if increasing the number of cores from 1 to m ?

Solution:

1) $(100 - x)\%$ can execute in parallel upon m cores

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on the execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Question:

What is a program execution speedup if increasing the number of cores from 1 to m ?

Solution:

- 1) $(100 - x)\%$ can execute in parallel upon m cores;
- 2) These $(100 - x)\%$ execute in parallel m times faster**

Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on the execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Question:

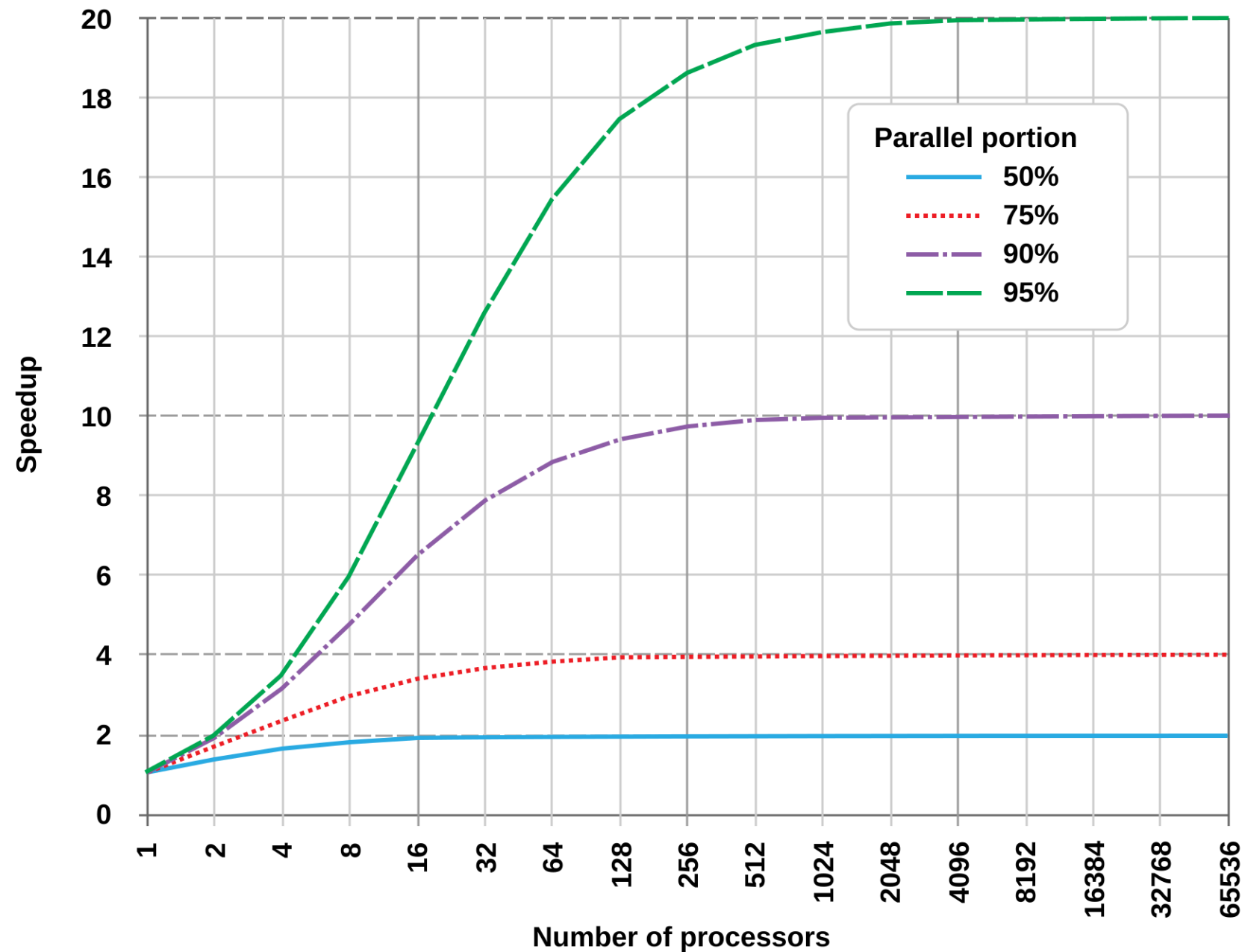
What is a program execution speedup if increasing the number of cores from 1 to m ?

Solution:

- 1) $(100 - x)\%$ can execute in parallel upon m cores;
- 2) These $(100 - x)\%$ execute in parallel m times faster;
- 3) **According to Amdahl's law, the runtime of such a program reduces according to:**

$$\textit{speedup} = \frac{t^{1 \text{ proc}}}{t^{m \text{ procs}}} = \frac{1}{x + \frac{1-x}{m}}$$

Amdahl's Law



Amdahl's Law

Provides a theoretical upper bound on the execution speedup thanks to multicores use

Program given:

- 1) $x\%$ of instructions depend on execution results of previous instructions;
- 2) $(100 - x)\%$ of instructions do not depend on previous ones;

Question:

What is a program execution speedup if increasing the number of cores from 1 to m ?

Solution:

- 1) $(100 - x)\%$ can execute in parallel upon m cores;
- 2) These $(100 - x)\%$ execute in parallel m times faster;
- 3) According to Amdahl's law, the runtime of such a program reduces according to:

$$\text{speedup} = \frac{t^{1 \text{ proc}}}{t^{m \text{ procs}}} = \frac{1}{x + \frac{1-x}{m}}$$

