CPU Architecture and Terminology Cheatsheet

A concise reference for understanding CPU architecture, key terminology, and performance metrics. Useful for students, developers, and anyone interested in the inner workings of computer hardware.

Core CPU Concepts

Fundamental Components

ALU (Arithmetic Logic Unit)	Performs arithmetic and logical operations.
Control Unit	Fetches instructions, decodes them, and controls the execution flow.
Registers	Small, high-speed storage locations used to hold data and instructions being processed.
Cache Memory	Fast memory used to store frequently accessed data, reducing access time to main memory.
Bus Interface	Connects the CPU to other components like memory and peripherals.
Clock	Provides timing signals to synchronize operations within the CPU. Measured in Hertz (Hz).

CPU Operation Cycle

1. **Fetch:** Retrieve the instruction from memory.
2. **Decode:** Interpret the instruction.
3. **Execute:** Perform the operation specified by the instruction.
4. **Store:** Write the result back to memory or a register.

Instruction Set Architecture (ISA)

Definition	Defines the set of instructions a CPU can execute. Examples: x86, ARM, RISC-V.
CISC (Complex Instruction Set Computing)	Features a large set of complex instructions. Example: x86.
RISC (Reduced Instruction Set Computing)	Features a smaller set of simpler instructions. Example: ARM.

CPU Performance Metrics

Clock Speed and IPC

Clock Speed	The rate at which a CPU executes instructions, measured in GHz. Higher clock speed generally means faster performance, but it's not the only factor.
IPC (Instructions Per Cycle)	The average number of instructions a CPU can execute per clock cycle. A higher IPC indicates a more efficient architecture.
Relationship	Performance is a product of both clock speed and IPC: `Performance ≈ Clock Speed * IPC`

Core Count and Multithreading

Core	An independent processing unit within a CPU. More cores generally allow for better multitasking and parallel processing.
Multithreading (e.g., Hyper-Threading)	Allows a single core to execute multiple threads concurrently, improving resource utilization. It makes the operating system recognize one physical core as two virtual cores.
Effect on Performance	More cores and efficient multithreading improve performance in multi-threaded applications and workloads. However, single-threaded applications may not benefit significantly.

Cache Levels

L1 Cache	Smallest and fastest cache, closest to the core. Usually split into L1i (instruction cache) and L1d (data cache).
L2 Cache	Larger and slower than L1, but still faster than main memory. Serves as a secondary cache for data not found in L1.
L3 Cache	Largest and slowest cache, shared by all cores. Further reduces access time to main memory.

Other Important Metrics

TDP (Thermal Design Power)	The maximum amount of heat a CPU is expected to dissipate under normal operating conditions. Indicates cooling requirements.
Power Consumption	The amount of power the CPU consumes during operation. Lower power consumption is desirable for energy efficiency.
Manufacturing Process (e.g., 7nm, 5nm)	Smaller manufacturing processes generally result in higher transistor density, improved performance, and lower power consumption.
Bandwidth	Rate at which data can be read from or stored into a storage unit. Represented as bits per second or bytes per second.

CPU Architecture Types

Desktop and Server CPUs

Characteristics	Designed for high performance and multitasking. Typically have higher clock speeds, more cores, and larger caches.
Examples	Intel Core i9, AMD Ryzen 9, Intel Xeon, AMD EPYC
Typical Use	Gaming, content creation, scientific computing, server applications.

Embedded CPUs

Characteristics	Designed for specific tasks in embedded systems. Often have low power consumption and real-time capabilities.
Examples	ARM Cortex-M series, Microchip PIC, Atmel AVR
Typical Use	Microcontrollers, IoT devices, industrial control systems, automotive electronics.

Mobile CPUs

Characteristics	Optimized for power efficiency and battery life. Typically have lower clock speeds and fewer cores compared to desktop CPUs.
Examples	ARM Cortex-A series, Qualcomm Snapdragon, Apple Silicon (M1, M2)
Typical Use	Smartphones, tablets, laptops.

GPU (Graphics Processing Unit) as a CPU

Characteristics	Specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device.
Examples	NVIDIA GeForce, AMD Radeon
Typical Use	Video and/or image processing and rendering.

Advanced CPU Features

Virtualization

Definition	Allows multiple operating systems to run concurrently on a single physical machine. CPU features like Intel VT-x and AMD-V provide hardware support for virtualization.
Benefits	Improved resource utilization, easier management, and increased flexibility.

Security Features

Examples	Intel SGX (Software Guard Extensions), AMD SEV (Secure Encrypted Virtualization), ARM TrustZone.
Purpose	Provide hardware-based security features to protect sensitive data and code from unauthorized access.

SIMD (Single Instruction, Multiple Data)

Definition	Allows a single instruction to operate on multiple data elements simultaneously, improving performance in multimedia and scientific applications. Examples: Intel SSE, AVX, ARM NEON.
Benefits	Faster multimedia processing, improved scientific computations, and enhanced gaming performance.

Out-of-Order Execution

Definition	A technique where the CPU executes instructions in a different order than they appear in the program, optimizing performance by avoiding stalls due to data dependencies.
How it Works	The CPU dynamically analyzes instructions and executes them in the order that maximizes resource utilization, improving overall performance.

Branch Prediction

Definition	A technique used to predict the outcome of conditional branch instructions (e.g., if-then-else statements) to avoid pipeline stalls.
Importance	Accurate branch prediction reduces the number of pipeline stalls, improving overall CPU performance.