Labang, Hudhaifah A.
BSCS 2B

What are L1, L2, and L3 caches and how are they organized?

Cache memory is a small, high-speed memory unit located close to the CPU. It's divided into **L1**, **L2**, and **L3** levels based on speed, size, and proximity to the processor.

Cache Level	Location	Size	Speed	Purpose
L1	Closest to CPU core	16KB - 128KB	Fastest	Stores most frequently used data/instructions
L2	Shared or per core	256KB - 2MB	Fast	Acts as a backup to L1
L3	Shared across cores	4MB - 64MB	Slower than L2	Reduces main memory access

Level 1 (L1) cache is often divided into an Instruction Cache (I-Cache) and a Data Cache (D-Cache).
Level 2 (L2) cache can either include the contents of L1 (inclusive) or store separate data (exclusive).
Level 3 (L3) cache is typically shared among multiple cores, especially in multi-core processors.

Why is cache memory faster than main memory?

Cache memory is faster than main memory because it is physically closer to the CPU and uses high-speed SRAM instead of the slower DRAM found in RAM. Its smaller size allows for quicker data access, reducing latency to just a few nanoseconds.

How does cache memory affect overall system performance? What strategies (like direct, associative, or set-associative mapping) are used in real CPUs (e.g., Intel, AMD, ARM)

Cache memory significantly boosts system performance by reducing the time the CPU spends waiting for data from the main memory. In gaming consoles, this leads to smoother gameplay, faster loading screens, and more responsive in-game actions.

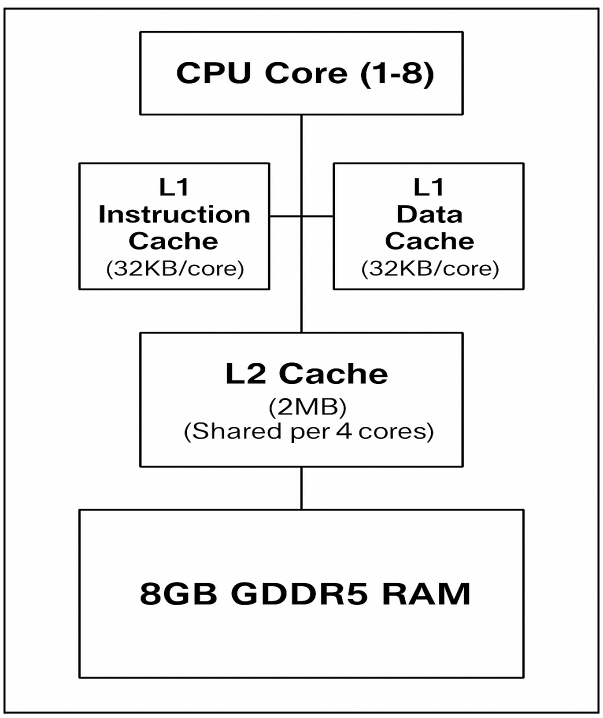
What are the trade-offs between cache size, speed, and cost?

Larger cache sizes improve performance by increasing data hit rates but are more expensive and slightly slower. Faster cache memory offers better speed but comes with higher costs and reduced capacity, making it essential to balance all three factors in console design.

Focus Area (PlayStation 4)

- **Summarize your findings**
  Cache memory plays a vital role in the performance of gaming consoles like the PlayStation 4 by acting as a high-speed data buffer between the CPU and main memory. The PS4 uses a multi-level cache system (L1 and L2) to reduce latency and improve game responsiveness. While it lacks an L3 cache, its architecture and fast GDDR5 RAM help maintain efficiency during high-demand gameplay.

- **Include diagrams or charts to support your explanation**



  Cache memory significantly helped the PS4 by reducing latency, allowing quick access to instructions and data, which minimized game stutter. It also improved the efficiency of AI and physics computations, as rapid instruction fetching boosted game logic and rendering performance. As a result, the PS4 was able to deliver smooth and consistent frame rates of 30 to 60 FPS in many AAA titles, despite having older hardware.

- **Highlight one real-world example where cache memory design made a big difference in performance**
  Horizon Forbidden West is a massive, visually stunning game, and it still runs surprisingly well on the PS4. That's thanks in part to how the console makes smart use of its L1 and L2 caches for real-time rendering. Even with huge open-world environments and complex enemy AI, the PS4 handles it smoothly by keeping memory access efficient. And while it doesn't have an L3 cache, the system manages a solid balance between its CPU caches and high-speed GDDR5 RAM, avoiding major slowdowns. It's a great example of how clever hardware design, especially with cache memory, where it can push older consoles to deliver an amazing gaming experience.