* **Recherche**

Es gibt unterschiedliche Arten von Caches:

* CPU (Prozessor) Cache:

-L1, L2, L3

* (Grafikprozessor) Cache
* Speicher-Cache (DRAM-Cache)
* Festplattencache
* Diese Caches benötigen natürlich unterschiedliche Größen, da Sie für jeweils andere Aufgaben gedacht sind bzw. Implementiert werde

Es ist ziemlich selten der Fall, dass man nur 4-fach assoziative Caches verwendet sondrn es variiert ziemlich stark von hersteller und dem eigentlichen Cache ab!

Allgemein gilt:

* Direct-mapped Caches werden seltener verwendet als assoziative Caches, hier sind Beispiele bei denen Sie trotzdem verwendet werden:
* Eingebettete Systeme (Embedded Systems):

In vielen eingebetteten Systemen, wie Mikrocontrollern und einfachen SoCs (System-on-Chip), werden häufig direct-mapped Caches verwendet. Diese Systeme haben oft beschränkte Ressourcen und benötigen einfache und kostengünstige Cache-Architekturen.

* Spezialprozessoren und einfache CPUs
* Spezialisierte Hardware-Komponenten
* Einfachere Cache-Implementationen in Forschung und Lehre

Fazit:

-Die Vorteile von direct-mapped Caches liegen in ihrer Einfachheit und Geschwindigkeit, da jede Speicheradresse genau einem Cache-Block zugeordnet wird. Diese Vorteile werden jedoch durch die erhöhte Wahrscheinlichkeit von Konflikt-Misses ausgeglichen, was bedeutet, dass sie weniger effizient sein können als ihre set-associative oder fully-associative Gegenstücke in vielen realen Anwendungen.

-Zusammengefasst: Direct-mapped Caches werden hauptsächlich in einfachen, kostensensitiven und ressourcenbeschränkten Umgebungen verwendet, wo ihre Vorteile die Nachteile überwiegen

Beispiele:

Intel® Core™ i9 Prozessoren der 13. Generation -> hat einen Cache der Größe: Prozessor 36 MB , ist ein “Smart-Cache” -> Prozessor Cache Typ

1. L1 Data cache (DCU) - 48KB (P-core) - 32KB (E-Core)

2. L1 Instruction cache (IFU) - 32KB (P-Core) - 64KB (E-Core)

3. MLC - Mid Level Cache - 1.25MB (P-Core) - 2MB (shared by 4 E-Cores)2

**Größen sind alle Pro Kern**

-P Cores 1st level cache is divided into a data cache (DFU) and an instruction cache

(IFU). The processor 1st level cache size is 48KB for data and 32KB for instructions.

The 1st level cache is an 12-way associative cache.2

-E Cores 1st level cache is divided into a data cache (DFU) and an instruction cache2

(IFU). The processor 1st level cache size is 32KB for data and 64KB for instructions.

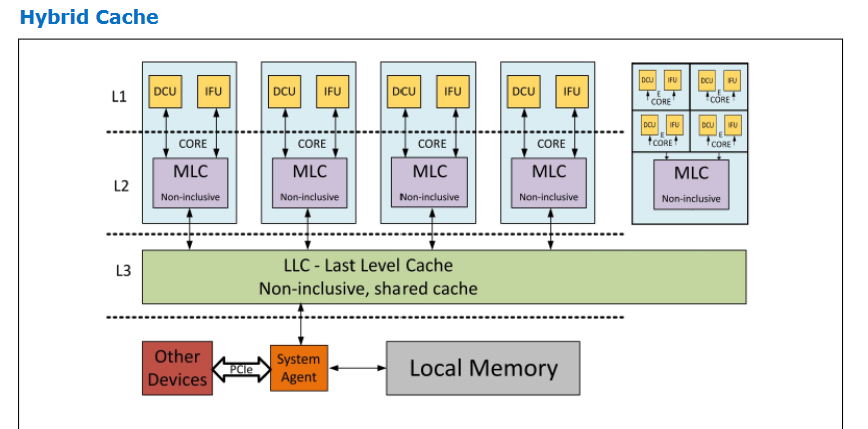
The 1st level cache is an 8-way associative cache.2

-The 2nd level cache holds both data and instructions. It is also referred to as mid-level

cache or MLC. The P Cores 2nd level cache size is 1.25MB and is a 10-way non-

inclusive associative cache., 4 E Cores processors share 2MB 2nd level cache and is a

16-way non-inclusive. associative cache2

(Aufbau eines Cache bei Intel nutzen auch das LLC-System)2

“L1 cache that can contain tens of thousands of bytes (32 KB is a typical size)“1

„Other private and shared caches are usually located on the path between the L1 cache and [main memory](https://en.wikipedia.org/wiki/Computer_data_storage) (although non-temporal loads and stores can bypass them). These caches range in size from the private 256-KB caches and many-MB shared caches in processors, to the many-GB caches stored in Multi-channel DRAM ([MCDRAM](https://en.wikipedia.org/wiki/MCDRAM)) and High-Bandwidth Memory ([HBM](https://en.wikipedia.org/wiki/High_Bandwidth_Memory)) memories or in dual inline-memory modules ([DIMMs](https://en.wikipedia.org/wiki/DIMM)).“ 1

* Direct-mapped Caches werden fast gar nicht verwendet → keine Üblichen Größen

Assuming you have a large processor (about 16 cores), the following summarizes, for 2016, approximate data totals present in and moving through the system:

| Memory | Size | Latency | Bandwidth |
| --- | --- | --- | --- |
| L1 cache | 32 KB | 1 nanosecond | 1 TB/second |
| L2 cache | 256 KB | 4 nanoseconds | 1 TB/second Sometimes shared by two cores |
| L3 cache | 8 MB or more | 10x slower than L2 | >400 GB/second |
| MCDRAM |  | 2x slower than L3 | 400 GB/second |
| Main memory on DDR DIMMs | 4 GB-1 TB | Similar to MCDRAM | 100 GB/second |
| Main memory on Cornelis\* Omni-Path Fabric | Limited only by cost | Depends on distance | Depends on distance and hardware |
| I/O devices on memory bus | 6 TB | 100x-1000x slower than memory | 25 GB/second |
| I/O devices on PCIe bus | Limited only by cost | From less than milliseconds to minutes | GB-TB/hour Depends on distance and hardware |

2

→ Typische Latenzen mit Größe für Caches laut Intel

Cache-Line Größe:

-bei x86 Architekturen sind es normalerweise 64 Bytes3 (Intel, AMD)

-Die Cache-Line-Größe variiert, aber viele moderne ARM-Prozessoren verwenden eine Cache-Line-Größe von 64 Bytes (Apple Prozessoren wie der M1, M2 oder M3)4

Quellen: <https://www.intel.com/content/www/us/en/developer/articles/technical/memory-performance-in-a-nutshell.html1>

<https://edc.intel.com/content/www/us/en/design/ipla/software-development-platforms/client/platforms/alder-lake-desktop/12th-generation-intel-core-processors-datasheet-volume-1-of-2/001/intel-smart-cache-technology/2>

<https://www.intel.com/content/www/us/en/developer/articles/technical/intel-sdm.html3>

<https://developer.arm.com/documentation4>