

Arsitektur Komputer

PRODI TEKNOLOGI REKAYASA INTERNET

Cache Memory

Tujuan Pembelajaran

Setelah mempelajari bab ini, Anda akan dapat:

Memberikan gambaran umum tentang fitur utama sistem memori computer dan penggunaan hierarki memori.

Mampu menjelaskan konsep dasar dan maksud dari memori cache.

Mampu menjelaskan elemen utama dari desain cache.

Mampu membedakan antara pemetaan langsung, pemetaan asosiatif, dan himpunan asosiatif pemetaan.

Mampu menjelaskan alasan untuk menggunakan beberapa level cache.

Mampu memahami implikasi kinerja dari berbagai tingkat memori.

Cache Memory

Location

Internal (e.g., processor registers, cache, main memory)

External (e.g., optical disks, magnetic disks, tapes)

Capacity

Number of words

Number of bytes

Unit of Transfer

Word

Block

Access Method

Sequential

Direct

Random

Associative

Performance

Access time

Cycle time

Transfer rate

Physical Type

Semiconductor

Magnetic

Optical

Magneto-optical

Physical Characteristics

Volatile/nonvolatile

Erasable/nonerasable

Organization

Memory modules

Internal Memori

RAM

- Statis/SRAM (Semiconductor based)
- Dinamis/DRAM (capasitor based)

ROM

- EPROM,
- EEPROM dan
- Memori flash

Access Methode

Word : unit "alami" dari organisasi memori. Panjang sebuah word biasanya sama dengan jumlah bit yang digunakan untuk mewakili bilangan bulat dan instruksi panjang. Sayangnya, ada banyak pengecualian. Misalnya, CRAY C90 (superkomputer CRAY model lama) memiliki panjang word 64-bit, tetapi menggunakan 46-bit seluruh representasi. Arsitektur Intel x86 memiliki berbagai macam instruksi panjang, dinyatakan sebagai kelipatan byte dan ukuran word 32-bit

Addressable Units : Pada beberapa sistem, unit yang dapat dialamatkan adalah word. Namun, banyak sistem memungkinkan pengalamatan tingkat byte. Bagaimanapun, hubungan antara panjang dalam bit A dari sebuah alamat dan jumlah N dari unit yang dapat dialamatkan adalah $2^A = N$.

Transfer of Units : Untuk memori utama, ini adalah jumlah bit yang dibaca dari atau ditulis ke memori pada suatu waktu. Satuan transfer tidak harus sama dengan word atau unit yang dapat dialamatkan. Untuk memori eksternal, data sering ditransfer dalam banyak unit yang lebih besar dari satu word dan disebut blok.

Metode Akses (Sequential access)

Akses sekuensial: memori diatur ke dalam unit data, yang disebut register.

Akses harus dalam urutan linier tertentu. Informasi pengalamatan yang disimpan digunakan untuk memisahkan catatan dan membantu dalam proses pengambilan. Mekanisme baca dan tulis bersama digunakan dan harus dipindahkan dari lokasi saat ini ke lokasi yang diinginkan, melewati dan menolak setiap record perantara. Oleh karena itu, waktu untuk mengakses catatan arbitrer sangat bervariasi.

Metode Akses (Direct access)

Seperti halnya akses sekuensial, akses langsung melibatkan pembagian mekanisme baca-tulis. Namun, blok atau catatan individu memiliki satu alamat berdasarkan lokasi fisik. Akses adalah melalui akses langsung untuk sampai ke lingkungan umum, selain pencarian berurutan, hitung atau tunggu sampai ke lokasi akhir. Sekali lagi, waktu akses bervariasi.

Metode Akses (Random Access)

Setiap lokasi yang dapat dialamatkan dalam memori memiliki keunikan, secara fisik mekanisme pengalamatan yang terhubung. Waktu akses ke lokasi tertentu tidak tergantung pada urutan akses sebelumnya dan konstan. Jadi lokasi mana pun dapat dipilih secara acak dan langsung ditangani dan diakses. memori utama dan beberapa sistem caching adalah akses acak

Metode Akses (Associative)

Asosiatif: Ini adalah jenis memori akses acak yang memungkinkan Anda untuk membandingkan lokasi bit yang diinginkan dalam sebuah word dengan kecocokan tertentu, dan melakukan ini untuk semua kata secara bersamaan. Jadi, sebuah kata diambil berdasarkan bagian dari isinya, bukan alamatnya. Seperti memori akses acak umum, setiap lokasi memiliki mekanisme pengalamatan sendiri dan waktu pengambilan konstan, terlepas dari lokasi atau pola akses sebelumnya. Memori cache dapat menggunakan akses asosiatif

Performance (Access time (latency))

Waktu akses (latensi): Untuk memori akses acak, ini adalah waktu yang diperlukan untuk melakukan operasi baca atau tulis, yaitu waktu dari saat alamat disajikan ke memori hingga saat data disimpan atau dibuat. tersedia untuk digunakan. Untuk memori akses non-acak, waktu akses adalah waktu yang diperlukan untuk memposisikan mekanisme baca dan tulis di lokasi yang diinginkan.

Performance (Memory cycle time)

Waktu Siklus Memori: Konsep ini terutama diterapkan pada memori akses acak dan terdiri dari waktu akses ditambah waktu tambahan yang diperlukan sebelum akses kedua dapat dimulai. Waktu tambahan ini mungkin diperlukan agar transien mati pada jalur sinyal atau untuk meregenerasi data jika dibaca secara destruktif. Perhatikan bahwa waktu siklus memori terkait dengan bus sistem, bukan prosesor.

Transfer rate

Kecepatan transfer: Ini adalah kecepatan di mana data dapat ditransfer ke dalam atau keluar dari unit memori. Untuk memori akses acak, sama dengan $1 / (\text{waktu siklus})$. Untuk memori akses non-acak, hubungan berikut berlaku:

$$T_n = T_A + \frac{n}{R}$$

where

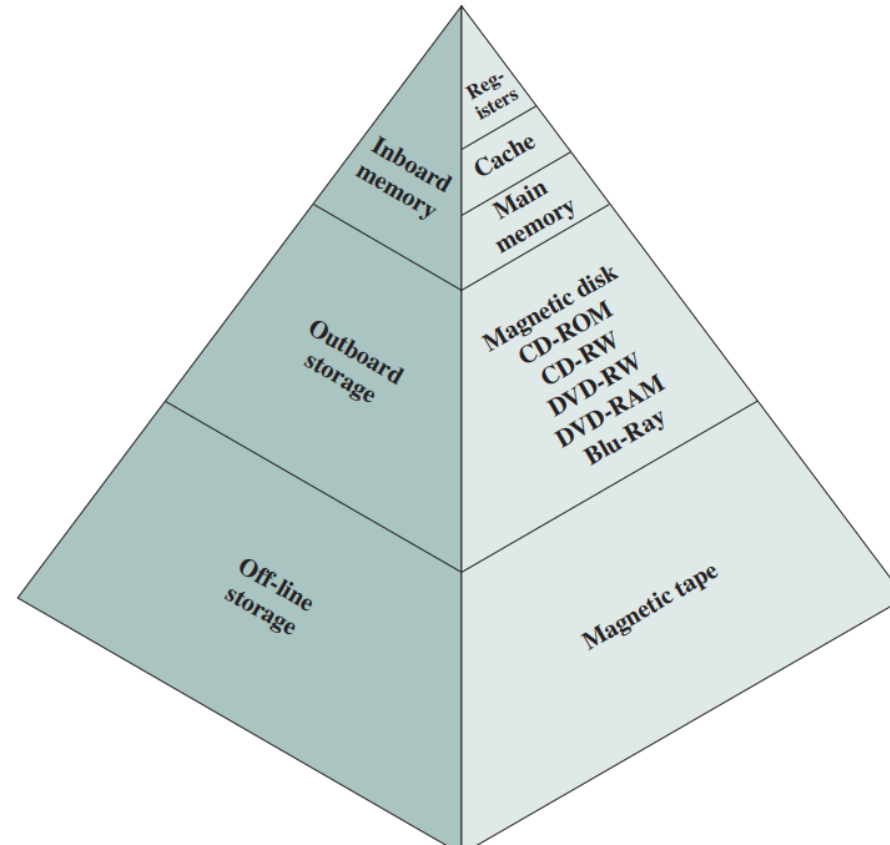
T_n = Average time to read or write n bits

T_A = Average access time

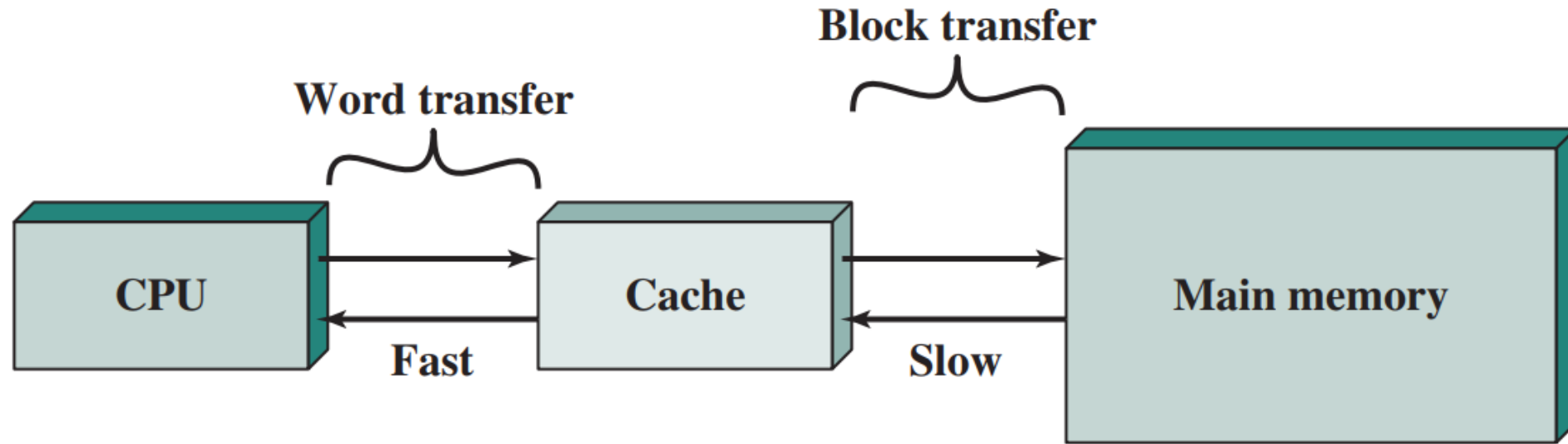
n = Number of bits

R = Transfer rate, in bits per second (bps)

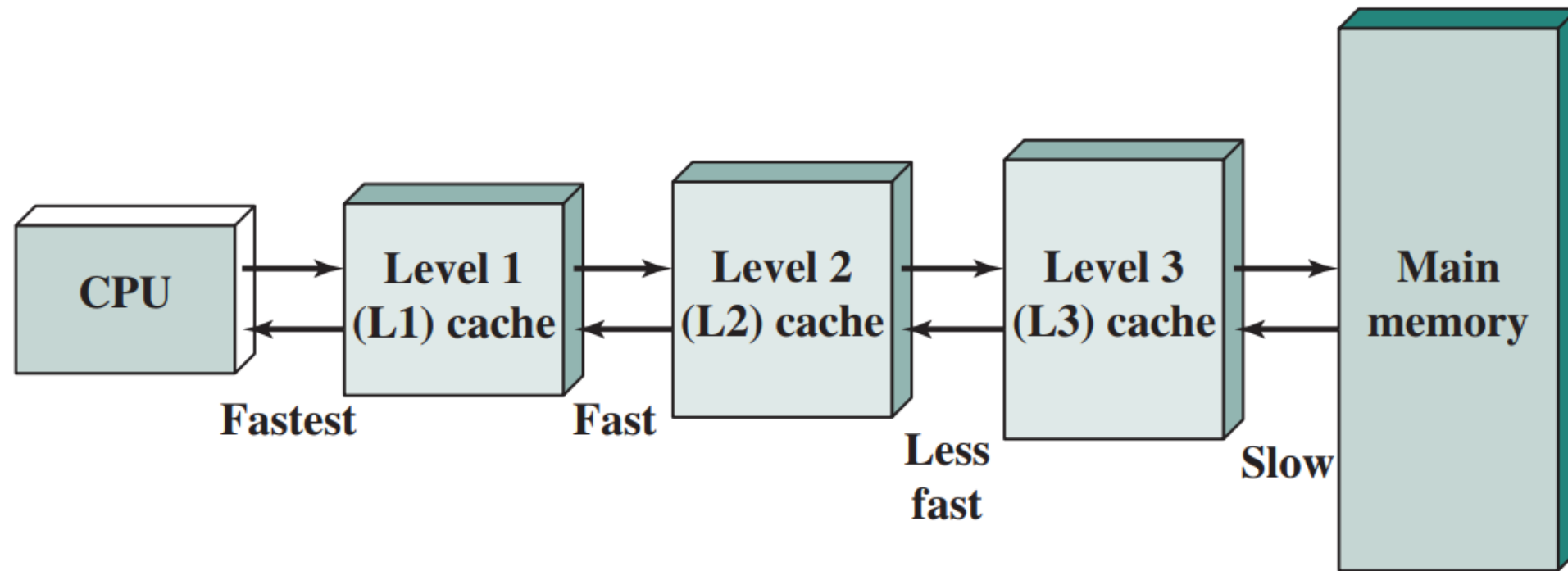
The Memory Hierarchy



Cache memory principles

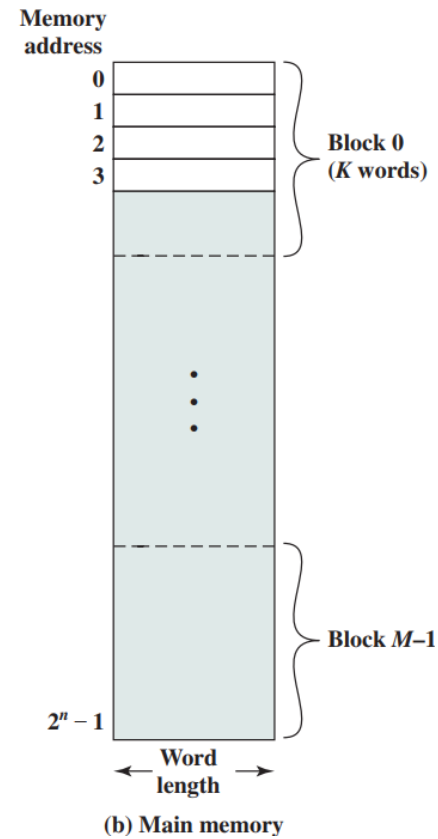
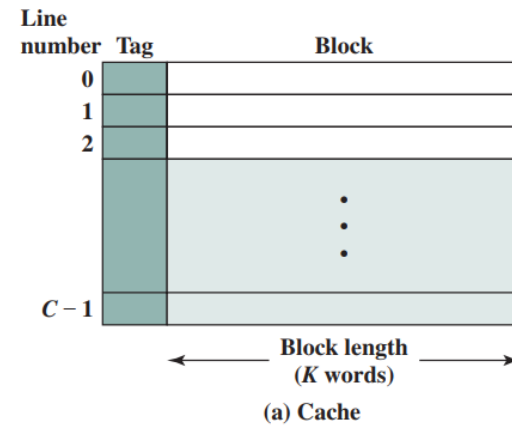


(a) Single cache

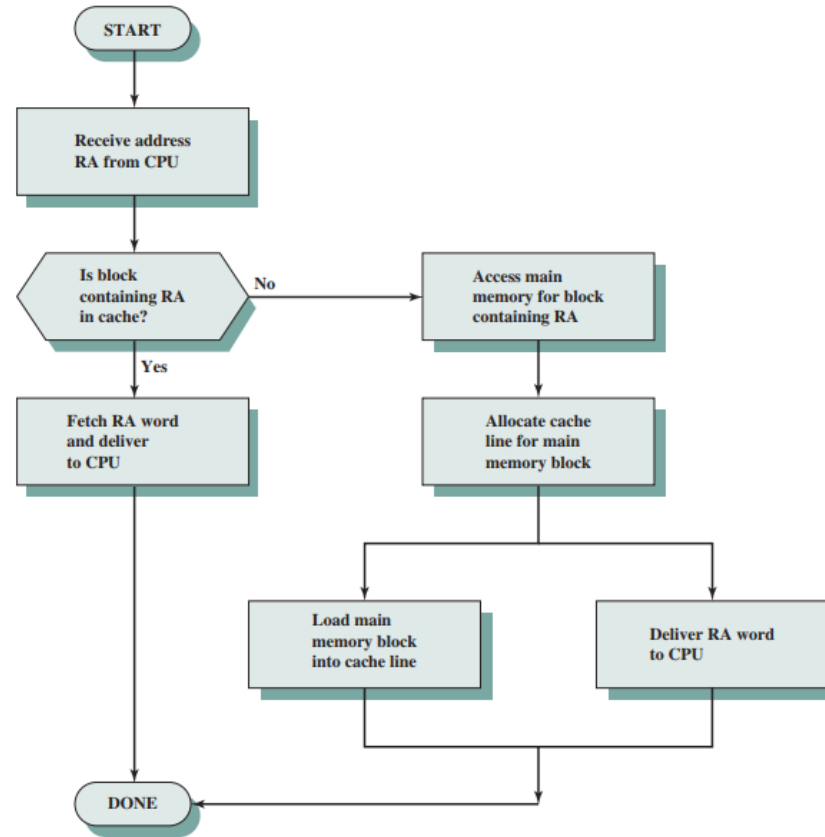


(b) Three-level cache organization

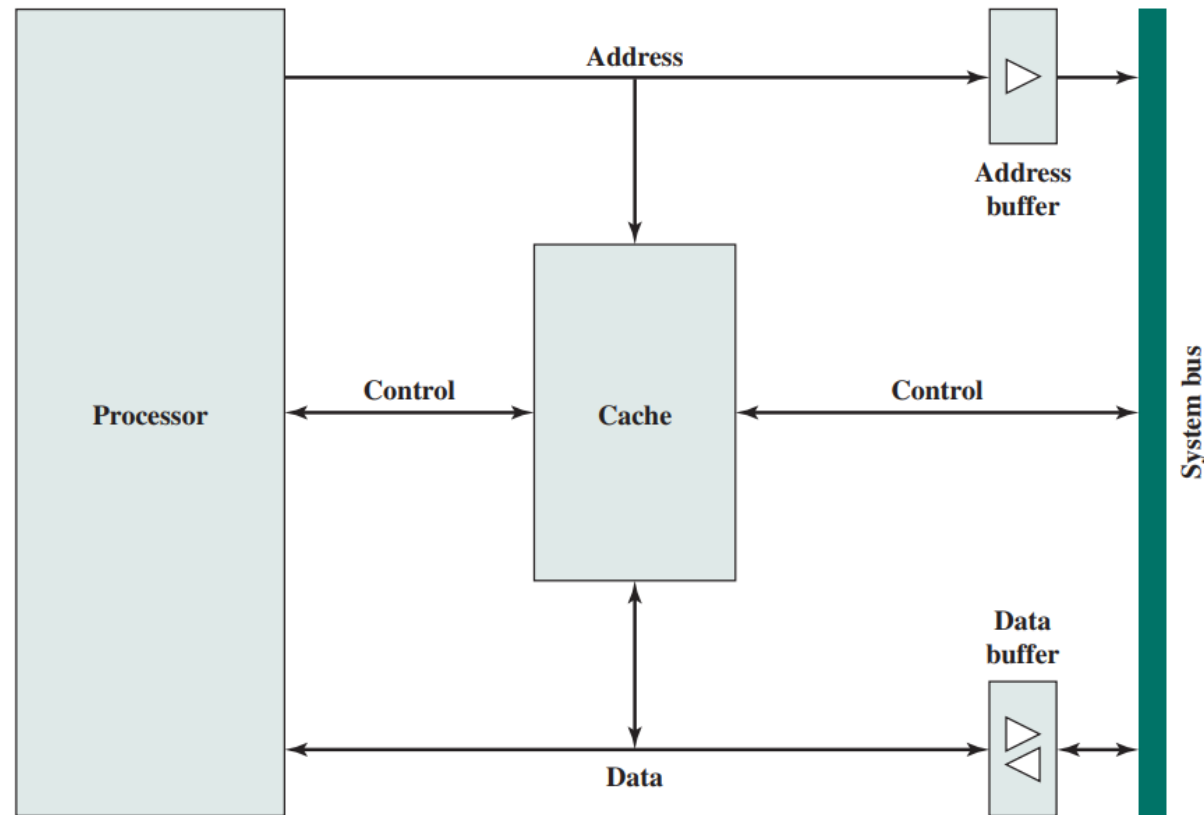
Cache/Main Memory Structure



Cache Read Operation



Typical Cache Organization



External Memory

Magnetic Disk

RAID (*Redundant Array of Independent Disks*)

Solid State Drives

Optical Memory

Magnetic Tape