2012-07-05

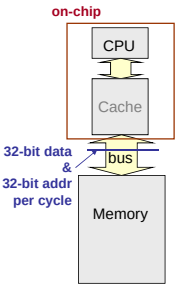
計算機アーキテクチャ 第一 (E)

半導体メモリシステムの補足

吉瀬 謙二 計算工学専攻
kise_at_cs.titech.ac.jp
W641講義室　木曜日13:20 ー 14:50

Memory Systems that Support Caches

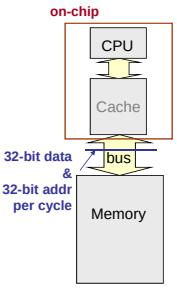
- The off-chip interconnect and memory architecture can affect overall system performance **in dramatic ways**



One word wide organization (one word wide bus and one word wide memory)

- **Assume**
  1. **1** clock cycle to send the address
  2. **25** clock cycles for DRAM **cycle time**, **8** clock cycles **access time**
  3. **1** clock cycle to return a word of data
- Memory-Bus to Cache **bandwidth**
  - **number of bytes transferred from memory to cache per clock cycle**

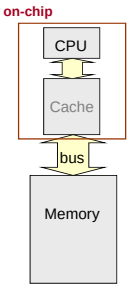
One Word Wide Memory Organization



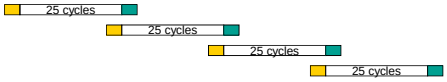
- The pipeline stalls the number of cycles for **one word** (32bit) from memory
  - **1** cycle to send address
  - **25** cycles to read DRAM
  - **1** cycle to return data
  - **27** **total clock cycles** miss penalty

- Number of bytes transferred per clock cycle (**bandwidth**) for a single miss
  - **4 / 27 = 0.148** bytes per clock

One Word Wide Memory Organization, con't



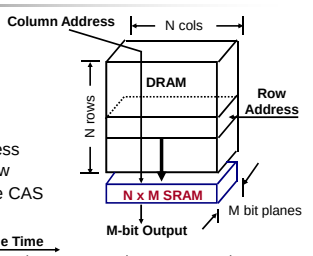
- What if the block size is **four words**?
  - **1** cycle to send 1st address
  - **4 * 25 = 100** cycles to read DRAM
  - **1** cycle to return last data word
  - **102** **total clock cycles** miss penalty

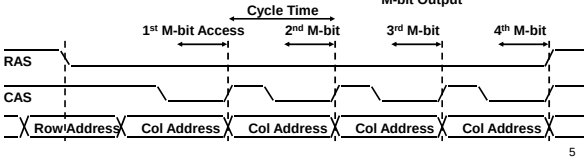


- Number of bytes transferred per clock cycle (**bandwidth**) for a single miss
  - **(4 x 4) / 102 = 0.157** bytes per clock

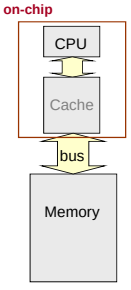
Page Mode DRAM Operation

- Page Mode DRAM
  - N x M SRAM to save a row
- After a row is read into the SRAM "register"
  - Only CAS is needed to access other M-bit words on that row
  - RAS remains asserted while CAS is toggled

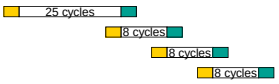




One Word Wide Memory Organization, con't

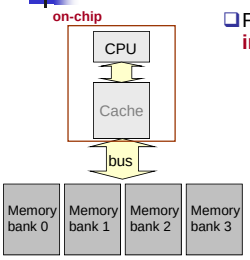


- What if the block size is **four words** and if a **page mode DRAM** is used?
  - **1** cycle to send 1st address
  - **25 + (3 * 8) = 49** cycles to read DRAM
  - **1** cycle to return last data word
  - **51** **total clock cycles** miss penalty

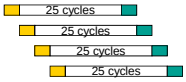


- Number of bytes transferred per clock cycle (**bandwidth**) for a single miss
  - **(4 x 4) / 51 = 0.314** bytes per clock

Interleaved(インターリーブ) Memory Organization



- For a block size of **four words** with **interleaved memory (4 banks)**
  - **1** cycle to send 1st address
  - **25 + 3 = 28** cycles to read DRAM
  - **1** cycle to return last data word
  - **30 total clock cycles** miss penalty



- Number of bytes transferred per clock cycle (**bandwidth**) for a single miss
  - **(4 x 4) / 30 = 0.533** bytes per clock

7

Exercise

- 1ワード(語)を4バイト，キャッシュとメモリ間の転送単位を示すブロック・サイズを2ワードとする．
  容量4Kバイトの**ダイレクト・マップ方式**と**2ウェイ・セットアソシアティブ方式**のキャッシュのハードウェア構成を図示せよ．エントリ数および，インデックスの生成方法(参照アドレスのどのビットを用いるか)を明らかにすること．

氏名，学籍番号，
学籍番号マーク欄

8

2012-07-05

2012年 前学期 TOKYO TECH

計算機アーキテクチャ 第一 (E)

磁気ディスク，RAID

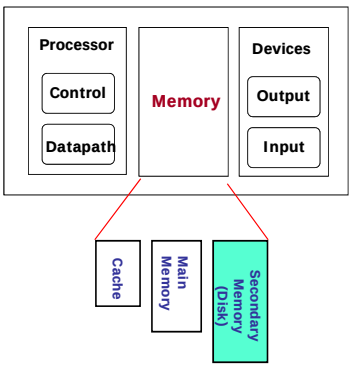
吉瀬 謙二 計算工学専攻
kise_at_cs.titech.ac.jp
W641講義室 木曜日13:20 ー 14:50

Acknowledgement

- Lecture slides for Computer Organization and Design, Third Edition, courtesy of **Professor Mary Jane Irwin**, Penn State University
- Lecture slides for Computer Organization and Design, third edition, Chapters 1-9, courtesy of **Professor Tod Amon,** Southern Utah University.

10

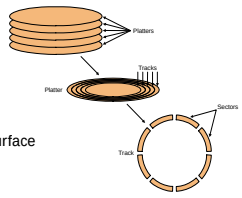
Major Components of a Computer



11

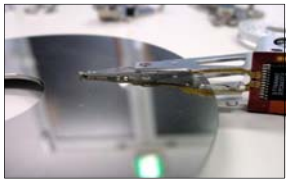
Magnetic Disk（磁気ディスク）

- Purpose
  - Long term, **nonvolatile**(不揮発性) storage
  - Lowest level in the memory hierarchy
    - slow, large, inexpensive
- General structure
  - A rotating **platter** coated with a magnetic surface
  - A moveable read/write **head** to access the information on the disk
- Typical numbers
  - 1 to 4 platters per disk of 1" to 5.25" in diameter (3.5" dominate in 2004)
  - Rotational speeds of 5,400 to 15,000 RPM (rotation per minute)
  - 10,000 to 50,000 **tracks** per surface
    - **cylinder** - all the tracks under the head at a given point on all surfaces
  - 100 to 500 **sectors** per track
    - the smallest unit that can be read/written (typically **512B**)



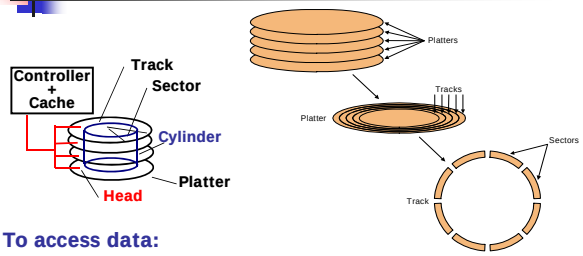
12

Magnetic Disk（磁気ディスク）



http://sougo057.aicomp.jp/0001.html

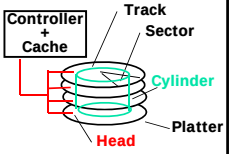
Disk Drives



To access data:

- **seek time（シーク時間）**: position the head over the proper track
- **rotational latency（回転待ち時間）**: wait for desired sector
- **transfer time（転送時間）**: grab the data (one or more sectors)
- **Controller time（制御時間）**: the overhead the disk controller imposes in performing a disk I/O access

Magnetic Disk Characteristic



- **Disk read/write components**
  1. **Seek time**: position the head over the proper track (**3 to 14 ms avg**)
     - due to locality of disk references the actual average seek time may be only 25% to 33% of the advertised number
  2. **Rotational latency**: wait for the desired sector to rotate under the head (½ of 1/RPM converted to ms)
     - 0.5/5400RPM = 0.5/90 rotations per second = **5.6 ms**
     - 0.5/15000RPM = 0.5/250 rotations per second = **2.0 ms**
  3. **Transfer time**: transfer a block of bits (one or more sectors) under the head to the disk controller's cache (**30 to 80 MB/s** are typical disk transfer rates)
  4. **Controller time**: the overhead the disk controller imposes in performing a disk I/O access (**typically < .2 ms**)

Typical Disk Access Time

- The average time to read or write a 512B sector for a disk rotating at 10,000RPM with average seek time of 6ms, a 50MB/sec transfer rate, and a 0.2ms controller overhead

Typical Disk Access Time

- The average time to read or write a 512B sector for a disk rotating at 10,000RPM with average seek time of 6ms, a 50MB/sec transfer rate, and a 0.2ms controller overhead

  Avg disk read/write time
  = 6.0ms + 0.5/(10000RPM/(60sec/minute)) + 0.5KB/(50MB/sec) + 0.2ms
  = 6.0 + 3.0 + 0.01 + 0.2
  = 9.21ms

  If the measured average seek time is **25%** of the advertised average seek time, then

  Avg disk read/write = 1.5 + 3.0 + 0.01 + 0.2 = 4.71ms

- The **rotational latency** is usually the largest component of the access time

Disk Latency & Bandwidth Milestones

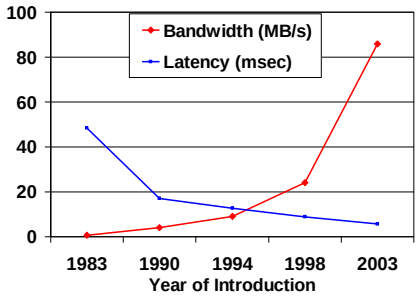
- Disk **latency** is one average seek time plus the rotational latency.
- Disk **bandwidth** is the peak transfer time of formatted data from the media (not from the cache).

	CDC Wren	SG ST41	SG ST15	SG ST39	SG ST37
Speed (RPM)	3600	5400	7200	10000	15000
Year	**1983**	**1990**	**1994**	**1998**	**2003**
Capacity (Gbytes)	0.03	1.4	4.3	9.1	73.4
Diameter (inches)	5.25	5.25	3.5	3.0	2.5
Interface	ST-412	SCSI	SCSI	SCSI	SCSI
Bandwidth (MB/s)	0.6	4	9	24	86
Latency (msec)	48.3	17.1	12.7	8.8	5.7

Patterson, CACM Vol 47, #10, 2004

Latency & Bandwidth Improvements

- In the time that the disk **bandwidth doubles** the **latency** improves by a factor of only **1.2** to **1.4**



19

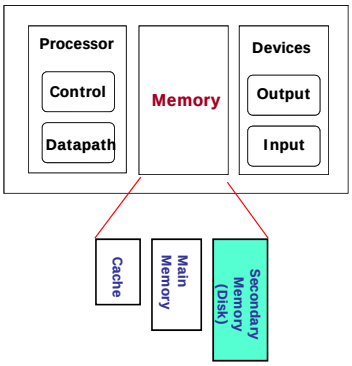
SSD (Solid State Drive)

20

SSD (Solid State Drive)

21

Major Components of a Computer



22

Reliability（信頼性）, Availability

- Reliability – measured by the mean time to failure (平均故障時間, MTTF).
- Service interruption is measured by mean time to repair (平均修復時間, MTTR)
- Availability（アベイラビリティ）

  Availability = MTTF / (MTTF + MTTR)

- To increase MTTF, either improve the quality of the components or design the system to continue operating in the presence of faulty components
  1. Fault avoidance: preventing fault occurrence by construction
  2. Fault tolerance: using redundancy to correct or bypass faulty components (hardware)

高信頼ディスクの典型的なMTTF は100万時間 (114年) 程度.

23

アナウンス

- 講義スライドおよびスケジュール
  - www.arch.cs.titech.ac.jp
  - 講義日程が変更になることがあるので頻繁に確認すること.

24