2010-05-20

計算機アーキテクチャ 第一 (E)

6. メモリ1：半導体メモリシステム

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.

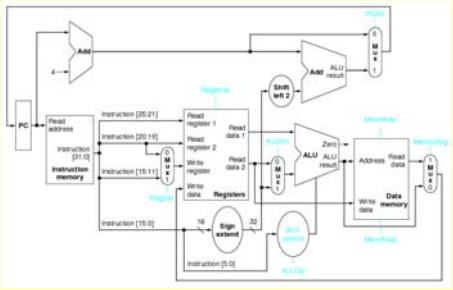
Adapted from *Computer Organization and Design*, Patterson & Hennessy, © 2005

プロセッサのデータパス(シングル・サイクル)

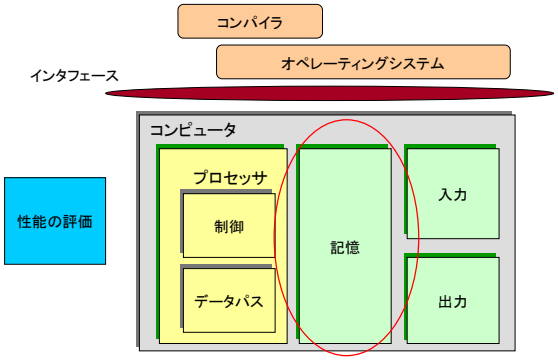
op	rs	rt	16 bit immediate

I format

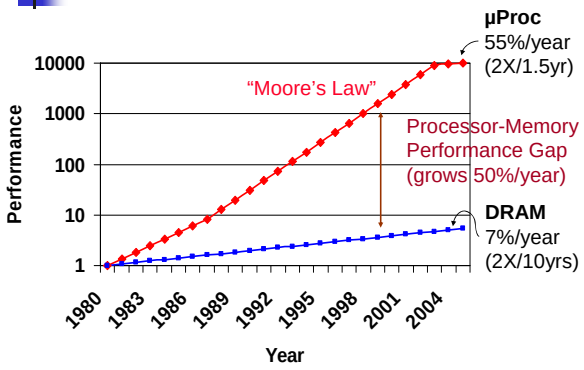
beq $s0, $s1, Label [beq $16, $17, Label]



コンピュータ(ハードウェア)の古典的な要素



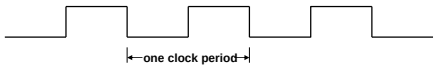
Processor-Memory Performance Gap



Machine Clock Rate

- **Clock rate** (MHz, GHz) is inverse of clock cycle time (**clock period**)

Clock period = 1 / (**clock rate**)



10 nsec clock cycle	=>	100 MHz clock rate
5 nsec clock cycle	=>	200 MHz clock rate
2 nsec clock cycle	=>	500 MHz clock rate
1 nsec clock cycle	=>	1 GHz clock rate
500 psec clock cycle	=>	2 GHz clock rate
250 psec clock cycle	=>	4 GHz clock rate
200 psec clock cycle	=>	5 GHz clock rate

Clock Cycles per Instruction, **CPI**

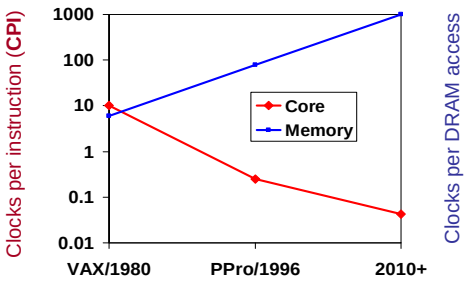
- Not all instructions take the same amount of time to execute

# CPU clock cycles for a program	=	# Instructions for a program	x	Average clock cycles per instruction

- **Clock cycles per instruction** (**CPI**) – the average number of clock cycles each instruction takes to execute
  - CPI = 10.0
  - CPI = 1.0
  - CPI = 0.5
  - CPI = 0.1

The "**Memory Wall**"

- **Arithmetic** vs **DRAM** speed gap continues to grow

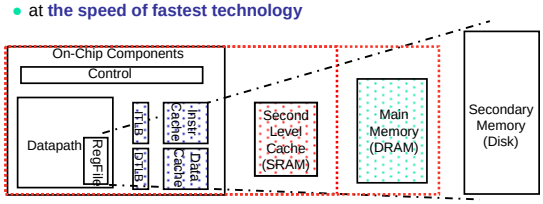


The Memory System **Goal**

- **Fact**:
  Large memories are slow and
  fast memories are small
- How do we create a memory that gives the illusion of being large, cheap and fast ?
  - With **hierarchy**（階層）
  - With **parallelism**（並列性）

A Typical Memory Hierarchy

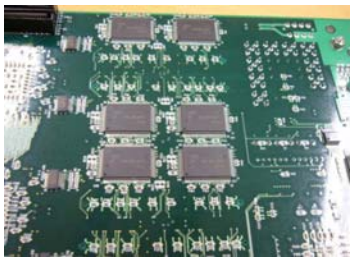
- By taking advantage of **the principle of locality**（局所性）
  - Present **much memory** in **the cheapest technology**
  - at **the speed of fastest technology**



Speed (%cycles):	½'s	1's	10's	100's	1,000's
Size (bytes):	100's	K's	10K's	M's	G's to T's
Cost:	highest				lowest

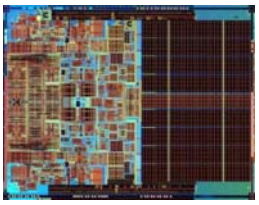
DRAM (dynamic random access memory)

SRAM (static random access memory)



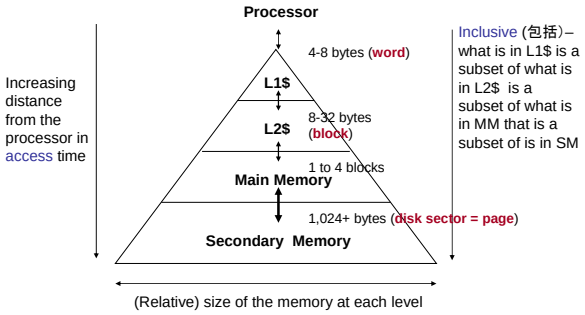
Cache

- *Cache memory* consists of a small, fast memory that acts as a buffer for the DRAM memory.
- The nontechnical definition of *cache* is a safe place for hiding things.



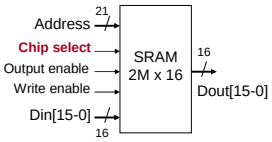
Intel Core 2 Duo

Characteristics of the Memory Hierarchy



Memory Hierarchy Technologies

- Caches use SRAM for speed and technology compatibility
  - **Low density** (**6 transistor cells**), high power, expensive, fast
  - **Static**: content will last "forever" (until power turned off)

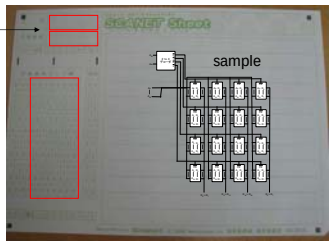


- Main Memory uses DRAM for size (density)
  - **High density** (**1 transistor cells**), low power, cheap, slow
  - **Dynamic**: needs to be "refreshed" regularly (~ every 8 ms)
    - 1% to 2% of the active cycles of the DRAM
  - Addresses divided into 2 halves (row and column)
    - **RAS** or Row Access Strobe triggering row decoder
    - **CAS** or Column Access Strobe triggering column selector

演習

- 512K x 8ビット（512KB）のSRAMを用いて、32ビットデータ幅の4MB のメモリを実現したい.
- 8個のメモリチップ，チップ選択信号CS， データ信号，アドレス信号の接続を示せ.

氏名，学籍番号，
学籍番号マーク欄(**右詰で**)



Memory Performance Metrics

- **Latency(レイテンシ，応答時間)**:
  Time to access one word
  - **Cycle time**: time between requests
  - **Access time**: time between the request and when the data is available (or written)
  - Usually **cycle time** > **access time**
- **Bandwidth(バンド幅，スループット)**:
  How much data from the memory can be supplied to the processor per unit time
  - width of the data channel * the rate at which it can be used

Classical RAM Organization (~Square)