Contents

1 Von Neumann Architecture

A simple abstraction of a general purpose digital computer

- All program instructions and data are stored in a **single memory unit**
- The CPU fetches instructions and executes them, in sequence, using the ALU
- The memory is **read-write** and random access, i.e. any location in memory can be accessed as easily/quickly **as any other**. Unlike, for example, tape storage

2 Memory hierarchy

2.1 Main Memory and Registers

Main memory is **separate** from CPU.

- Explicitly addressed by program instructions
- Contains all program code and almost all data
- Is **slow to access** compared to one ALU operation

Registers are built into CPU

- **Explicitly named** by program instructions
- Contain just a few local (local) variables / parameters
- **Very fast to access but very limited**

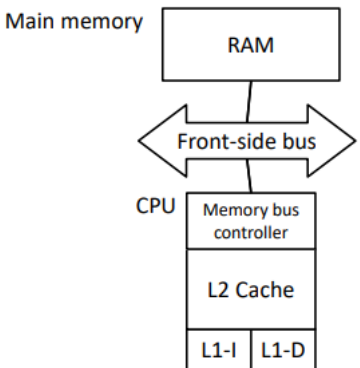
2.2 Cache

Cache is used to improve main memory performance

- Built into CPU (like registers)
- **Not directly visible** to program code (unlike registers)
- Temporarily stores code/data **fetched from main memory**
- Limited size: only stores recently requested, used or anticipated code/data
- Faster to access than main memory **once loaded**
  - Repeating code (e.g. loop) is faster than sequential code
  - Repeated data access is faster than one-off data access

2.2.1 Cache levels

- Level 1 Instruction (L1-I) cache: program code being executed (a few kB)
- Level 1 Data (L1-D) cache: data from main memory being used right now (a few kB)
- Level 2 Unified (Instruction and Data) cache (a few MB)
- (may be more levels)



Reference section

placeholder