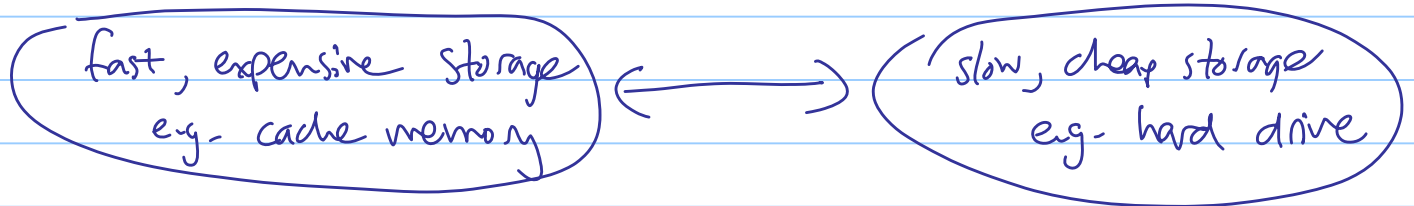


Memory hierarchy

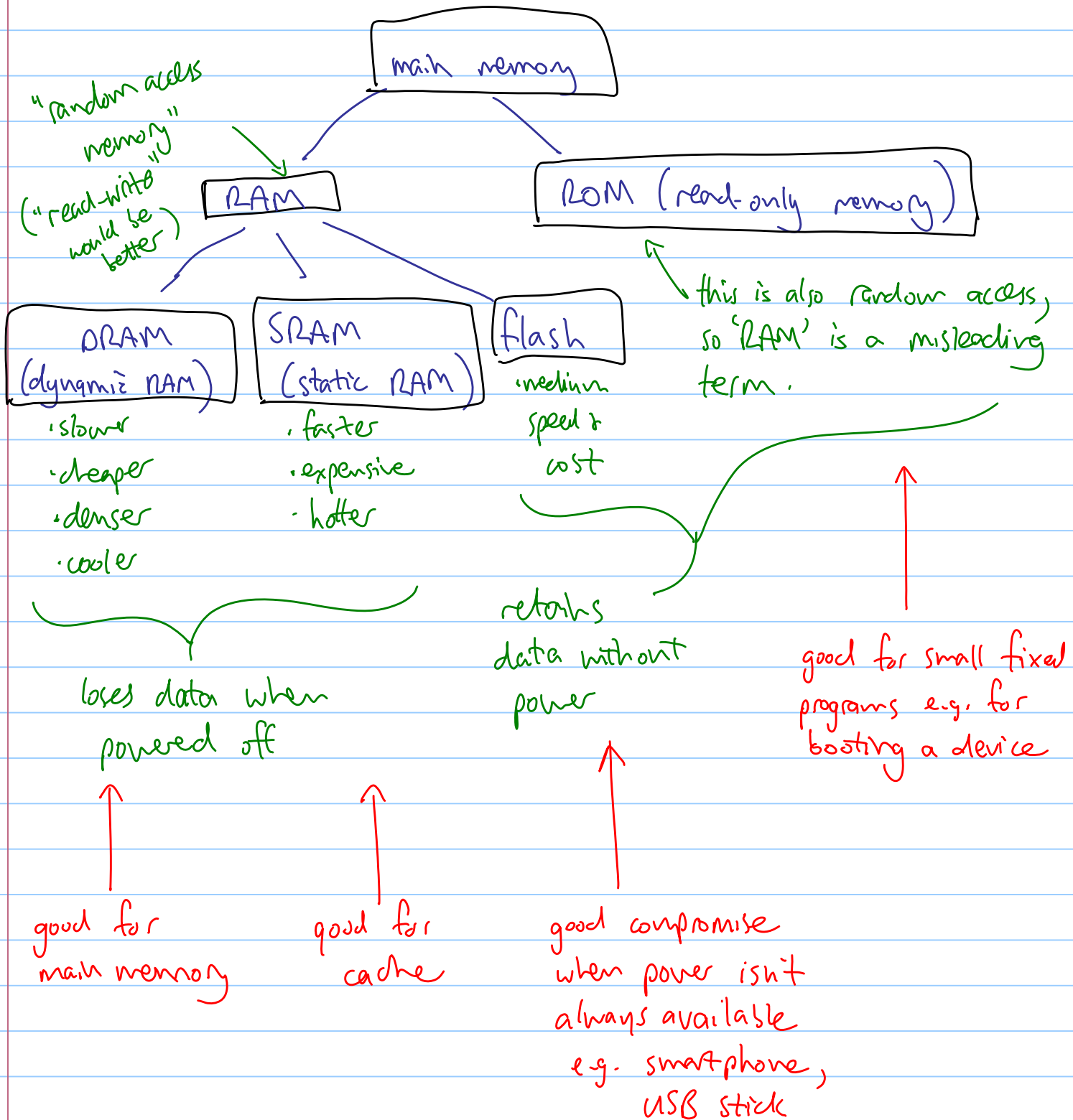
Note Title

[Typewritten portions of these notes are copied verbatim from the Powerpoint slides provided by the textbook authors]

All computers use a hierarchy of hardware for storing data, ranging from

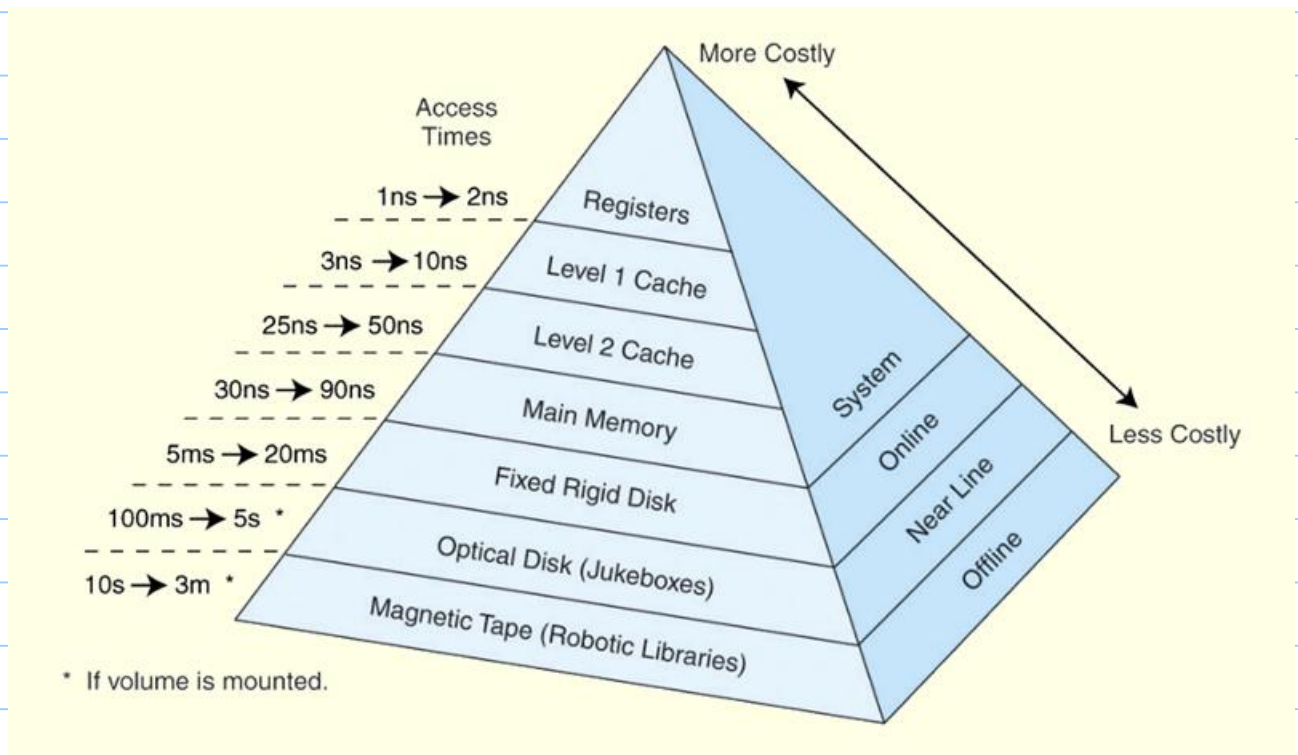


Types of main memory



Memory hierarchy

From textbook :



The hierarchy uses the concept of caching : store information that is likely to be needed soon in a place where it can be accessed quickly.

When searching for a given piece of data, start from the top and move down until it is found.

A simplified version of the above pyramid is:

| Storage type | typical access time | typical amount available |
|--------------|------------------------------------|--------------------------|
| register | 1 ns | 50 bytes |
| cache | 10 ns | 5 MB |
| main memory | 100 ns | 5 GB |
| hard drive | 10 000 ns for <u>random</u> access | 100 GB |

Definitions from textbook authors' slides:

- A *hit* is when data is found at a given memory level.
- A *miss* is when it is not found.
- The *hit rate* is the percentage of time data is found at a given memory level.
- The *miss rate* is the percentage of time it is not.
- Miss rate = $1 - \text{hit rate}$.
- The *hit time* is the time required to access data at a given memory level.
- The *miss penalty* is the time required to process a miss, including the time that it takes to replace a block of memory plus the time it takes to deliver the data to the processor.

Defn : Expected value

Recall that the expected value of an event with probabilities p_1, p_2 and outcomes v_1, v_2 is just $p_1 v_1 + p_2 v_2$

e.g. biased coin with $\text{prob}(\text{head}) = 0.7$
 $\text{prob}(\text{tail}) = 0.3$

play game where head wins \$10, tail wins \$2.
What is expected value of playing once?

Answer: exercise

Similar formula for many outcomes: $p_1 v_1 + p_2 v_2 + \dots + p_n v_n$.

Expected access time (EAT)

Above formula applies when calculating expected access time for a piece of data.

e.g. 2-level hierarchy with 70% hit rate.
Access times are 10ns and 100ns for the two levels. What is the EAT?

Answer: exercise

Mimlas : We look at results from CacheTimer.java and estimate size of cache on lab machines.

