Read Ch. 9

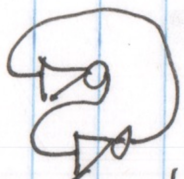
Kinds of memory

RAM - Random Access Memory

Read-write memory

Random = All accesses take the same time

SRAM - S = static

An array of latches  6 transitors per cell

Any delay element is a potential memory

Dynamic RAM

3 transistors per cell

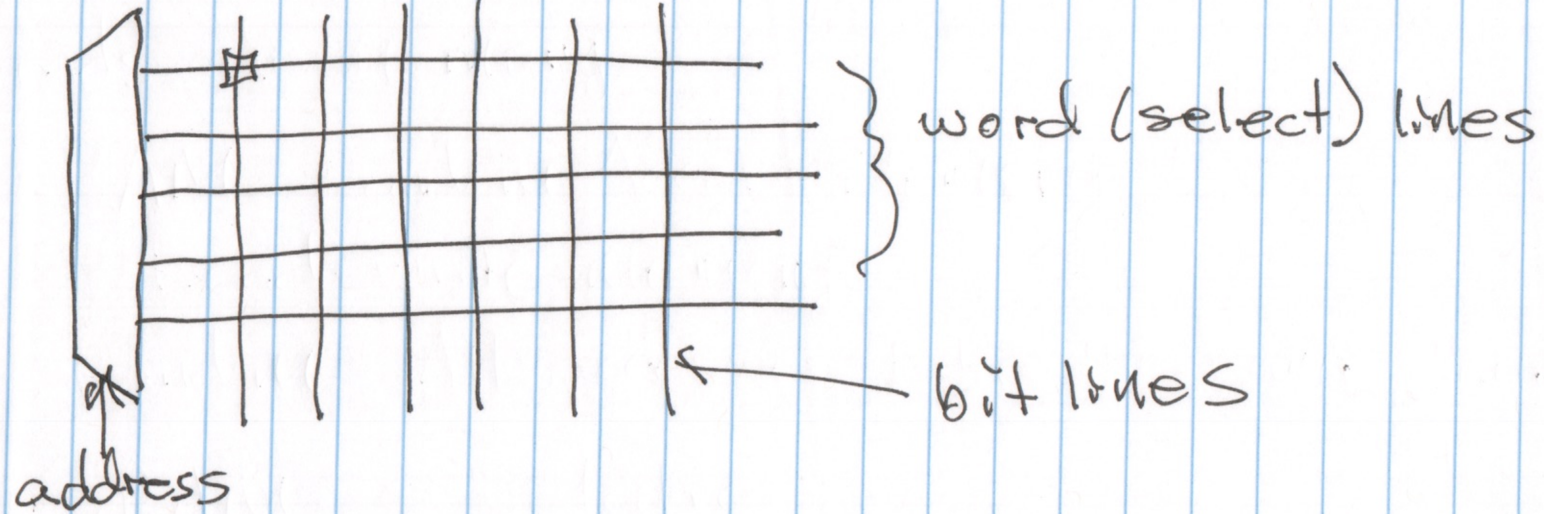
Forgets!

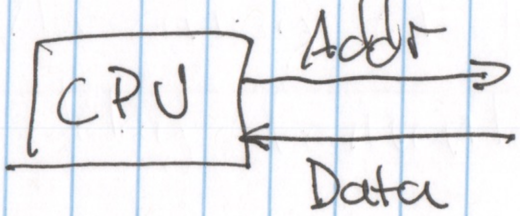
Periodically refresh it

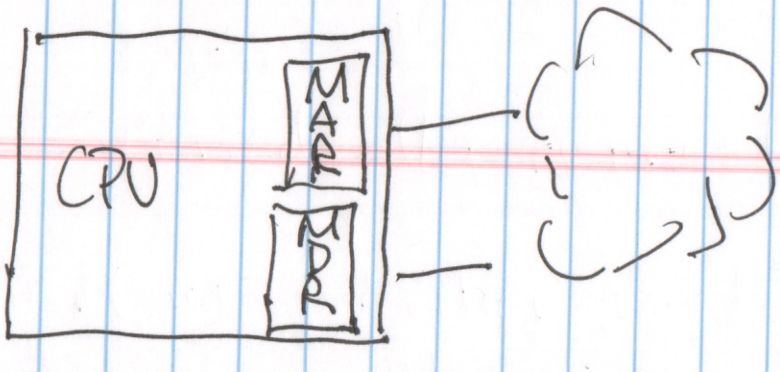
Cheaper than SRAM

Memory is either fast or cheap but never both!

Speed $\sim$ cost

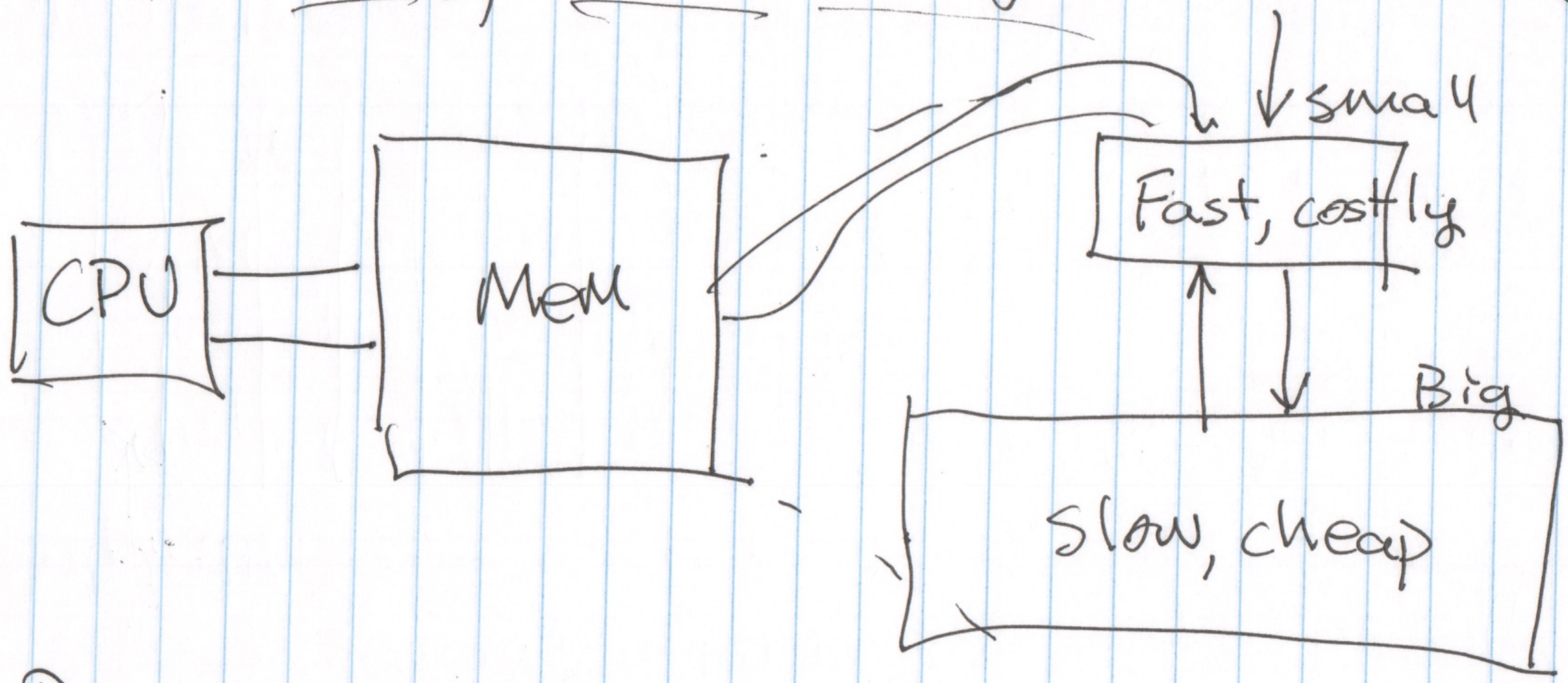






Goal cheap, fast memory!

Cache ~~Buffer~~ memory



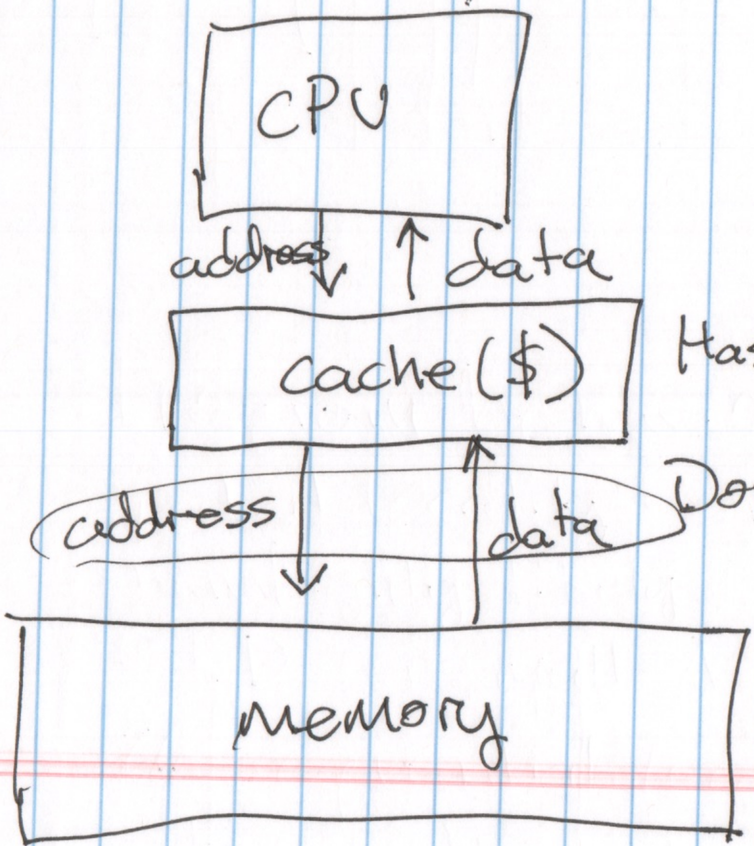
Program behavior

Spatial locality - if access X at time t, then highly likely to access X ± A at time t ± D for small A and D

special case ~~→~~ Sequential locality if access X at time t then highly likely to access X + B at time t+1 for constant B

Temporal locality

if access X at time t, then highly likely re-access X at time $t \pm D$ for small D



Miss repair

Has it! HIT

Doesn't have it
:(MISS

time for a miss repair is the MISS PENALTY

Effective Memory Access Time

Average Access Time

=

Time for a hit — cost in time of a hit

+

Miss rate × Miss penalty — average cost of a Miss (also in time)

$$\text{Miss rate} = \frac{\#\text{ of misses by a program}}{\text{Total }\#\text{ of program accesses}}$$

2^{10} kilo

2^{20} mega

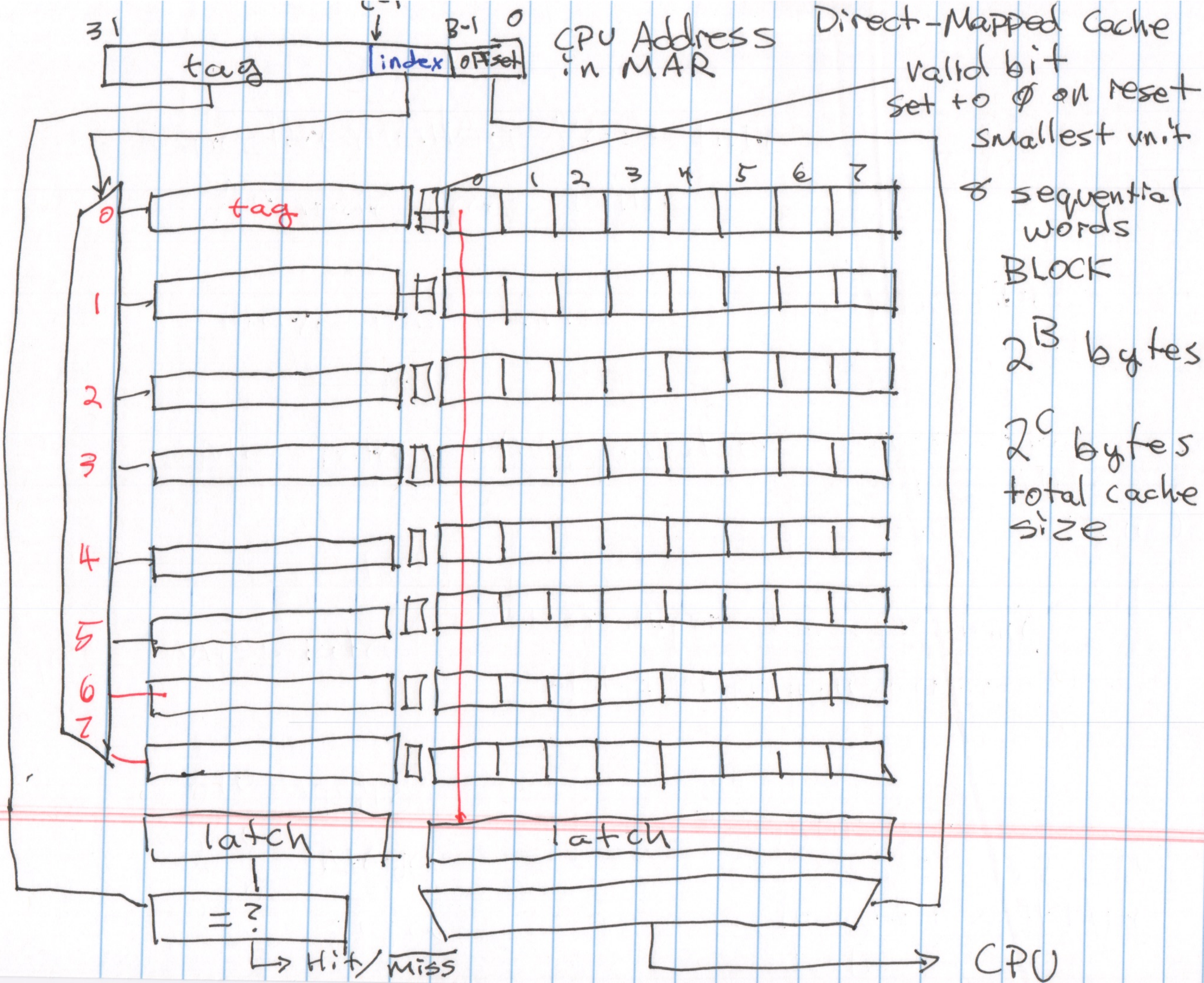
2^{30} giga

$2^{32} = 2^{30} \cdot 2^{2}$

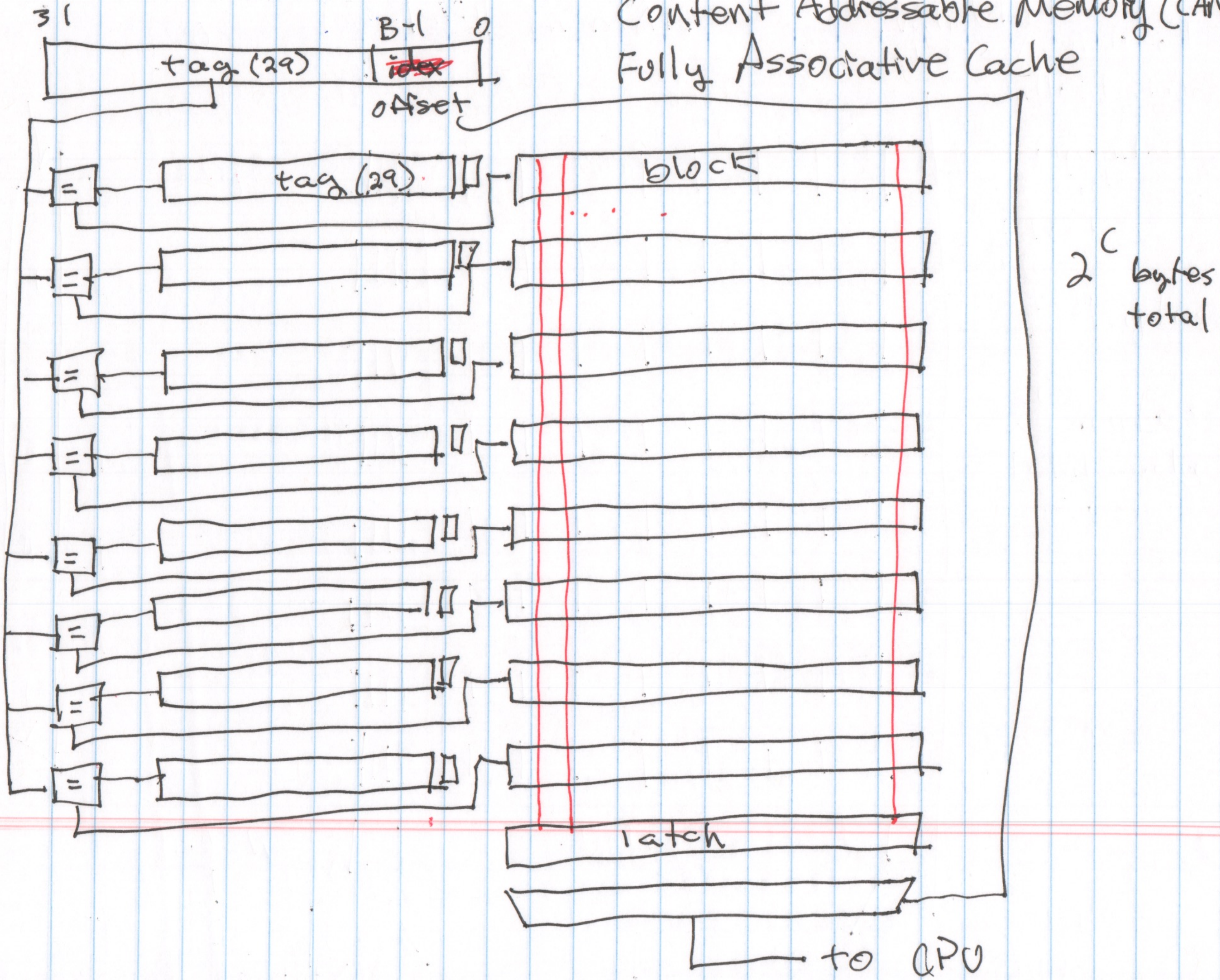
$= 1\text{ Gigabyte} \times 4\text{ bytes}$

$= 4\text{GB}$

Direct-Mapped Cache



Content Addressable Memory (CAM)
Fully Associative Cache



Direct Mapped

Any address has only one "slot" in cache

Bad

Fully Associative

Any address can use any of the

$2^C/2^B$ blocks in the cache

Bad idea because - more hardware

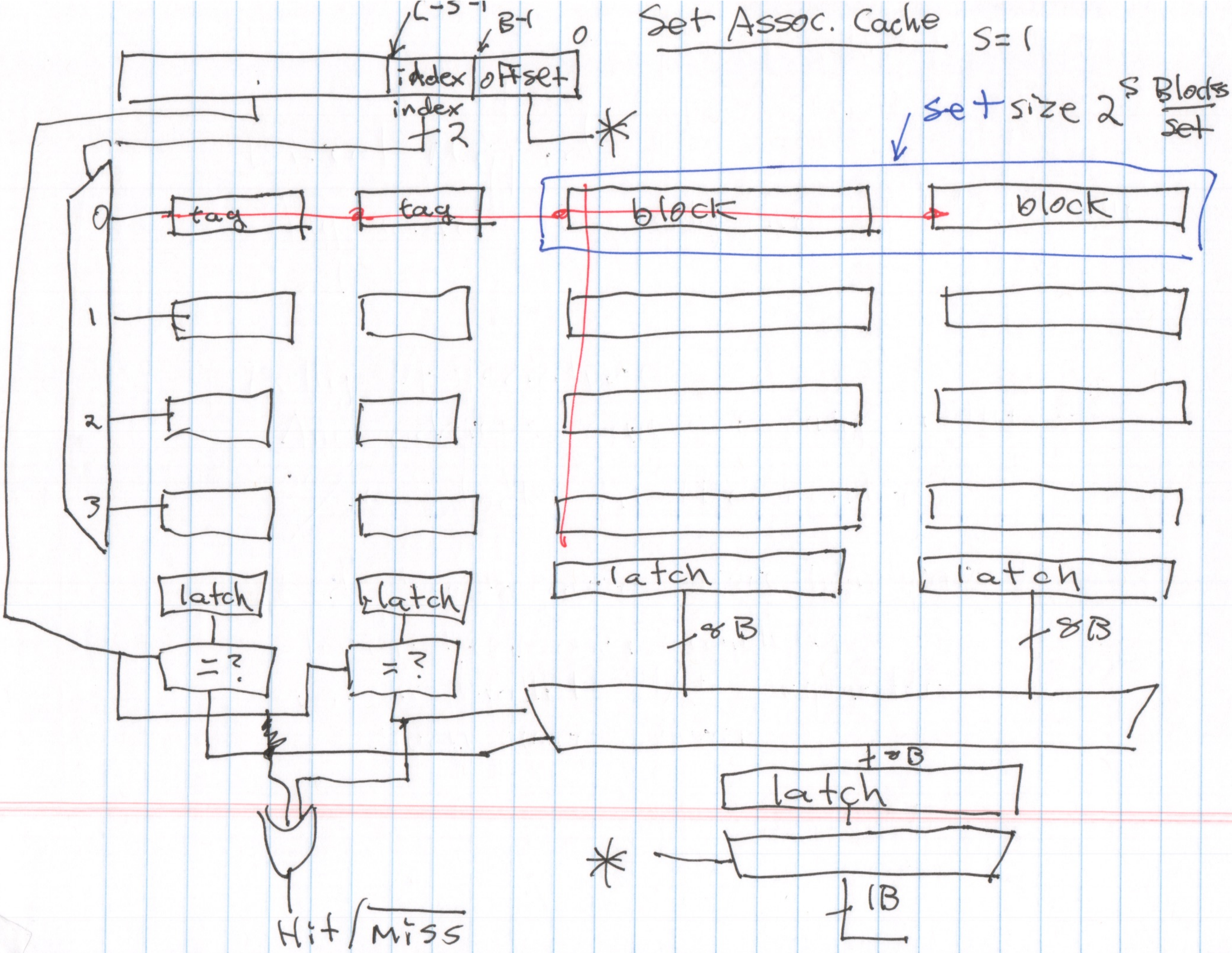
slower because of searching

or if in parallel $2^C/2^B$ comparators

= much more power

Set Assoc. Cache

$S = 1$



In general, 2^C bytes total storage
2^B bytes per block
2^S blocks per set
are the cache dimensions
(C, B, S)

Direct-mapped $S=0$

Set-associative : 2^S-way set assoc. cache

Fully associative has only one set

$$\text{size of set} = 2^C / 2^B = 2^{C-B}$$

Fully associative is when $S = C-B$