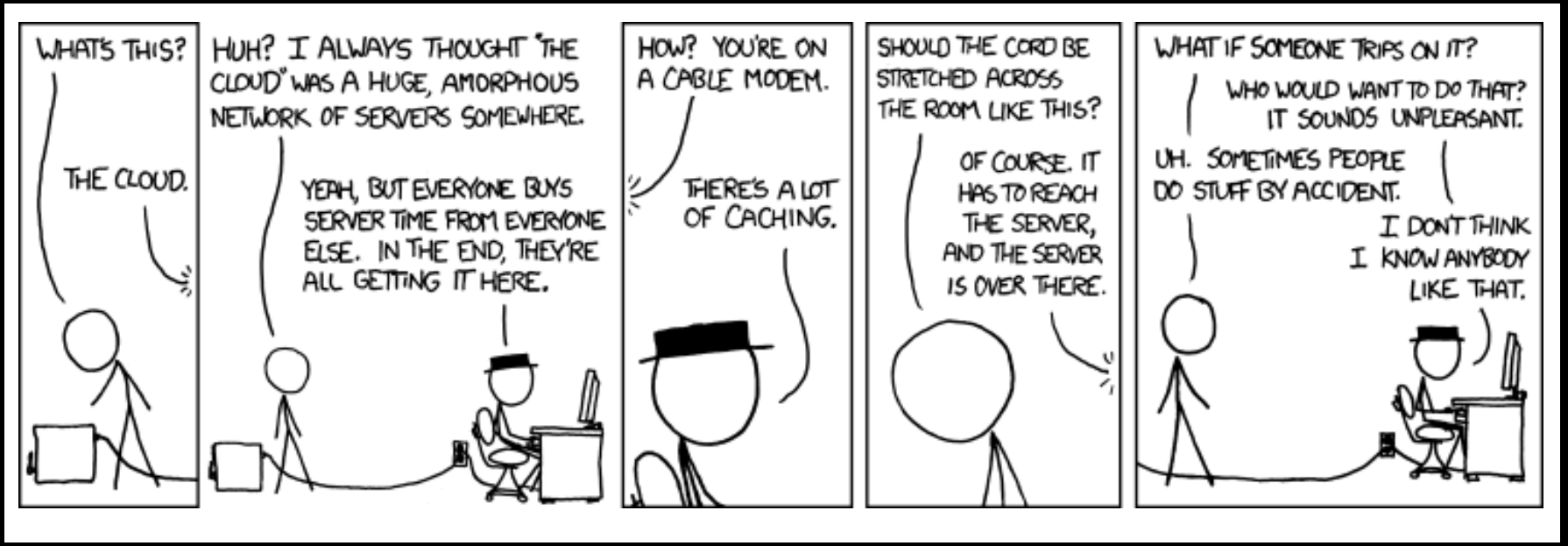
EECS 370 Discussion



xkcd.com

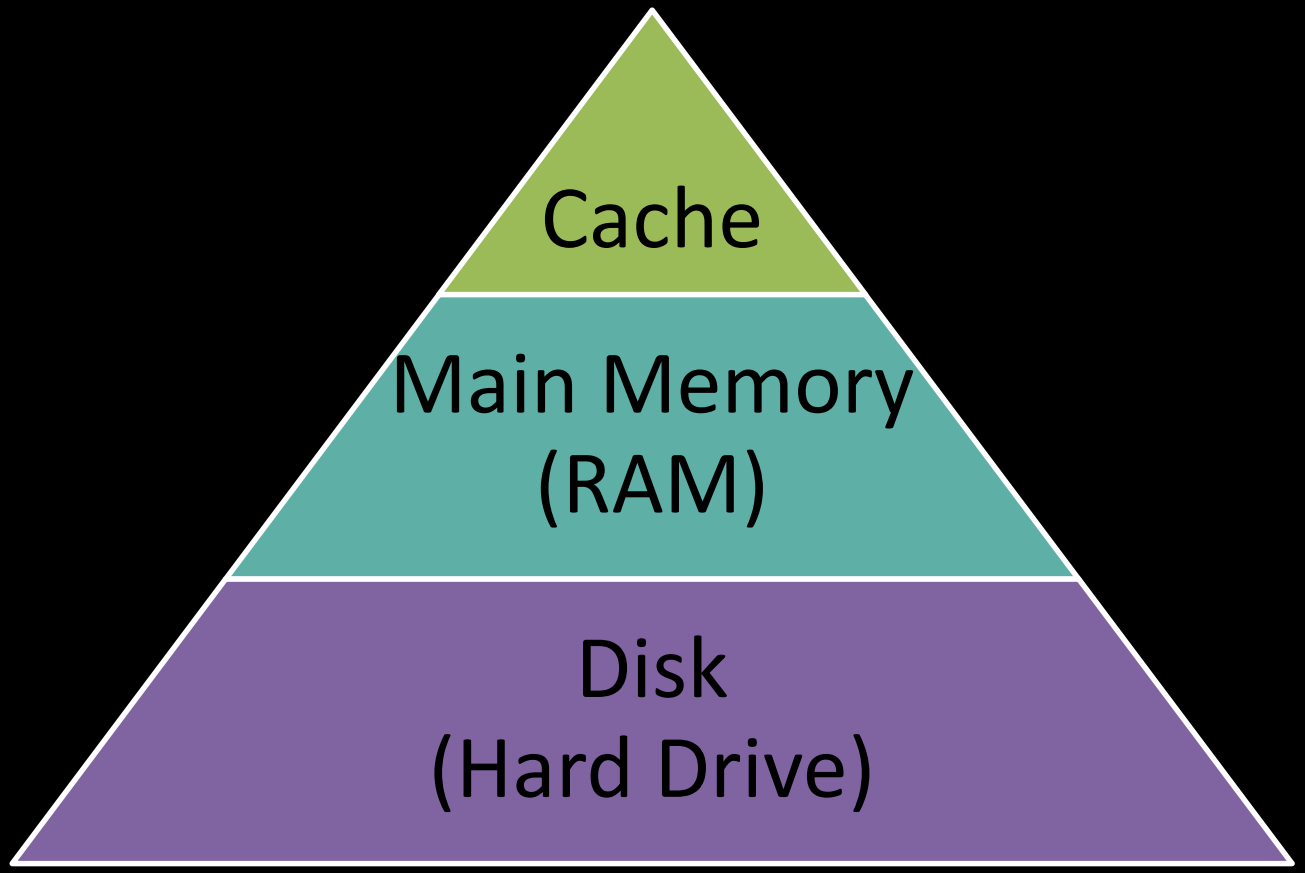
EECS 370 Discussion

Topics Today:

- Caches!!
  - Theory
  - Design
  - Examples

EECS 370 Discussion

Memory Hierarchy



EECS 370 Discussion

Memory Hierarchy

Real World Example - Intel i7

Cache Level	Size	Access Time
L1	64 kB	4 cycles
L2	256 kB	10 cycles
L3	8192 kB	40 cycles

RAM	8388608 kB	200 cycles
Disk	1073741824 kB	20000000 cycles

EECS 370 Discussion

Memory Hierarchy

Problem: Caches are very tiny, but memory is quite large

If memory accesses are totally random, caches are useless

Solution: Memory accesses really aren't random at all!

EECS 370 Discussion

Memory Hierarchy

Temporal Locality

You are likely to access memory locations multiple times

Spatial Locality

You are likely to access memory locations near each other

EECS 370 Discussion

Memory Hierarchy

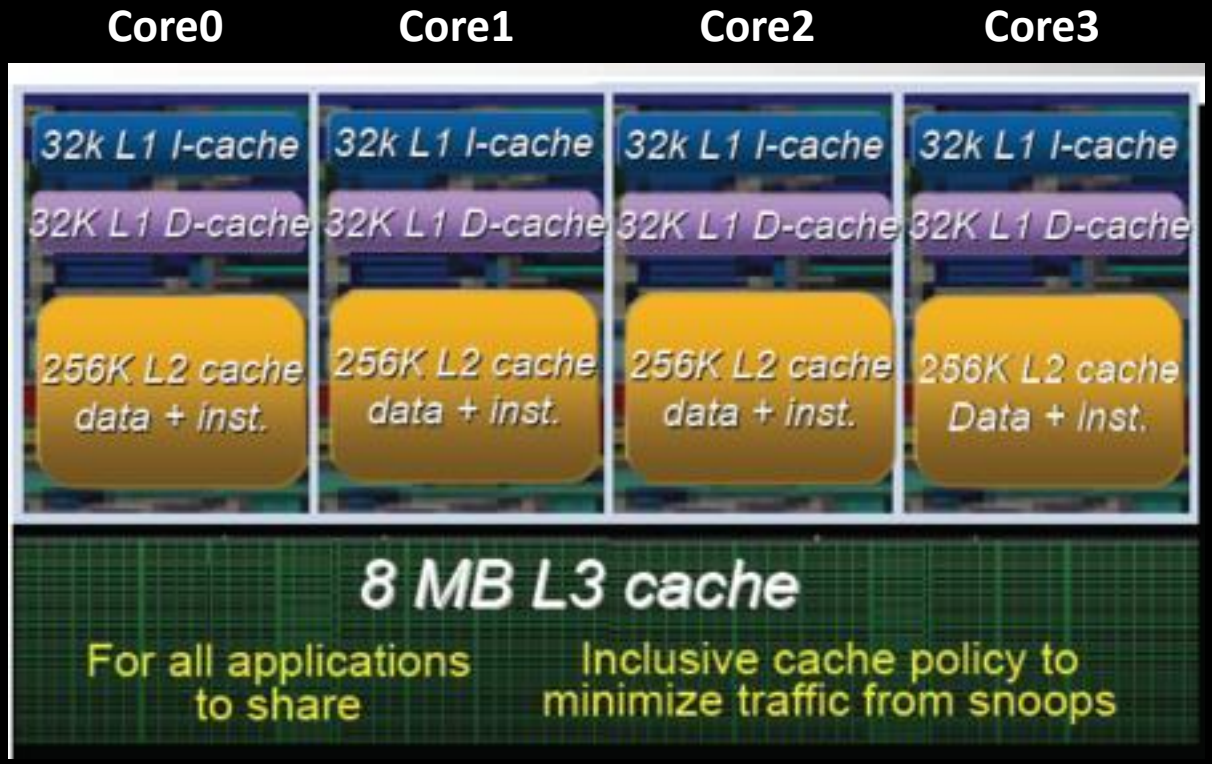
Example:

```
int data[10];
int sum = 0;
…
for (int i=0; i<10; i++) {
   sum += data[i];
}
```

EECS 370 Discussion

Memory Hierarchy

Real World Example - Intel i7



EECS 370 Discussion

Cache Design

Split Cache – portion your cache into two halves

I-Cache: Instruction Cache

D-Cache: Data Cache

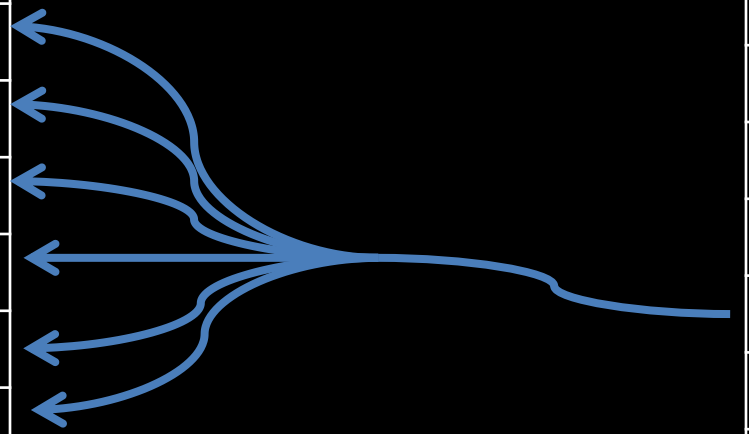
Why would we want to do this?

EECS 370 Discussion

Types of Caches

Fully Associative

Cache	
Tag	Data



Memory	
0x1000	10
0x1004	20
0x1008	30
0x100C	40
0x1010	50
0x1014	60
0x1018	70
0x101C	80
0x1020	90
0x1024	100
0x1028	110

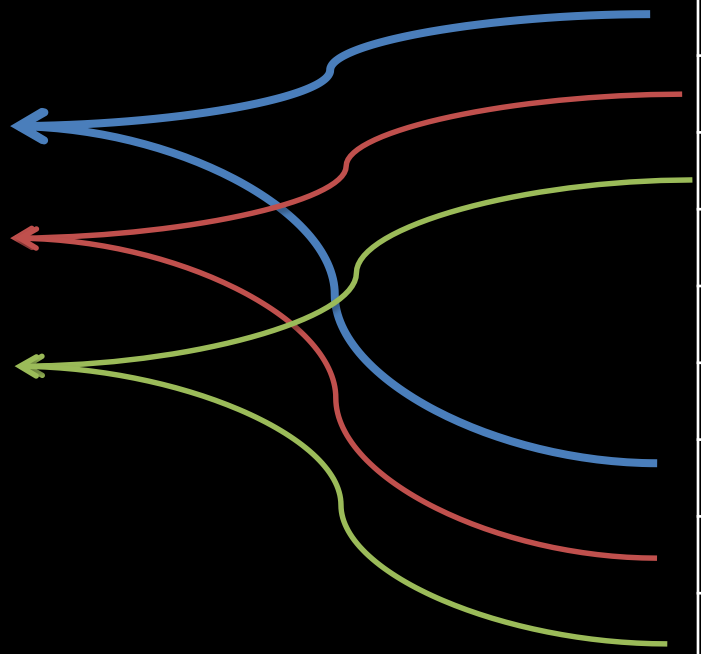
Blocks map to any cache line

EECS 370 Discussion

Types of Caches

Direct Mapped

Cache	
Tag	Data



Memory	
0x1000	10
0x1004	20
0x1008	30
0x100C	40
0x1010	50
0x1014	60
0x1018	70
0x101C	80
0x1020	90
0x1024	100
0x1028	110

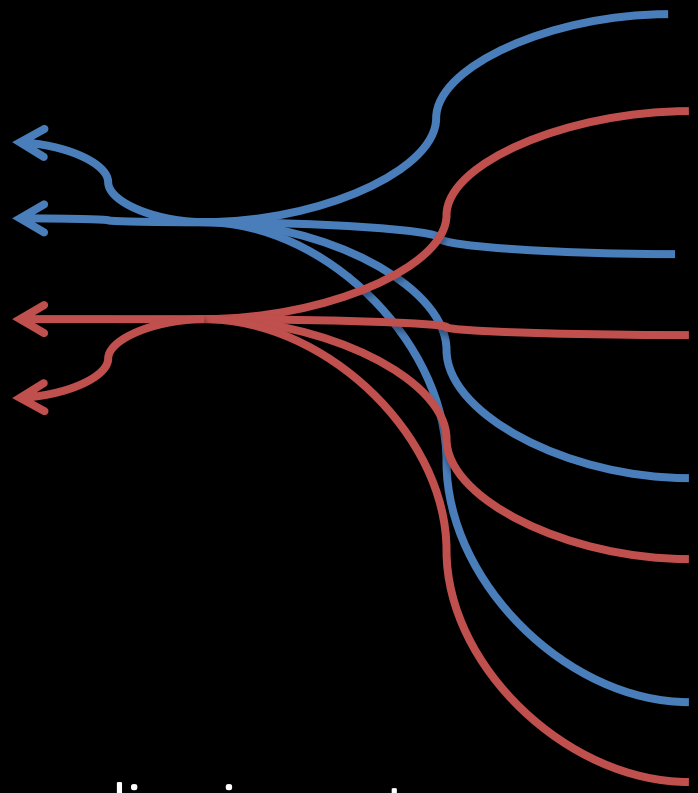
Blocks map to one cache line

EECS 370 Discussion

Types of Caches

Set Associative

Cache	
Tag	Data



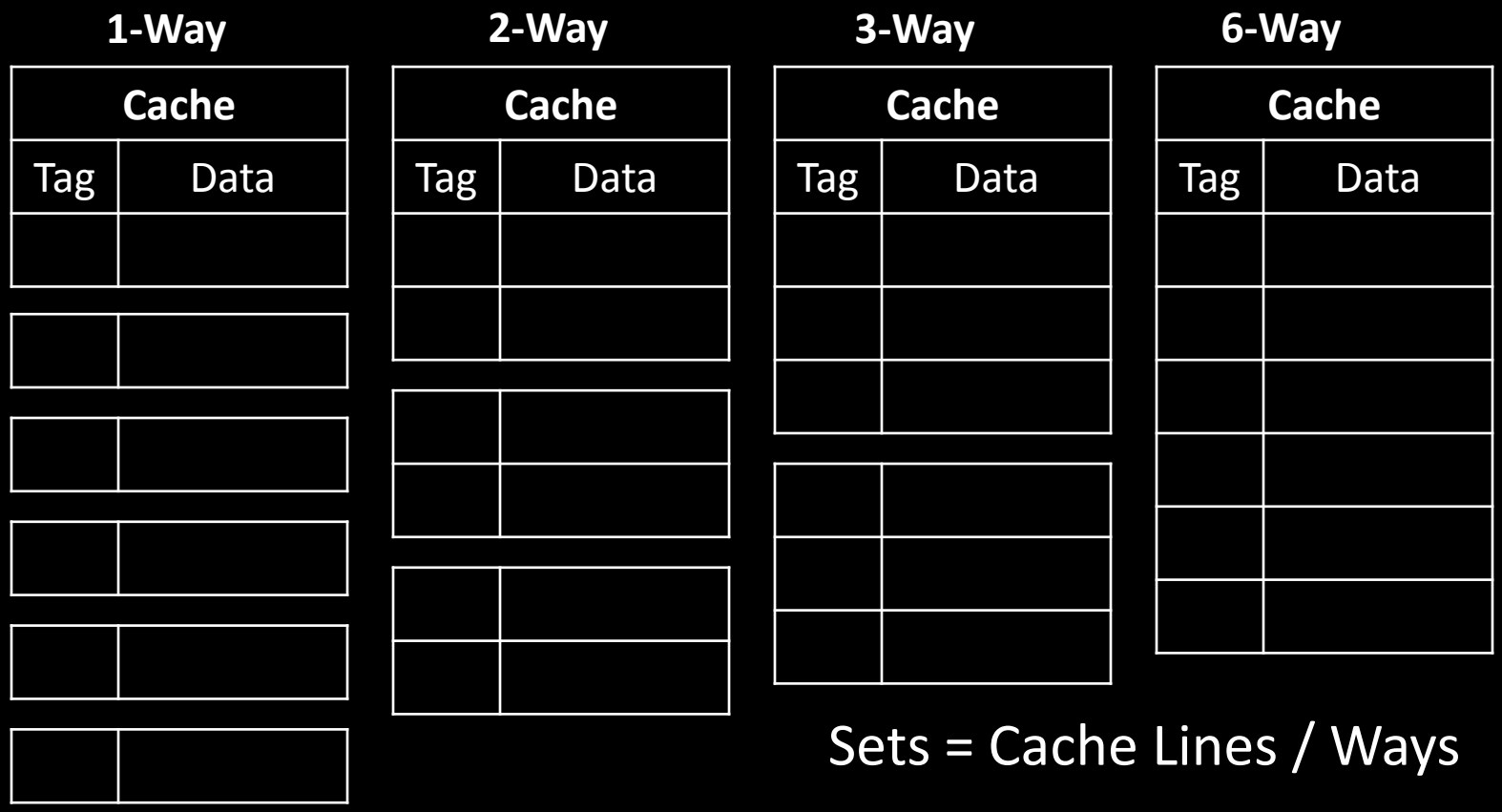
Memory	
0x1000	10
0x1004	20
0x1008	30
0x100C	40
0x1010	50
0x1014	60
0x1018	70
0x101C	80
0x1020	90
0x1024	100
0x1028	110

Blocks map to any line in a set

EECS 370 Discussion

Types of Caches

Set Associative



Sets = Cache Lines / Ways

EECS 370 Discussion

Three C's

There are three reasons a cache miss occurs

1) Compulsory – never been loaded before

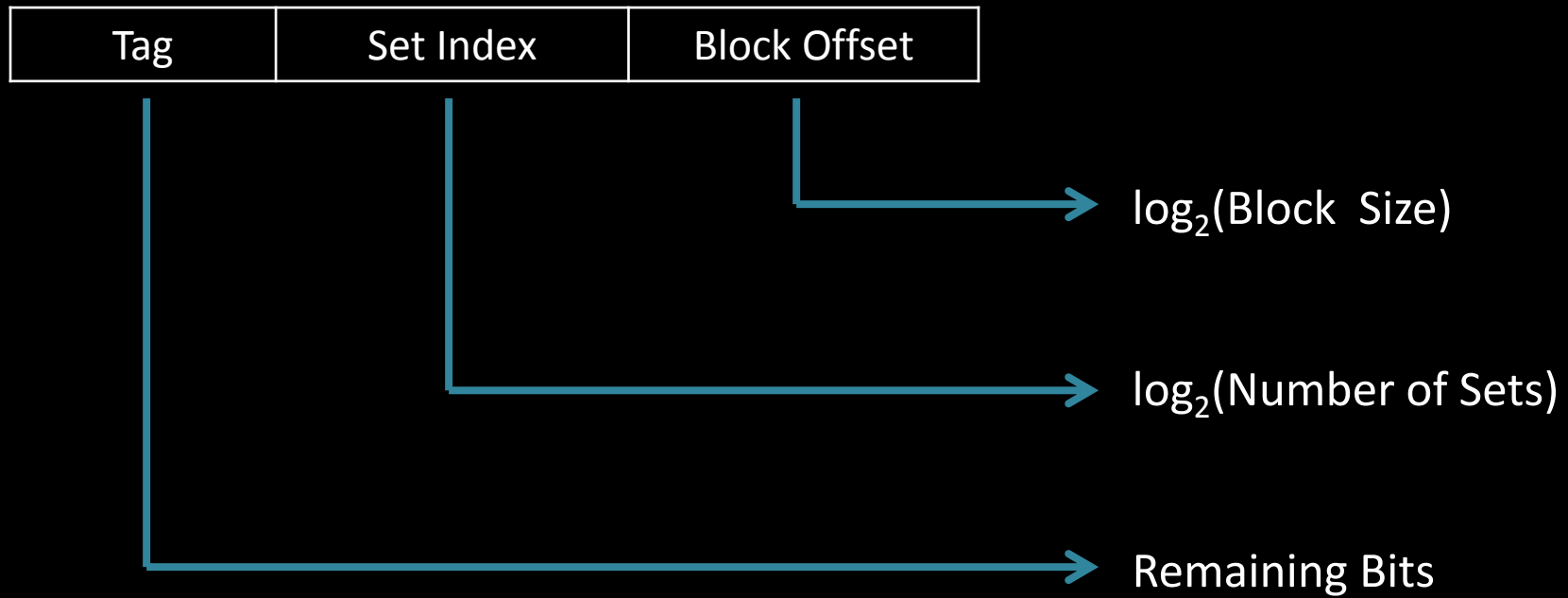
2) Capacity – evicted due to small cache size

3) Conflict – evicted due to overlap with another block

EECS 370 Discussion

Cache Addressing

Addresses split into:



EECS 370 Discussion

Cache Writing Policy

For Writes Only

On misses:

- Write Allocate – add to cache
- No Write Allocate – don't add to cache

On hits:

- Write Through – always write to memory
- Write Back – only write to cache

EECS 370 Discussion

Caches

Cache Examples

EECS 370 Discussion

Block size = 1 Word, Address = 16 bits

2-way Set Associative

Cache	
Tag	Data

Memory	
0x1000	10
0x1004	20
0x1008	30
0x100C	40
0x1010	50
0x1014	60
0x1018	70
0x101C	80
0x1020	90
0x1024	100
0x1028	110
0x102C	120

EECS 370 Discussion

Caches

Cache Examples

Options:

- Type of Cache
- Cache Size
- Block Size
- Address Bits
- Write Policy