

last time

implementing stalling and squashing

- pipeline registers: where do instructions need to go
- need to fill holes with nops ('bubbles')
- logic based on icodes in each stage

latency gap between registers, cache, main memory

temporal locality

- accessed now — accessed again soon
- so caches should store recently accessed things

spatial locality

- accessed now — nearby accessed soon
- so caches should store adjacent things together

Quiz: Question 1

Question 1:

cycle N: E (fetch), D (decode), C (execute), B (memory), A (writeback)

cycle N+1: E, D, C, (nop), (nop)

need to set `stall_D` and `stall_E`

disable changing `fD` and `dE` register bank outputs

not mentioned in question:

squashing B, A involve doing things in addition to pipeline registers

example: for A: MUX to override `reg_dstE/reg_dstM`

Quiz: Question 4: temporal locality

A. iterating through a large array on the stack

accessing particular array element: won't access again
(accessing machine code for loop: will access again soon)

B. accesses to machine code of small, frequently used function

access piece of machine code: will access again next call (which is soon:
"frequent")

should have specified I meant only data accesses for A, but I don't think it changes the answer

comparing mix of very good + very bad temporal locality with good temporal locality

Quiz: Question 5: temporal locality

repeated accesses to hashtable key:

- particular element (or small number of elements with chaining, etc.)
accessed now and will be accessed very quickly on next hashtable lookup

repeated linear search:

- particular element accessed now won't be accessed again on next search
until all previous elements are accessed again

probably similar temporal locality between both for machine code

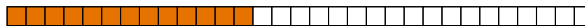
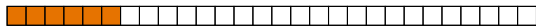
- both running loops without that much code
similar operations for each array element accessed (check if key matches)

(but I should have specified to only compare the data accesses)

Quiz: Question 6: spatial locality

accesses to an array and/or temporary arrays during a merge sort operation

merge operation: scanning two input arrays, writing output array in order
accessing element X of array now: will access element $X+1$ very soon



accesses to a sorted array during repeated binary search operations

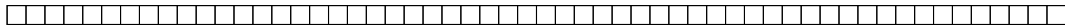
first access of search: middle of array

second access: half the array size away from that

third access: quarter the array size away from that

...

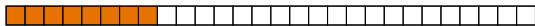
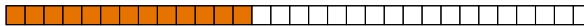
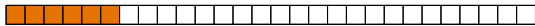
except for last few accesses: next access will be far away



Quiz: Question 6: spatial locality

accesses to an array and/or temporary arrays during a merge sort operation

merge operation: scanning two input arrays, writing output array in order
accessing element X of array now: will access element $X+1$ very soon



accesses to a sorted array during repeated binary search operations

first access of search: middle of array

second access: half the array size away from that

third access: quarter the array size away from that

...

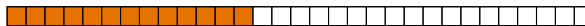
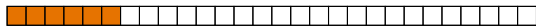
except for last few accesses: next access will be far away



Quiz: Question 6: spatial locality

accesses to an array and/or temporary arrays during a merge sort operation

merge operation: scanning two input arrays, writing output array in order
accessing element X of array now: will access element $X+1$ very soon



accesses to a sorted array during repeated binary search operations

first access of search: middle of array

second access: half the array size away from that

third access: quarter the array size away from that

...

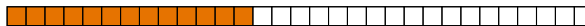
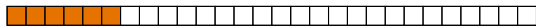
except for last few accesses: next access will be far away



Quiz: Question 6: spatial locality

accesses to an array and/or temporary arrays during a merge sort operation

merge operation: scanning two input arrays, writing output array in order
accessing element X of array now: will access element $X+1$ very soon



accesses to a sorted array during repeated binary search operations

first access of search: middle of array

second access: half the array size away from that

third access: quarter the array size away from that

...

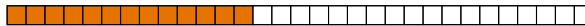
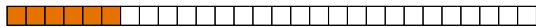
except for last few accesses: next access will be far away



Quiz: Question 6: spatial locality

accesses to an array and/or temporary arrays during a merge sort operation

merge operation: scanning two input arrays, writing output array in order
accessing element X of array now: will access element $X+1$ very soon



accesses to a sorted array during repeated binary search operations

first access of search: middle of array

second access: half the array size away from that

third access: quarter the array size away from that

...

except for last few accesses: next access will be far away



building a (direct-mapped) cache

Cache

value
00 00
00 00
00 00
00 00

cache block: 2 bytes

Memory

addresses	bytes
00000-00001	00 11
00010-00011	22 33
00100-00101	55 55
00110-00111	66 77
01000-01001	88 99
01010-01011	AA BB
01100-01101	CC DD
01110-01111	EE FF
10000-10001	F0 F1
...	...

building a (direct-mapped) cache

read byte at 01011?

Cache

value
00 00
00 00
00 00
00 00

cache block: 2 bytes

Memory

addresses	bytes
00000-00001	00 11
00010-00011	22 33
00100-00101	55 55
00110-00111	66 77
01000-01001	88 99
01010-01011	AA BB
01100-01101	CC DD
01110-01111	EE FF
10000-10001	F0 F1
...	...

building a (direct-mapped) cache

read byte at 01011?

exactly **one place** for each address
spread out what can go in a block

Cache

Memory

index

00

01

10

11

value

00 00

00 00

00 00

00 00

addresses

00000-00001

00010-00011

00100-00101

00110-00111

01000-01001

01010-01011

01100-01101

01110-01111

10000-10001

...

bytes

00 11

22 33

55 55

66 77

88 99

AA BB

CC DD

EE FF

F0 F1

...

cache block: 2 bytes

direct-mapped

building a (direct-mapped) cache

read byte at 01011?

exactly **one place** for each address
spread out what can go in a block

Cache

Memory

index	value	addresses	bytes
00	00 00	00000-00001	00 11
01	00 00	00010-00011	22 33
10	00 00	00100-00101	55 55
11	00 00	00110-00111	66 77
		01000-01001	88 99
		01010-01011	AA BB
		01100-01101	CC DD
		01110-01111	EE FF
		10000-10001	F0 F1
		...	...

cache block: 2 bytes

direct-mapped

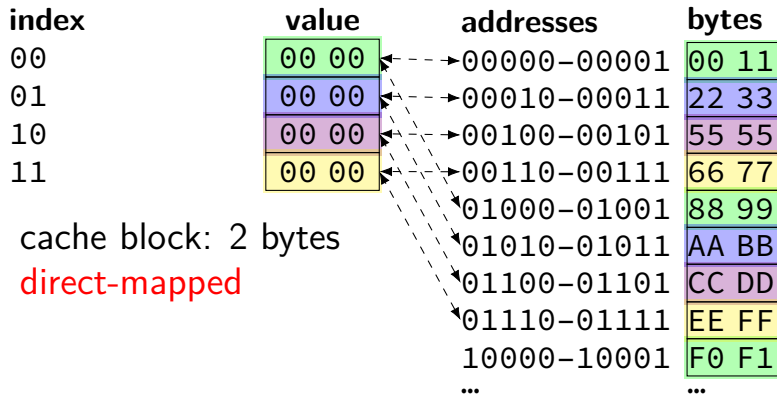
building a (direct-mapped) cache

read byte at 01011?

exactly **one place** for each address
spread out what can go in a block

Cache

Memory



building a (direct-mapped) cache

read byte at 01011?

Cache			Memory	
index	valid	value	addresses	bytes
00	0	00 00		1
01	0	00 00	00010-00011	22 33
10	0		0-00101	55 55
11	0	00 00	00110-00111	66 77
			01000-01001	88 99
			01010-01011	AA BB
			01100-01101	CC DD
			01110-01111	EE FF
			10000-10001	F0 F1
			...	...

cache block: 2 bytes
direct-mapped

is this even a value?

need extra bit to know

building a (direct-mapped) cache

read byte at 01011?

invalid, fetch

Cache

index	valid	value
00	0	00 00
01	1	AA BB
10	0	00 00
11	0	00 00

cache block: 2 bytes

direct-mapped

Memory

addresses	bytes
00000-00001	00 11
00010-00011	22 33
00100-00101	55 55
00110-00111	66 77
01000-01001	88 99
01010-01011	AA BB
01100-01101	CC DD
01110-01111	EE FF
10000-10001	F0 F1
...	...

building a (direct-mapped) cache

read byte at 01011?

invalid, fetch

Cache				Memory	
index	valid	tag	value	addresses	bytes
00	0	00	00 00	00000-00001	00 11
01	1	01	AA BB	00010-00011	22 33
10	0	00	00 00	00100-00101	55 55
11	0			00110-00111	66 77
				01000-01001	88 99
				01010-01011	AA BB
				01100-01101	CC DD
				01110-01111	EE FF
				10000-10001	F0 F1
				...	...

value from 01010 or 00010?

need tag to know

cache block: 2 bytes
direct-mapped

building a (direct-mapped) cache

read byte at 01011?

invalid, fetch

Cache

index	valid	tag	value
00	0	00	00 00
01	1	01	AA BB
10	0	00	00 00
11	0	00	00 00

cache block: 2 bytes

direct-mapped

Memory

addresses	bytes
00000-00001	00 11
00010-00011	22 33
00100-00101	55 55
00110-00111	66 77
01000-01001	88 99
01010-01011	AA BB
01100-01101	CC DD
01110-01111	EE FF
10000-10001	F0 F1
...	...

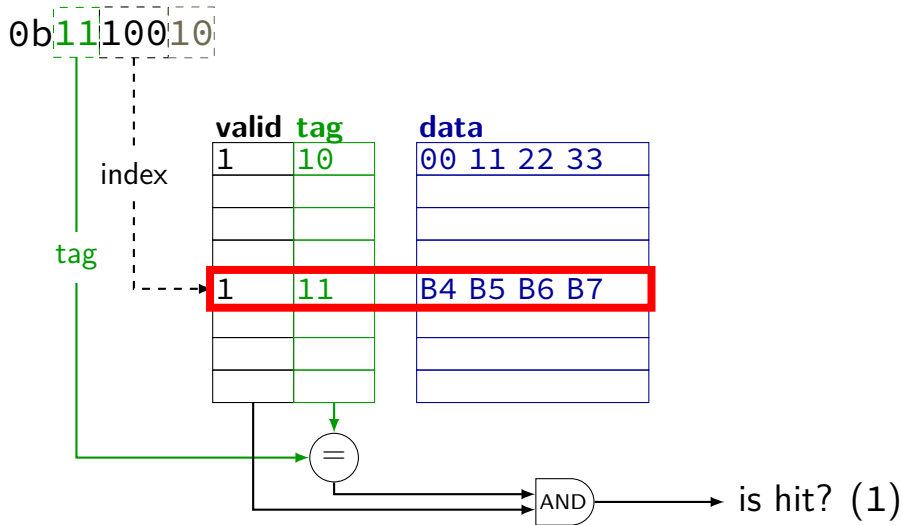
cache operation (read)

0b1110010

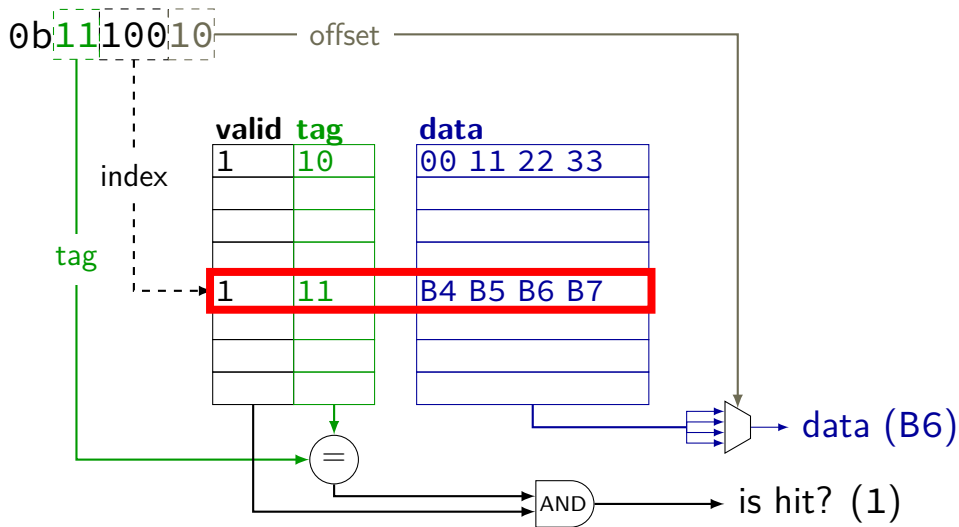
index

valid	tag	data
1	10	00 11 22 33
1	11	B4 B5 B6 B7

cache operation (read)



cache operation (read)



terminology

row = set

preview: change how much is in a row

Tag-Index-Offset (TIO)

address 001111 (stores value 0xFF)

cache tag index offset

2 byte blocks, 4 sets

2 byte blocks, 8 sets

4 byte blocks, 2 sets

2 byte blocks, 4 sets

index	valid	tag	value
00	1	000	00 11
01	1	001	AA BB
10	0	--	-- --
11	1	001	EE FF

4 byte blocks, 2 sets

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

2 byte blocks, 8 sets

index	valid	tag	value
000	1	00	00 11
001	1	01	F1 F2
010	0	--	-- --
011	0	--	-- --
100	0	--	-- --
101	1	00	AA BB
110	0	--	-- --
111	1	00	EE FF

Tag-Index-Offset (TIO)

address 00111**1** (stores value 0xFF)

cache tag index offset

2 byte blocks, 4 sets

1

2 byte blocks, 8 sets

1

4 byte blocks, 2 sets

2 byte blocks, 4 sets

index	valid	tag	value
00	1	000	00 11
01	1	001	AA BB
10	0	--	-- --
11	1	001	EE FF

4 byte blocks, 2 sets

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

2 byte blocks, 8 sets

index	valid	tag	value
000	0	--	-- --
001	0	--	-- --
010	1	00	AA BB
011	0	--	-- --
100	1	00	EE FF
101	0	--	-- --
110	0	--	-- --
111	1	00	EE FF

2 = 2¹ bytes in block
1 bit to say which byte

Tag-Index-Offset (TIO)

address 0011**11** (stores value 0xFF)

cache tag index offset

2 byte blocks, 4 sets	1
2 byte blocks, 8 sets	1
4 byte blocks , 2 sets	11

2 byte blocks, 4 sets

index	valid	tag	value
00	1	000	00 11
01	1	001	AA BB
10	0		
11	1		

4 byte

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

4 = 2² bytes in block
2 bits to say which byte

2 byte blocks, 8 sets

index	valid	tag	value
000	1	00	00 11
001	1	01	F1 F2
	0	--	-- --
	0	--	-- --
	0	--	-- --
	1	00	AA BB
	0	--	-- --
	1	00	EE FF

Tag-Index-Offset (TIO)

address 001111 (stores value 0xFF)

cache	tag	index	offset
2 byte blocks, 4 sets		11	1
2 byte blocks, 8 sets			1
4 byte blocks, 2 sets		1	11

2 byte blocks, 4 sets

index	valid	tag	value
00	1	000	00 11
01	1	001	AA BB
10	0	--	-- --
11	1	001	EE FF

2 byte blocks, 8 sets

index	valid	tag	value
000	1	00	00 11
			F1 F2
			-- --
			-- --
100	0	--	-- --
101	1	00	AA BB
110	0	--	-- --
111	1	00	EE FF

$2^2 = 4$ sets
2 bits to index set

4 byte blocks, 2 sets

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

Tag-Index-Offset (TIO)

address 001111 (stores value 0xFF)

cache	tag	index	offset
2 byte blocks, 4 sets	11	1	
2 byte blocks, 8 sets	111	1	
4 byte blocks, 2 sets	1	11	

2 byte blocks, 4 sets

index	valid	tag	value
00	1	000	00 11
01	1	001	AA BB
10	0	--	-- --
11	1		

$2^3 = 8$ sets
3 bits to index set

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

2 byte blocks, 8 sets

index	valid	tag	value
000	1	00	00 11
001	1	01	F1 F2
010	0	--	-- --
011	0	--	-- --
100	0	--	-- --
101	1	00	AA BB
110	0	--	-- --
111	1	00	EE FF

Tag-Index-Offset (TIO)

address 001111 (stores value 0xFF)

cache	tag	index	offset
2 byte blocks, 4 sets	11	1	
2 byte blocks, 8 sets	111	1	
4 byte blocks, 2 sets	1	11	

2 byte blocks, 4 sets

index	valid	tag	value
00	1	000	00 11
01	1	001	AA BB
10	0	--	-- --
11	1	001	EE FF

4 byte blocks, 2 sets

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

2 byte blocks, 8 sets

index	valid	tag	value
000	1	00	00 11
001	1	01	F1 F2
010	0	--	-- --
011	0	--	-- --
100	0	--	-- --
101	0	--	-- --
110	0	--	-- --
111	1	00	EE FF

$2^1 = 2$ sets

1 bit to index set

Tag-Index-Offset (TIO)

address 001111 (stores value 0xFF)

cache	tag	index	offset
2 byte blocks, 4 sets	001	11	1
2 byte blocks, 8 sets	00	111	1
4 byte blocks, 2 sets	001	1	11

tag — whatever is left over

00	1	000	00 11
01	1	001	AA BB
10	0	--	-- --
11	1	001	EE FF

4 byte blocks, 2 sets

index	valid	tag	value
0	1	000	00 11 22 33
1	1	001	CC DD EE FF

2 byte blocks, 8 sets

index	valid	tag	value
000	1	00	00 11
001	1	01	F1 F2
010	0	--	-- --
011	0	--	-- --
100	0	--	-- --
101	1	00	AA BB
110	0	--	-- --
111	1	00	EE FF

Tag-Index-Offset formulas (direct-mapped only)

m memory addresses bits (Y86-64: 64)

$S = 2^s$ number of sets

s (set) index bits

$B = 2^b$ block size

b (block) offset bits

$t = m - (s + b)$ tag bits

$C = B \times S$ cache size (if direct-mapped)

TIO: exercise

64-byte blocks, 128 set cache

stores $64 \times 128 = 8192$ bytes

if addresses 32-bits, then how many tag/index/offset bits?

which bytes are stored in the same block as byte from 0x1037?

- A. byte from 0x1011
- B. byte from 0x1021
- C. byte from 0x1035
- D. byte from 0x1041

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

index	valid	tag	value
00	0		
01	0		
10	0		
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

index	valid	tag	value
00	0		
01	0		
10	0		
11	0		

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

index	valid	tag	value
00	0		
01	0		
10	0		
11	0		

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	0		
10	0		
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	0		
10	0		
11	0		

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	1	01100	mem[0x62] mem[0x63]
10	0		
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	01100	mem[0x60] mem[0x61]
01	1	01100	mem[0x62] mem[0x63]
10	0		
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	01100	mem[0x60] mem[0x61]
01	1	01100	mem[0x62] mem[0x63]
10	0		
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	miss
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	1	01100	mem[0x62] mem[0x63]
10	0		
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	miss
01100100 (64)	miss

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	1	01100	mem[0x62] mem[0x63]
10	1	01100	mem[0x64] mem[0x65]
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	miss
01100100 (64)	miss

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	1	01100	mem[0x62] mem[0x63]
10	1	01100	mem[0x64] mem[0x65]
11	0		

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	miss
01100100 (64)	miss

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	1	01100	mem[0x62] mem[0x63]
10	1	01100	mem[0x64] mem[0x65]
11	0		

miss caused by conflict

exercise

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

4 byte blocks, 4 sets

index	valid	tag	value
00			
01			
10			
11			

exercise

4 byte blocks, 4 sets

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

index	valid	tag	value
00			
01			
10			
11			

how is the 8-bit address 61 (01100001) split up into tag/index/offset?

b block offset bits;

$B = 2^b$ byte block size;

s set index bits; $S = 2^s$ sets ;

$t = m - (s + b)$ tag bits (leftover)

exercise

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

4 byte blocks, 4 sets

index	valid	tag	value
00			
01			
10			
11			

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 4 = 2^b$ byte block size

$b = 2$ (block) offset bits

$t = m - (s + b) = 4$ tag bits

exercise

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 4 = 2^b$ byte block size

$b = 2$ (block) offset bits

$t = m - (s + b) = 4$ tag bits

4 byte blocks, 4 sets

index	valid	tag	value
00			
01			
10			
11			

exercise

address (hex)	result
00000000 (00)	
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

4 byte blocks, 4 sets

index	valid	tag	value
00			
01			
10			
11			

exercise: which accesses are hits?

example access pattern (1)

2 byte blocks, 4 sets

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	miss
01100100 (64)	miss

tag index offset

$m = 8$ bit addresses

$S = 4 = 2^s$ sets

$s = 2$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 5$ tag bits

index	valid	tag	value
00	1	00000	mem[0x00] mem[0x01]
01	1	01100	mem[0x62] mem[0x63]
10	1	01100	mem[0x64] mem[0x65]
11	0		

miss caused by conflict

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	0			0		
1	0			0		

multiple places to put values with same index
avoid conflict misses

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	0		set 0	0		
1	0		set 1	0		

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	0			0		
1	0			0		

Diagram illustrating a 2-way set associative cache structure with 2 byte blocks and 2 sets. The cache is divided into two sets, each containing two entries (index 0 and index 1). Each entry has three fields: valid, tag, and value. The first set is labeled "way 0" and the second set is labeled "way 1". In both sets, the valid bit is 0 for both index 0 and index 1, indicating that the cache is currently empty.

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	0			0		
1	0			0		

$m = 8$ bit addresses

$S = 2 = 2^s$ sets

$s = 1$ (set) index bits

$B = 2 = 2^b$ byte block size

$b = 1$ (block) offset bits

$t = m - (s + b) = 6$ tag bits

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	0		
1	0			0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	0		
1	0			0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	0		
1	1	011000	mem[0x62] mem[0x63]	0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	1	011000	mem[0x60] mem[0x61]
1	1	011000	mem[0x62] mem[0x63]	0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	
00000000 (00)	
01100100 (64)	

tag index offset

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	1	011000	mem[0x60] mem[0x61]
1	1	011000	mem[0x62] mem[0x63]	0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	
01100100 (64)	

tag index offset

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	1	011000	mem[0x60] mem[0x61]
1	1	011000	mem[0x62] mem[0x63]	0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	hit
01100100 (64)	

tag index offset

adding associativity

2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	1	011000	mem[0x60] mem[0x61]
1	1	011000	mem[0x62] mem[0x63]	0		

address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	hit
01100100 (64)	miss

needs to replace block in set 0!

tag index offset

adding associativity

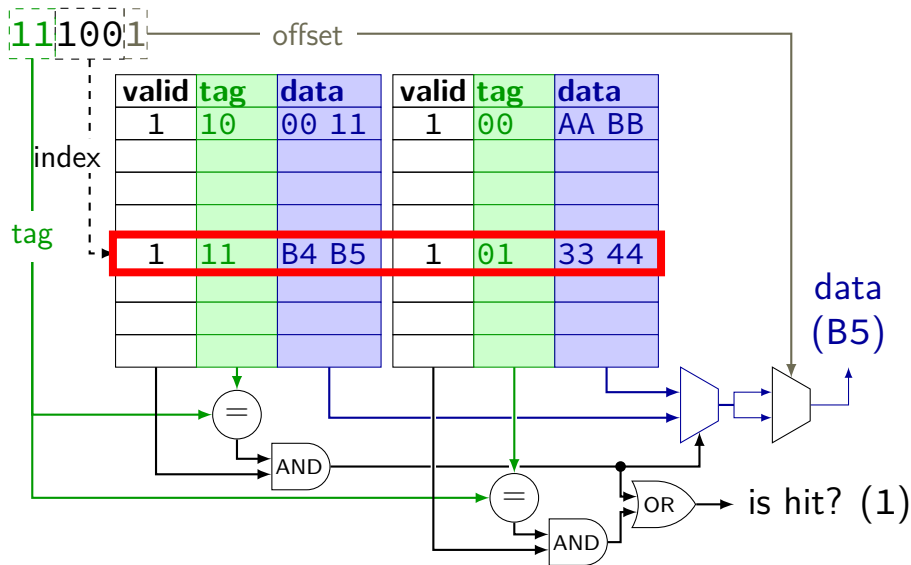
2-way set associative, 2 byte blocks, 2 sets

index	valid	tag	value	valid	tag	value
0	1	000000	mem[0x00] mem[0x01]	1	011000	mem[0x60] mem[0x61]
1	1	011000	mem[0x62] mem[0x63]	0		

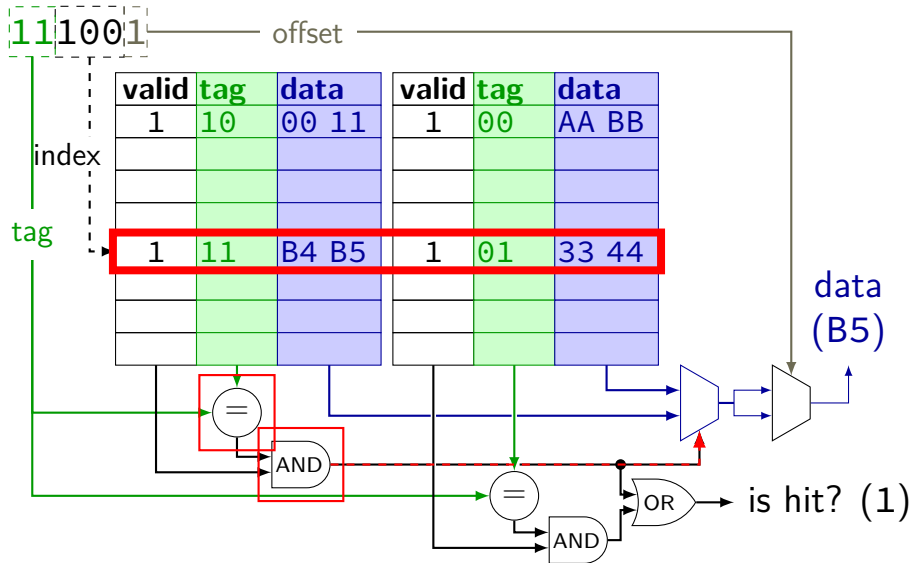
address (hex)	result
00000000 (00)	miss
00000001 (01)	hit
01100011 (63)	miss
01100001 (61)	miss
01100010 (62)	hit
00000000 (00)	hit
01100100 (64)	miss

tag indexoffset

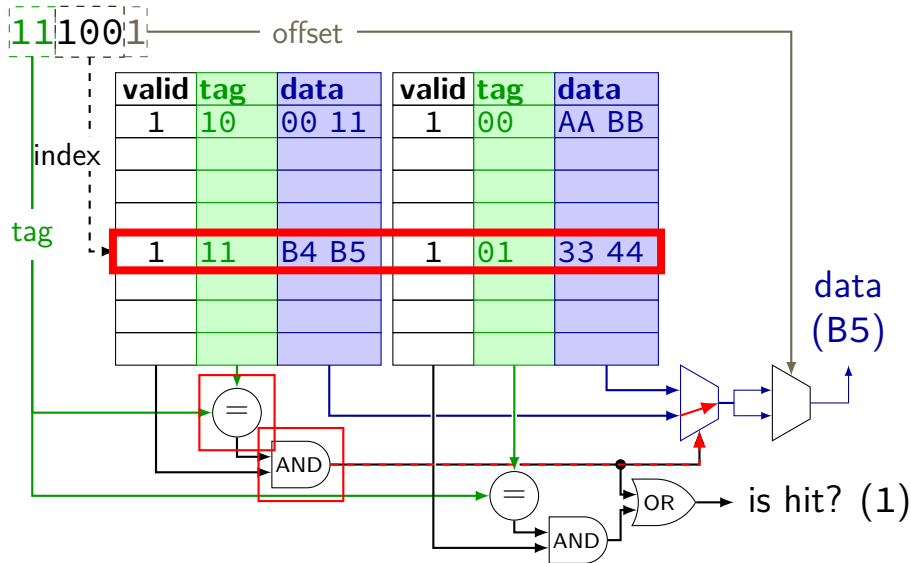
cache operation (associative)



cache operation (associative)



cache operation (associative)



associative lookup possibilities

none of the blocks for the index are valid

none of the valid blocks for the index match the tag
something else is stored there

one of the blocks for the index is valid and matches the tag

backup slides