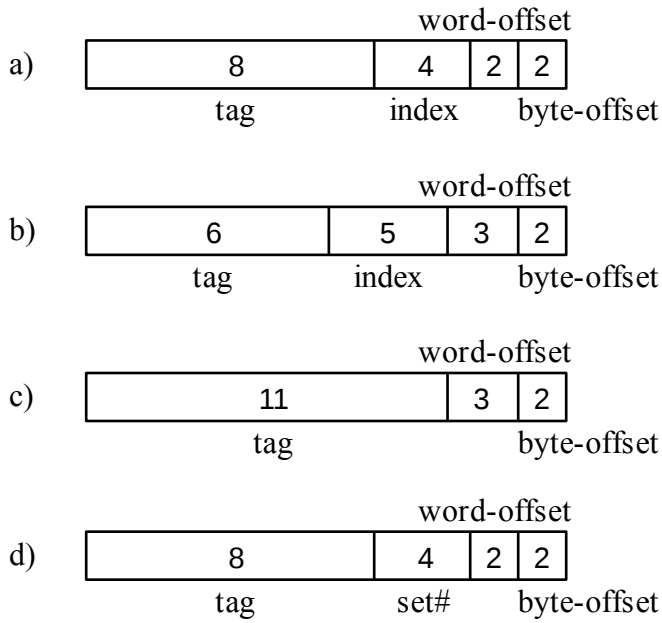


1. Temporary storage constructed from fast memory that exploits the principle of locality.
2. L1 cache is closest to the processor and part of the CPU whereas L2 cache may be part of the CPU or a separate component on the motherboard. L1 cache is constructed of faster and more expensive memory chips than L2 cache.
3. A tag is used to identify a specific block of memory that is stored in the cache.
4. Temporal locality refers to time. When a word of memory is referenced it will tend to be referenced again relatively soon. Program loop constructs and variables used within loops are examples that exhibit temporal locality.
5. a) **2192 bits.**
 - 1 bit valid flag
 - 8 bit tag field
 - + 128 data bits: 4 words/block @ 32 bits/word
 - 137 bits x 16 blocks = 2192 bits
- b) **8448 bits.**
 - 1 bit valid flag
 - 1 bit dirty flag
 - 6 bit tag
 - + 256 data bits: 8 words/block @ 32 bits/word
 - 264 bits x 32 blocks = 8448 bits
- c) **4304 bits.**
 - 1 bit valid flag
 - 1 bit dirty flag
 - 11 bit tag
 - + 256 data bits: 8 words/block @ 32 bits/word
 - 269 bits x 16 blocks = 4304 bits
- d) **8768 bits.**
 - 1 bit valid flag
 - 8 bit tag
 - + 128 data bits: 4 words/block @ 32 bits/word
 - 137 bits/entry
 - x 4 entries/set
 - 548 bits per set x 16 sets = 8768 bits

6.



7. The final cache contents is shown below. The data block entries shows the address of the last reference from the given cache entry. The hit ratio is 1/4.

	valid	tag	data
0	1	000	address 2
1	1	001	address 41
2	1	110	address 213
3	1	110	address 221

8. The final cache contents is shown below. The data block entries shows the address of the last reference from the given cache entry. The hit ratio is 5/12.

	V	LRU	tag	data
0	1	1	101	161
	1	0	000	2
1	1	0	001	41
	0			
2	1	0	110	213
	0			
3	1	0	110	221
	0			