3 Computer Design (tmj32)

A processor contains two cores, each with an L1 cache connected via a shared bus to an L2 cache, which is then connected to main memory. Each L1 is a direct-mapped, 4 KiB, write-back cache. The L2 is a 4-way set-associative 16 KiB cache with the least-recently-used replacement policy. All cache lines are 16 B long. The hierarchy is inclusive, runs the MSI cache coherence protocol and is initially empty.

(*a*) Considering this cache hierarchy,

(*i*) Explain whether it would be suitable for a system-on-chip that is only running single-threaded applications. [2 marks]

(*ii*) Explain whether it would be suitable for a system-on-chip where each core processes a small (e.g. $\leqslant$ 1 KiB) array of data at a time. [2 marks]

(*b*) Show the cache contents and coherence state of cache lines after each access in the following sequence of physical addresses, stating any assumptions you have made. All accesses are 4 bytes long and entirely complete before the next one starts.

```
Core 1:  Read   0xab18
Core 2:  Read   0xdb14
Core 1:  Write  0x1b10
Core 1:  Read   0xab14
Core 1:  Read   0xbb1c
Core 2:  Write  0xa010
Core 1:  Read   0x2b10
Core 2:  Read   0x1b10
Core 1:  Read   0xa018
Core 1:  Write  0x1b14
```

[8 marks]

(*c*) Describe the impact of each change below (in isolation) on the cache hierarchy.

(*i*) Increasing the L1 cache size.

(*ii*) Increasing the line size in all caches.

(*iii*) Increasing the associativity of the L1 caches.

(*iv*) Changing to an exclusive cache hierarchy.

[2 marks each]