

1. Describe and draw a sketch of the von Neumann architecture. Extend the drawing to show multiple cores and a cache and memory hierarchy with three levels of cache.
2. Explain the concept of pipelining. Given the specification for a pipeline, calculate r_{∞} and $n_{1/2}$.
3. Explain cache associativity and work an example showing how associativity affects cache misses for a code example.
4. Analyze and optimize a code example with respect to efficient use of cache memory.
5. Explain the concept of vectorization and how it can improve code performance.
6. Given the specifications for a multicore processor and its measured memory bandwidth, draw a roofline model for the processor. Evaluate the operational intensity of a piece of code and use the result to place it on the roofline model.
7. Compare and contrast shared memory and distributed memory parallel architectures. Explain what type of parallel programming paradigm(s) will work on each.
8. Given an MPI point-to-point communication example code, evaluate whether or not the code is safe (i.e., no possibility of deadlock).
9. Given an OpenMP code example, evaluate whether a race condition exists and if so, rewrite the code to be correct.
10. Use MPI collective communications calls correctly to carry out a given communication task.
11. Given the portion of a code that can be parallelized, calculate the maximum possible speedup of the entire code using Amdahl's law.
12. Given the runtimes for serial and parallel versions of a code, calculate parallel speedup and efficiency.