

Home ► My courses ► Parallel ► Lecture 5 - Roofline and Performance Modeling ► Quiz 5

Started on Sunday, 2 February 2020, 8:46 PM

State Finished

Completed on Sunday, 2 February 2020, 8:50 PM

Time taken 3 mins 42 secs

Marks 9.00/9.00

Grade 10.00 out of 10.00 (100%)

Question 1

Correct

Mark 1.00 out of 1.00

Which computational kernels are considered "compute limited"?

Select one:

- ☐ a. Kernels whose machine balance is lower than the architecture's arithmetic intensity
- ☐ b. Kernels whose arithmetic intensity is lower than the machine balance of the architecture
- ☒ c. Kernels whose arithmetic intensity is higher than the machine balance of the architecture ✓
- ☐ d. Kernels whose machine balance is larger than the architecture's arithmetic intensity

Your answer is correct.

The correct answer is: Kernels whose arithmetic intensity is higher than the machine balance of the architecture

Question 2

Correct

Mark 1.00 out of 1.00

Mark all reasons of using performance modeling and tools

Select one or more:

- ☒ a. To identify performance bottlenecks ✓
- ☒ b. To motivate software optimizations ✓
- ☒ c. To determine when it is time to stop optimizing the code ✓
- ☒ d. To predict performance on future machines ✓
- ☐ e. To identify bugs affecting correctness

Your answer is correct.

The correct answers are: To identify performance bottlenecks, To motivate software optimizations, To determine when it is time to stop optimizing the code, To predict performance on future machines

2/2/20, 8:50 PM

Question 3

Correct

Mark 1.00 out of
1.00

Mark all valid definitions of computational **depth** <https://moodle.xsede.org/mod/quiz/review.php?at...>

Select one or more:

- ☒ a. The critical path of a parallel computation ✓
- ☒ b. The fastest possible execution time of a parallel computation assuming we have an infinite number of processors ✓
- ☒ c. The longest series of sequential operations in a parallel computation ✓
- ☐ d. the gain in speed made by parallel execution compared to sequential execution

Your answer is correct.

The correct answers are: The longest series of sequential operations in a parallel computation, The critical path of a parallel computation, The fastest possible execution time of a parallel computation assuming we have an infinite number of processors

Question 4

Correct

Mark 1.00 out of
1.00

Mark all latency hiding mechanisms developed by hardware vendors

Select one or more:

- ☒ a. massive thread parallelism ✓
- ☒ b. prefetching ✓
- ☐ c. very-long instruction word (VLIW)
- ☒ d. out of order execution ✓

Your answer is correct.

The correct answers are: massive thread parallelism, prefetching, out of order execution

Question 5

Correct

Mark 1.00 out of
1.00

Roofline is a ...

Select one:

- ☐ a. latency-oriented performance model
- ☒ b. throughput-oriented performance model ✓

Your answer is correct.

The correct answer is: throughput-oriented performance model

Question 6

Correct

Mark 1.00 out of
1.00<https://moodle.xsede.org/mod/quiz/review.php?at...>

The most basic description of the roofline model considers the bandwidth to DRAM as its memory-bound line. What other types of memory accesses can be the limiting factor in a hierarchical roofline model? Mark all valid answers.

Select one or more:

- ☒ a. Bandwidth to L2 cache ✓
- ☒ b. Bandwidth to MCDRAM ✓
- ☒ c. Bandwidth to L1 cache ✓
- ☐ d. Translation Lookaside Buffer (TLB) misses

Your answer is correct.

The correct answers are: Bandwidth to L1 cache, Bandwidth to MCDRAM, Bandwidth to L2 cache

Question 7

Correct

Mark 1.00 out of
1.00

Consider the application of Little's Law to performance modeling. Assume that the system does not have a cache.

If the memory latency (i.e. latency to DRAM) is 98 nanoseconds, and the bandwidth to DRAM is 61 GB/s, then the processor-memory system must support outstanding memory requests for how many bytes of data?

Answer: 5978.00



The correct answer is: 5978.00

Quiz 5
Question 8

Correct

Mark 1.00 out of
1.00

<https://moodle.xsede.org/mod/quiz/review.php?at...>

Instruction mix of the code that is being executed affects its roofline in superscalar architectures. For example, Intel KNL is a 2-issue superscalar architecture with 2 floating-point (FP) data paths. This means that in order to achieve the peak flops performance on the roofline, all (100%) of the code's instructions need to be floating point.

Intel Haswell, on the other hand, is a 4-issue superscalar architecture with 2 FP data paths. Consequently, **at least what percentage of the code needs to be floating point** in order to achieve the peak flops performance on Haswell?

Select one:

- ☐ a. 100%
- ☐ b. 200%
- ☐ c. 25%
- ☒ d. 50% ✓

Your answer is correct.

The correct answer is: 50%

Question 9

Correct

Mark 1.00 out of
1.00

Modern CPUs often have variable clock rates that can change on the fly. This is also known as dynamic frequency scaling. **Mark all** possible reasons of **dynamically lowering the frequency of a CPU**.

Select one or more:

- ☒ a. To avoid overheating ✓
- ☒ b. To conserve energy in mobile devices ✓
- ☐ c. To accelerate computation

Your answer is correct.

The correct answers are: To conserve energy in mobile devices, To avoid overheating

◀ Lecture Video: Roofline and
Performance Modeling

Jump to...

Lecture Video: Sources of
Parallelism and Locality in Simulation
(Part 1a) ▶