Reading #3 Quiz

Started: Feb 10 at 6:07pm

Quiz Instructions

This quiz is based on **domain_specific_hw_accelerators.pdf (https://canvas.asu.edu/courses/212341/files/102624391?wrap=1)**

Question 1 1 pts

ASICs provide higher efficiency than GPUs for domain-specific acceleration but come at the cost of higher nonrecurring engineering (NRE) costs.

- (●) True
- ( ) False

Question 2 1 pts

In modern technology nodes, the energy required for moving data (such as reading from memory) and instruction processing (such as fetching an instruction) is significantly lower than performing a simple arithmetic operation.

- ( ) True
- (●) False

Question 3 1 pts

Which of the following is NOT a major technique used by domain-specific accelerators to achieve performance gains?

- ( ) Data specialization
- ( ) Parallelism
- (●) Increased clock speeds
- ( ) Optimized memory

○ Reduce overhead

Question 4 1 pts

Which of the following statements best describes the trade-offs between ASICs, FPGAs, and GPUs in domain-specific acceleration?

○ GPUs always outperform ASICs in domain-specific tasks due to their general-purpose nature.

○ FPGAs offer better efficiency than ASICs in all applications.

◉ ASICs provide the best efficiency but are difficult to reconfigure, whereas FPGAs offer reconfigurability at the cost of some efficiency, and GPUs balance efficiency and programmability.

○ ASICs are more flexible than GPUs in handling a variety of applications.

Question 5 1 pts

Consider an accelerator designed for a deep learning application. The accelerator employs a memory compression technique that reduces the memory footprint by a factor of 10x. If global memory bandwidth was the previous bottleneck, how would this impact performance?

○ Performance would decrease since compression requires additional processing power

◉ Performance would likely increase significantly since less data needs to be transferred

○ Performance would remain unchanged since compression does not affect computation

○ Performance would increase, but only if the number of processing elements is also increased

Question 6 1 pts

A company is considering an FPGA-based accelerator for genomic sequence alignment instead of an ASIC-based solution. Which of the following factors would MOST justify this decision?

○ The need for maximum efficiency regardless of flexibility

◉ The likelihood that genomic alignment algorithms will evolve over time

○ The need for higher single-threaded performance

Question 7 1 pts

Which of the following design strategies would BEST ensure that a domain-specific accelerator remains useful even as new algorithms emerge?

- ○ Implementing fixed-function hardware with no software programmability
- ○ Using a GPU to maintain generality
- ◉ Adding specialized instructions to a programmable processor

Question 8 1 pts

Your company is designing a DSA for natural language processing (NLP) tasks. Given the large memory footprint of transformer-based models, how would you optimize your accelerator to handle such workloads efficiently? Write a few techniques you'd apply. Answer in no more than 25 words.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

I would apply model pruning, quantization, mixed-precision arithmetic, memory pooling, and layer fusion to reduce memory usage and enhance performance in transformer-based NLP tasks.

p 24 words </>

Question 9 1 pts

You are designing a DSA for real-time speech recognition (such as one in Google Home or Alexa). This accelerator must process continuous audio streams with low latency. What architectural decisions would you take? Write a few techniques you'd apply. Answer in no more than 25 words.

Edit View Insert Format Tools Table

12pt Paragraph

I'd use pipelining, stream-based processing, low-latency neural networks, real-time data buffering, and hardware-level acceleration with custom DSPs to ensure fast, continuous audio processing.

p 23 words

Question 10 1 pts

List names of the some DSAs mentioned in the paper.

Edit View Insert Format Tools Table

12pt Paragraph

The List of DSAs:
Darwin
Efficient Inference Engine
Sparse Convolutional Neural Network Accelerator
NVIDIA Deep Learning Accelerator
Google Tensor Processing Unit
CHARM
MARS
MSE

p

24 words

No new data to save. Last checked at 6:18pm

Submit Quiz