# CSE 490/590 Spring 2025 Homework 5 Solution

1. Discuss the applications in these two categories and explain how they architecturally differ from each other.

   Vector processors
   GPUs

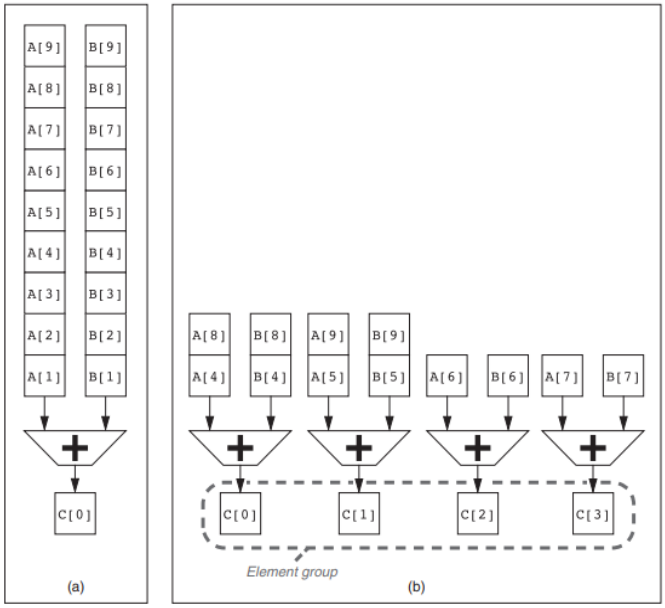
   - Similarities to vector machines:
     - Works well with data-level parallel problems
     - Scatter-gather transfers
     - Mask registers
     - Large register files

   - Differences:
     - No scalar processor
     - Uses multithreading to hide memory latency
     - Has many functional units, as opposed to a few deeply pipelined units like in a vector processor.

2. Explain Convoy and Chime

   Convoy - A set of vector instructions that could potentially execute together
   Chime - Unit of time taken to execute a convoy

3. In the figure given below, which architecture is better in terms of performance? Explain the reason for your answer. (Hint: Figures link for Chapter 4)




(b) is better because it has four add pipelines which can complete four additions per clock cycle, while (a) can complete only one addition per clock cycle.

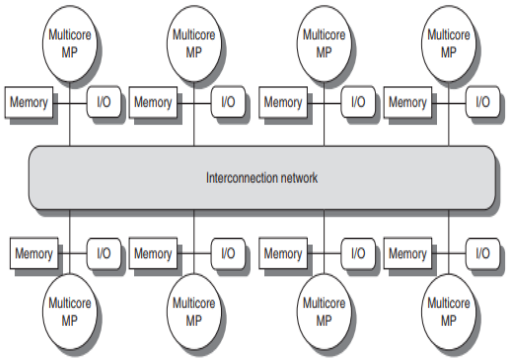
(This example describes how using multiple functional units can be used to improve the performance of a single vector add instruction)

# CSE 490/590 Spring 2025 Homework 5 Solution

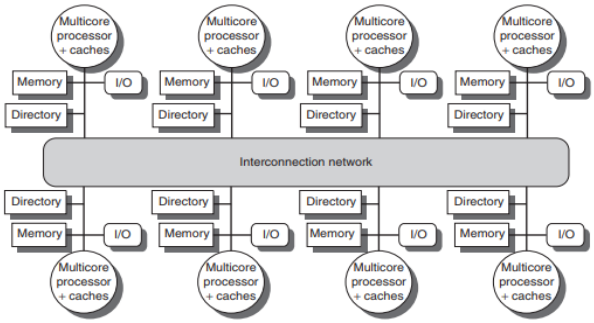
4.

a. What is cache coherence? Explain the term briefly.

b. Considering the figures below (a) and (b), ~~which is better in terms of cache coherence? Explain the reason for your answer.~~
Instead, consider the question as discuss how cache coherence works in each case.
Answer: Thread level parallelism chapter slide 5 & following and slide 19 and following.



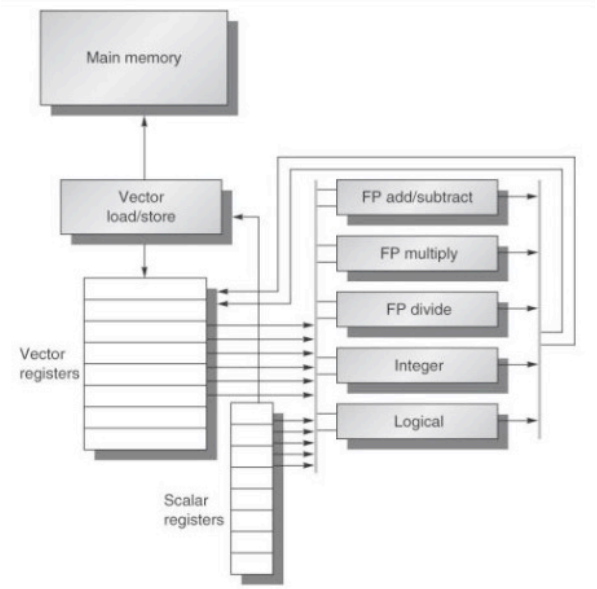

(a)




(b)

5. Considering the VMIPS structure below, explain why there are multiple read and write ports connected to the vector registers.

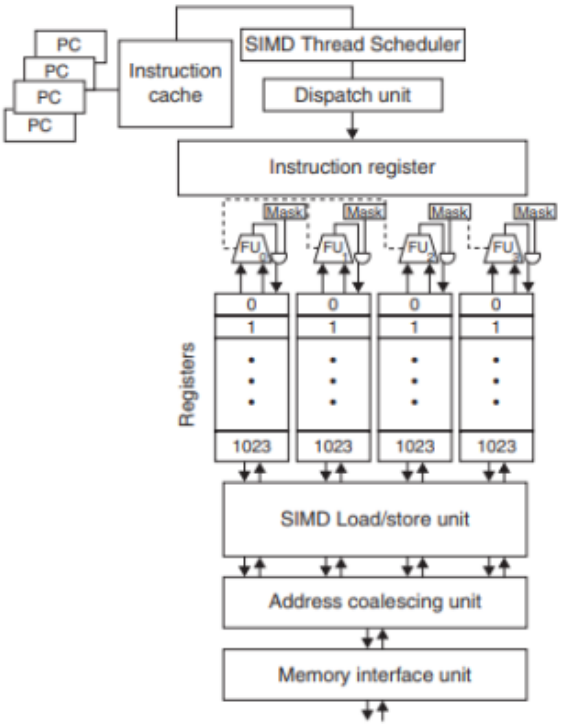
# CSE 490/590 Spring 2025 Homework 5 Solution




To allow multiple simultaneous reads/writes

To allow multiple simultaneous vector operations

6. Consider a multithreaded SIMD Processor of a GPU with four SIMD Lanes. Explain why there are multiple Program Counters.




The processor has one PC per SIMD thread to help with multithreading.

7. Explain the concept of scatter-gather operations

# CSE 490/590 Spring 2025 Homework 5 Solution

Scatter-gather operations support moving between a compressed representation (i.e., zeros are not included) and normal representation (i.e., the zeros are included) of a sparse matrix. A gather operation takes an index vector and fetches the vector whose elements are at the addresses given by adding a base address to the offsets given in the index vector. The result is a dense vector in a vector register. After these elements are operated on in dense form, the sparse vector can be stored in expanded form by a scatter store, using the same index vector. GPUs don't have separate instructions for sequential data transfers, strided data transfers, and gather-scatter data transfers. All data transfers are gather-scatter.