

**CS516: Parallelization of Programs**  
**Hands on Session - 4**

**Date: 15-Feb-2024**

**Notes:**

1. The goal of this session is to make you familiar with GPU coalesced memory accesses. Please try to submit the solutions during the lecture hour itself.

**Task-1:**

Consider a Matrix  $M$  of size  $m \times m$  and a Vector  $V$  of size  $1 \times m$ . Write a CUDA program to perform Matrix-Vector multiplication  $Z = V * M$  under the following scenarios discussed in the class.

- (a) Without coalesced memory accesses
- (b) With coalesced memory accesses.

Try to understand the performance difference between them by running for large values of  $m$ .

Note: Make sure the output of the programs is the same.

**Task-2:**

Write a program using AOS and SOA as discussed in the class. Try to understand the performance difference by running on large arrays.