



VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGINEERING)

GRADE IMPROVEMENT/MAKE UP EXAMINATIONS, AUGUST 2021

SUBJECT: PARALLEL COMPUTER ARCHITECTURE AND PROGRAMMING [CSE 3252]

REVISED CREDIT SYSTEM

(05/08/2021)

Time: 2 Hours

MAX. MARKS: 40

Instructions to Candidates:

- ❖ Answer **ANY FOUR FULL** questions.
- ❖ Missing data may be suitably assumed.

- 1A** Explain three architectural configurations of parallel computers with neat diagrams. **6M**
- 1B** What is data parallelism? How do threads in CUDA exploit data parallelism? Explain with an example. **4M**
- 2A** With help of MPI code snippet explain any 4 variants of deadlock which may occur in point to point communication. **6M**
- 2B** How will you measure the performance of a MPI application? What is the problem in benchmarking a section of code in MPI Application? How will you address this issue? Explain with example. **4M**
- 3A** Write a MPI program to compute the dot product of two vectors A and B having n number of elements. Use m processes to solve this problem where n is evenly divisible by m . Display the final summation result in process 0. Use collective communication functions wherever possible. **6M**
- $$A.B = \sum_{i=1}^n (a_i b_i)$$
- 3B** Illustrate Error Handling in MPI with a complete program. **4M**
- 4A** Consider a word W having N characters (excluding NULL), where N is divisible by 2. Write an OpenCL kernel code which accepts the word W. Every work item deals with combination of 2 characters. It checks the first character say R1 and finds its next alphabet in A-Z series say A1 (Eg: R1=J, then A1=K). Similarly, it checks the second character say R2 and finds its next alphabet in A-Z series say A2. Then, each work-item will produce the resultant string R1A1R2A2 which will be received in the host. What is the global work size set for this? Create buffers for the global arrays used in the kernel. Note: if character is 'z' then next alphabet will be 'a'. **6M**

Input:
aXYBZDEA
Output:
abXYYZBCZADEEFAZ

4B Explain the steps needed to find the time taken by the OpenCL kernel with the code segment. **4M**

5A CUDA supports several additional types of memory that can be used to increase the CGMA ratio for a kernel. Justify the statement with enough reasons. **5M**

5B A CUDA host program reads an integer matrix A of size $M \times N$ and will generate a character Matrix B of size $M \times N$ from matrix A as follows: Put every prime matrix element with a character 'P', every composite element with a character 'C', element 0 with 'Z', element 1 with 'O' and Negative element with 'N'. Write a CUDA kernel which performs the above task. Every row of output matrix B is generated in parallel. Kernel should accept only required arguments. Also write the kernel invocation with configuration parameters in the host code.

Sample Input/Output:
Enter M & N: 3 2

5M

Input Matrix A

8 2
0 -2
13 1

Output Matrix B

C P
Z N
P O

6A Write a tiled matrix multiplication CUDA kernel to multiply two $N \times N$ matrices. **5M**

6B Explain various types of CUDA program variables and their memory types, scope and lifetime. What happens when we use `__host__` and `__device__` in a function declaration? **5M**