Register number ______________________________

SRM Institute of Science and Technology
College of Engineering and Technology
School of Computing
SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamilnadu
Academic Year: 2023-24 (EVEN)
B.Tech-Computer Science & Engineering **SET - D**

Test: CLA-T2 **Date: 28.03.2024**
Course Code & Title: 18CSE419T & GPU Programming **Duration: 2 periods**
Year & Sem: III Year /VI Sem **Max. Marks: 50**

Course articulation matrix:

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO-1	3														3
CO-2		3	2												3
CO-3		3	3												3
CO-4		3	3												3
CO-5			3	1									2		3

Part – A(1*10=10 Marks)
Answer All the Questions

Q. No	Questions	Marks	BL	CO	PO	PI Code
1	Calling a kernel is typically referred to as ----------- a) Kernel Thread b) Kernel initialization c) Kernel termination d) Kernel invocation	1	2	CO 3	2 & 3	**4.2.1**
2	a) A code consisting of GRID ,which runs on GPU consisting of a set of a) 32 thread b) 32 block c) Unit block d) Thread block	1	2	CO 3	2 & 3	**4.2.1**
3	The maximum number of threads that can be launched in a specific block is a) 8 b) 32 c) 256 d) 1024	1	2	CO 3	2 & 3	**4.2.1**
4	NVDIA CUDA warp is made up of how many threads? a) 512 b) 1024 c) 312 d) 32	1	2	CO 3	2 & 3	**4.2.1**
5	For a vector addition, assume that the vector length is 2000, each thread calculates one output element and the thread block size is 512 threads. How many threads will be in the grid? a) 2000 b) 2024 c) 2048 d) 2096	1	2	CO 3	2 & 3	**4.2.1**

Register number ______________________________

6	The smallest CUDA thread block dimension is a) 8 b) 16 c) 32 d) 64	1	2	CO 3	2 & 3	**4.2.1**
7	Which of the following memory locations is common for all the SMs in a typical CUDA GPU? a) Thread-local memory b) L1 cache c) L2 cache d) Shared memory	1	3	CO 3	2 & 3	**4.2.1**
8	What is the term used for the combination of CPU and GPU in a hybrid computing system? a) Homogenous computing b) Many-core architecture c) Hardware accelerator d) Heterogenous computing	1	3	CO 3	2 & 3	**4.2.1**
9	The scope of a constant memory is a) Thread b) Block c) Warp d) Grid	1	3	CO 3	2 & 3	**4.2.1**
10	If each CUDA block can hold a maximum of 512 threads then how many CUDA blocks would be created to process 4000 vector elements a) 7 b) 8 c) 10 d) 16	1	3	CO 3	2 & 3	**4.2.1**

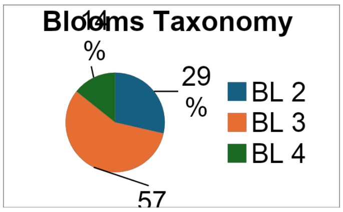
Part – B (4*4=16 marks)
Answer any four Questions

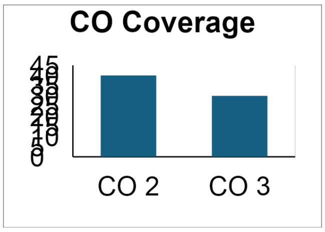
Q. No	**Question**	**Marks**	**BL**	**CO**	**PO**	**PI Code**
11	Write a CUDA code to add two numbers and explain the CUDA API functions involved in it.	4	2	CO 2	2	**4.2.1**
12	What is Texture memory? State the uses of it.	4	3	CO 3	2	**4.2.1**
13	Name two tools provided by NVIDIA to debug CUDA applications and describe its uses.	4	2	CO 3	2	**4.2.1**
14	What is zero-copy memory in GPU? What is the need for it?	4	3	CO 3	3	**4.2.1**
15	List the guidelines for optimizing the performance of the kernel through blocks and grid design.	4	3	CO 3	3	**4.2.1**

Part – C (2*12=24 marks)
Answer any Two Questions

16	Brief on the CUDA grid organization as a 3D array of blocks and state the functions involved to access a particular thread.	12	2	CO 2	2	**4.2.1**
17	Interpret a static shared memory allocation kernel through a CUDA Histogram calculation.	12	3	CO 3	3	**4.2.1**
18	Describe the role of constant memory in GPU and recognize how the values are cached and broadcasted using constant memory with an	12	3		3	**4.2.1**

	example.			CO 3		





Approved by Audit Professor/ Course Coordinator