Register Number									

Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603 110 (An Autonomous Institution, Affiliated to Anna University, Chennai)						
Computer Science and Engineering **Continuous Assessment Test -1** **Question Paper**						
Degree & Branch	B.E				**Semester**	VII
Subject Code & Name	UCS1727– GPU Computing				**Regulation:**	**2018**
Academic Year	2023-2024 ODD	**Batch**	2020-2024	**Date**	**12.09.2023**	**FN**
Time: 08:10 - 09:40 AM (90 Minutes)	**Answer All Questions**				**Maximum: 50 Marks**	

(K1: Remembering, K2: Understanding, K3: Applying, K4: Analyzing, K5: Evaluating)

CO1:	Understand GPU architecture (K2)
CO2:	Write programs using CUDA, identify issues and debug them (K3)
CO3:	Implement efficient algorithms in GPUs for common application kernels such as matrix multiplication (K3)
CO4:	Write simple programs using OpenCL (K3)
CO5:	Write an efficient parallel program for a given problem(K3).

Part – A (6×2 = 12 Marks)

1	K1	What are Warps? What is the size of a Warp defined by nVidia?	CO1	1.3.1 2.2.2 13.1.1
2	K1	What is the bandwidth of PCI-E lanes in Core-2 series GPU architecture?	CO1	1.3.1 1.4.1 2.2.2
3	K2	What are PTX instructions? Explain the fields of the PTX instruction format.	CO1	1.3.1 2.1.2
4	K1	List the differences between CPU and GPU	CO1	1.3.1 2.2.2
5	K1	How does thread divergence impact the performance of CUDA applications?	CO1	1.3.1 2.1.2 2.2.2 13.1.1 13.3.1
6	K3	Identify the CUDA API call that make sure that all previous kernel executions and memory copies have been completed.	CO2	1.4.1 2.1.2 13.3.1

Part – B (3×6 = 18 Marks)

7	K3	1) It is required to use each thread to calculate two(adjacent) elements of vector addition. Assume that variable **i** is the index of the first element to be processed by a thread. Identify the correct expression for mapping the thread/block indices to data index? Justify your Response..	CO1	1.3.1 1.4.1 2.2.2 13.1.2 13.3.1
8	K2	Explain the Memory Structure of GPU processor.	CO1	1.3.1 1.4.1 2.2.2 13.1.1
9	K3	Assume that the GPU is processing vector **A** that has 64 million data. Assume that one thread is launched for each element of a vector. If they group 1024 elements in a block and assign one block to each streaming multiprocessor, a)Estimate the number of threads, warps, blocks needed to process the data. b)Estimate the number of streaming multiprocessors needed in the system. c) Estimate the storage space requirements for storing: i)single precision floating point data. ii) Double precision floating point data	CO1	1.3.1 1.4.1 **2.2.2** 13.3.1

Part – C (2×10 = 20 Marks)

10	K2	Explain with a neat block diagram, the architecture of Core 2 series GPU processor.	CO1	1.3.1 1.4.1 2.2.2
		(OR)		
11	K2	Explain various CUDA compute levels in detail	CO1	1.3.1 1.4.1 2.2.2
12	K3	Explain the process of performing vector addition using CUDA programming. Provide a step-by-step explanation of how to parallelize the addition of two large vectors on a GPU, Discuss the advantages of using CUDA for vector addition compared to a sequential CPU approach. Develop a CUDA program to perform addition of two vectors.	CO2	1.4.1 2.1.2 13.3.1 13.3.2
		(OR)		
13	K3	Describe the key steps involved in implementing Radix Sort in GPU systems, including how you would handle parallelization and memory management Develop a CUDA program to sort list of elements in an array using Radix sort.	CO2	1.4.1 2.1.2 13.3.1 13.3.2

Prepared By **PAC Members** **Approved By**

(HOD/CSE)