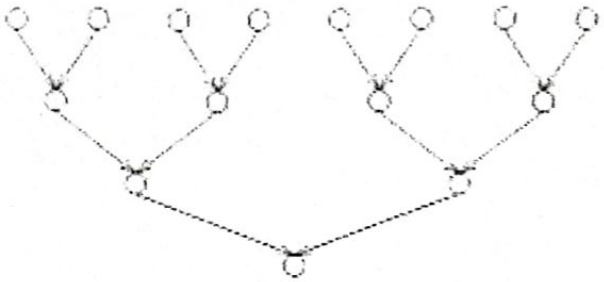
	RV COLLEGE OF ENGINEERING® Department of Computer Science and Engineering CIE-I : Question Paper	
Course : (Code)	HIGH PERFORMANCE COMPUTING ARCHITECTURES (MCE361T)	Semester : III(MTech)
Date : Dec 2024	Duration : 120 minutes	Staff : MM
Name :	USN :	Section :

Answer all questions

Sl.no	Questions	Marks	L1-L6	CO
1	Imagine you are using a web browser to navigate a website. The browser accesses data such as HTML files, CSS stylesheets, JavaScript scripts, and images to render a webpage. For the following scenarios, identify type of locality: i. If you visit a news website, the browser might reuse the cached logo image for every page on the site. ii. If you're reading an online article, the browser preloads text, images, and videos located in the next section of the article.	2	3	3
2	A program has 80% of its workload parallelizable (P=0.8) and is executed on 4 processors (N=4). What is the speedup?	2	2	2
3	Classify types of parallelism for following type of scenarios: i. Multiplying two large matrices by processing rows and columns in parallel. ii. A server handles multiple tasks concurrently, such as user authentication, database query, and serving web pages, each handled by a separate process.	2	3	3
4	Find the die yield for die that is 1.5 cm on a side assuming defect density of 0.2 per cm2 and ά=2	2	4	4
5	For a High-Performance Application following details are given, **Instruction Count**: 5×10^8 instructions **CPI**: 1.5 **Clock Speed**: 3 GHz, Calculate the total CPU time.	2	4	4

Part B

Sl.no	Questions	Marks	L1-L6	CO
1. a.	List and explain the Quantitative Principles of computer design.	6	L2	CO1
b.	Assume a **server subsystem** with the following components and MTTF values: • 8 CPUs, each rated at 2,000,000-hour MTTF • 2 power supplies, each rated at 1,500,000-hour MTTF • 4 memory modules, each rated at 500,000-hour MTTF • 2 cooling fans, each rated at 1,200,000-hour MTTF	4	L3	CO3

	• 1 network card, 3,000,000-hour MTTF Using the simplifying assumptions that the lifetimes are exponentially distributed and that failures are independent, compute the **MTTF** of the entire server subsystem.			
2	For the task graphs given below, determine the following  i. Maximum degree of concurrency. ii. Critical path length. iii. Maximum achievable speedup over one process assuming that an arbitrarily large number of processes is available. iv. The minimum number of processes needed to obtain the maximum possible speedup. v. The maximum achievable speedup if the number of processes is limited to (a)2, (b) 4, and (c) 8.	10	L3	CO3
3 a.	Differentiate between static power and dynamic power. List the different techniques to reduce dynamic power.	6	L2	CO2
b.	Consider the Frequency of FP operations to be 25%, Average CPI of FP operations to be 4.0, Average CPI of other instructions is 1.33, Frequency of FPSQR to be 2%, CPI of FPSQR as 20. Assume that the two design alternatives are to decrease the CPI of FPSQR to 2 or to decrease the average CPI of all FP operations to 2.5. Compare these two design alternative using the processor performance equation.	4	L4	CO4
4	Classify the different types of parallelism based on Flynn's taxonomy, providing suitable examples for each category.	10	L1	CO1
5 a.	Explain fork-join model used in OpenMP to implement parallelism. Write OpenMP C code to print "Hello World" with thread id from each thread.	10	L4	CO4

CO1 : Explore the fundamental concepts of parallel computer architecture
CO2 : Analyze the performance of parallel programming
CO3 : Design parallel computing constructs for solving complex problems.
CO4 : Demonstrate parallel computing concepts for suitable applications

	L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4
Marks	10	12	20	18	-	-	10	12	20	18

	RV COLLEGE OF ENGINEERING® Department of Computer Science and Engineering CIE-I : Scheme	
Course : (Code)	**HIGH PERFORMANCE COMPUTING ARCHITECTURES (MCE361T)**	**Semester : III(MTech)**

Date : Dec 2024	**Duration :** 120 minutes	**Staff : MM**	
Name :	**USN** :	**Section :**	

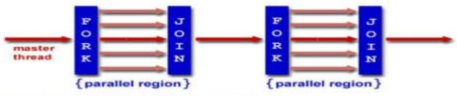
Answer all questions

Sl.no	Questions	Marks	L1-L6	CO
1	Temporal locality Spatial locality 2*1M	2	3	3
2	The speedup is **2.5**., Correct answer 1M Formula 1M	2	2	2
3	i. DLP ii. TLP 2*1M	2	3	3
4	The die yield is approximately **47.57%**. Correct answer 1M Formula 1M	2	4	4
5	The total CPU time is **0.25 seconds**. Correct answer 1M Formula 1M	2	4	4

Part B

Sl.no	Questions	Marks	L1-L6	CO
1. a.	Take Advantage of Parallelism , Principle of Locality , Amdahl's Law Focus on Common cases. Any 3*2	6	L2	CO1
b.	$\lambda_{\text{Total}} = \lambda_{\text{CPUs}} + \lambda_{\text{PSs}} + \lambda_{\text{Mems}} + \lambda_{\text{Fans}} + \lambda_{\text{Net}}$ Substitute the calculated values: $\lambda_{\text{Total}} = \frac{1}{250,000} + \frac{1}{750,000} + \frac{1}{125,000} + \frac{1}{600,000} + \frac{1}{3,000,000}$ Convert the fractions to a common denominator or decimal form for summation: $\lambda_{\text{Total}} = 0.000004 + 0.00000133 + 0.000008 + 0.00000167 + 0.00000033$ $\lambda_{\text{Total}} = 0.00001533$ $\text{MTTF}_{\text{System}} = \frac{1}{\lambda_{\text{Total}}}$ $\text{MTTF}_{\text{System}} = \frac{1}{0.00001533} \approx 65,250 \text{ hours}$	4	L3	CO3
2	i. 8 ii. 4 iii. 15 / 4 = 3.75 iv. 8 v. 8 -> 15/4=3.75, 4 -> 15/5=3, 2 -> 15/8 = 1.875	10	L3	CO3
3 a.	Differentiate between static power and dynamic power. List the different techniques to reduce dynamic power.	6	L2	CO2
b.	• **Design Alternative 2** (decreasing the average CPI of all FP operations to 2.5) provides a greater performance improvement, offering a speedup of about 1.43 compared to the current system, compared to **Design Alternative 1**	4	L4	CO4

	(decreasing the CPI of FPSQR to 2), which offers a speedup of about 1.18. Thus, **Design Alternative 2** is the better option for improving processor performance.			
4	SISD, SIMD, MISD, MIMD: with example 1+9M	10	L1	CO1

5 a.	`#include <omp.h>`
	`#include <stdio.h>`
	`#include <stdlib.h>`
	
	`int main (int argc, char *argv[])`
	`{`
	`int nthreads, tid;`
	
	`/* Fork a team of threads giving them their own copies of variables */`
	`#pragma omp parallel private(nthreads, tid)`
	`{`
	
	`/* Obtain thread number */`
	`tid = omp_get_thread_num();`
	`printf("Hello World from thread = %d\n", tid);`
	
	`/* Only master thread does this */`
	`if (tid == 0)`
	`{`
	`nthreads = omp_get_num_threads();`
	`printf("Number of threads = %d\n", nthreads);`
	`}`
	
	`} /* All threads join master thread and disband */`
	
	`}`
	• **Shared Memory, Thread Based Parallelism**
	• **Explicit Parallelism**
	• **Fork - Join Model**
	 
	
	• **Compiler Directive Based**
	• **Nested Parallelism Support**
	• **Dynamic Thread**
	• **Memory Model : Flush often**
	5+5M
		10	L4	CO4

CO1 : Explore the fundamental concepts of parallel computer architecture
CO2 : Analyze the performance of parallel programming
CO3 : Design parallel computing constructs for solving complex problems.
CO4 : Demonstrate parallel computing concepts for suitable applications

	L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4
Marks	10	12	20	18	-	-	10	12	20	18

	RV COLLEGE OF ENGINEERING® **Department of Computer Science and Engineering** **CIE-II : Question Paper**	
Course : (Code)	**HIGH PERFORMANCE COMPUTING ARCHITECTURES (MCE361T)**	**Semester : III(MTech)**
Date : Dec 2024	**Duration :** 120 minutes	**Staff : MM**
Name :	**USN :**	**Section :**

Answer all questions

Sl.no	Questions	Marks	L1-L6	CO
1	An NVIDIA GPU has 32,768 registers, with 16 SIMD lanes per processor. How many registers are allocated per lane?	2	3	3
2	A program has 60% of its workload parallelizable (P=0.8) and is executed on 4 processors (N=4). What is the speedup?	2	2	2
3	A database query decomposition has a critical path length of 30 time units and a maximum degree of concurrency of 6. How many processors are required to achieve the shortest parallel execution time?	2	3	3
4	What is the relation between Thread, Block and Grid in CUDA programming.	2	4	4
5	If a message of 10 integers (4 bytes each) is sent using MPI_Send, how much data (in bytes) is transferred?	2	4	4

Part B

Sl.no	Questions	Marks	L1-L6	CO
1. a.	Discuss the differences between recursive decomposition and data decomposition techniques. Use the matrix-vector multiplication problem as an example to demonstrate how each technique is applied.	6	L2	CO1
b.	In a task dependency graph, Task A takes 5 time units, Task B takes 10 time units, and Task C takes 15 time units. Task A and Task B must complete before Task C starts. What is the critical path length?	4	L3	CO3
2	Explain the message-passing programming paradigm in parallel computing. Discuss the key principles, the role of send and receive operations	10	L3	CO3
3 a.	Discus the characteristics of a parallel task.	6	L2	CO2
b.	Compare centralized and distributed dynamic mapping techniques. Highlight their advantages and limitations with suitable examples,	4	L4	CO4
4	Describe the CUDA programming model for GPUs. Explain the role of threads, blocks, and grids in executing parallel programs on NVIDIA GPUs, and how the hardware supports efficient memory access.	10	L1	CO1
5 a.	Discuss the differences between traditional vector processors and NVIDIA GPUs. Highlight the architectural innovations introduced by the Fermi GPU architecture and how they address the challenges of parallel computing.	10	L4	CO4

CO1 : Explore the fundamental concepts of parallel computer architecture
CO2 : Analyze the performance of parallel programming
CO3 : Design parallel computing constructs for solving complex problems.
CO4 : Demonstrate parallel computing concepts for suitable applications

	L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4
Marks	10	12	20	18	-	-	10	12	20	18

	RV COLLEGE OF ENGINEERING® Department of Computer Science and Engineering CIE-III : Question Paper	
Course : (Code)	HIGH PERFORMANCE COMPUTING ARCHITECTURES (MCE361T)	Semester : III(MTech)
Date : March 2025	Duration : 120 minutes	Staff : MM
Name :	USN :	Section :

Answer all questions

Sl.no	Questions	Marks	L1-L6	CO
1	The OpenACC compute directives are __, _____,_____ and _____.	2	3	3

2	Consider the following MPI code:
	if (rank == 0) {
	MPI_Send(&data, 1, MPI_INT, 1, 0, MPI_COMM_WORLD); MPI_Recv(&data, 1, MPI_INT, 1, 0, MPI_COMM_WORLD, &status); } else if (rank == 1) {
	MPI_Send(&data, 1, MPI_INT, 0, 0, MPI_COMM_WORLD); MPI_Recv(&data, 1, MPI_INT, 0, 0, MPI_COMM_WORLD, &status);
	}
	Does this code lead to deadlock? Why?
		2	2	2

3	A program sends a 2D matrix of size 100×100 with MPI_Send. Each element is a double (8 bytes). How much data (in bytes) is transferred in a single MPI_Send call?How much data would be sent if MPI_Scatter is used with 4 processes	2	3	3
4	What is difference between OpenACC and CUDA?	2	4	4
5	An FPGA has **10,000** logic elements, and each logic element consumes **1.2 mW** of power. What is the total power consumption of the FPGA?	2	4	4

Sl.no	Part B Questions	Marks	L1-L6	CO
1. a.	Discuss the following system called used in MPI: MPI_Bcast(), MPI_Scatter(), MPI-Gather()	6	L2	CO1
b.	Write a noir on Groups and Communicators	4	L3	CO
2	Discuss the architecture of an FPGA.	10	L3	CO
3 a.	Discus the characteristics of a parallel task.	6	L2	CO
b.	Compare centralized and distributed dynamic mapping techniques. Highlight their advantages and limitations with suitable examples,	4	L4	CO
4	Explain the key differences between the parallel and kernels constructs in OpenACC, providing examples of when to use each.	10	L1	CO
5	Compare and contrast the architectures of CPUs, GPUs, and FPGA focusing on their suitability for machine learning tasks.	10	L4	CO

CO1 : Explore the fundamental concepts of parallel computer architecture
CO2 : Analyze the performance of parallel programming
CO3 : Design parallel computing constructs for solving complex problems.
CO4 : Demonstrate parallel computing concepts for suitable applications

	L1	L2	L3	L4	L5	L6	CO1	CO2	CO3	CO4
Marks	10	12	20	18	-	-	10	12	20	18