U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

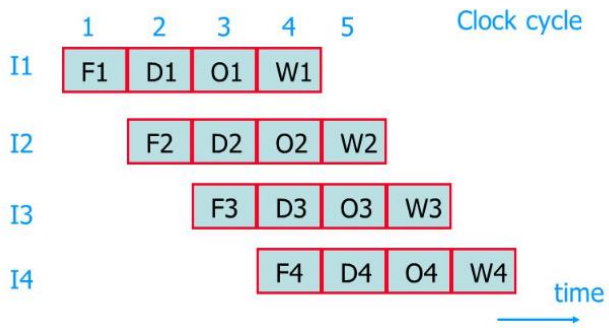
Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

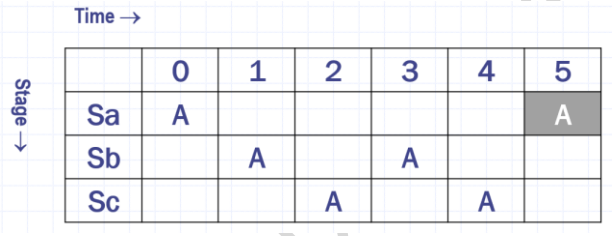
Programme: B.E. **Semester: VI**
Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**
Course Code: 24AM6PCPAP **Max Marks: 100**
Course: Parallel Architectures and Programming

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Illustrate Von-Neumann's computer architecture with a neat diagram.	*CO1*	*PO1*	**06**
	b)	Differentiate Flynn's and Feng's parallel architecture classification in detail. Justify which classification is best suitable for parallel programming.	*CO1*	*PO2*	**10**
	c)	Consider computation task where 70% of the program can be parallelized. Running this computation on a system with 8 processors. If it scale the size of the problem so that the parallelizable portion dominates, calculate the speedup according to Gustafson's Law.	*CO2*	*PO2*	**04**
		OR			
2	a)	Calculate the time taken and latency (in clock cycles) to execute the instructions for the given stage diagram in both non-pipelined architectures. How do you handle the same in a pipelined architecture? Provide one such stage diagram solution if a data hazard occurs at instruction 2. 	*CO1*	*PO1*	**06**
	b)	A program consists of 70% of parallelizable code fraction *f*. Without any loss of generality, the workload is also fixed for all the time. Asses the speedup when using 2, 16, 32, 128, 256 processors. Plot the results on a cartesian plane and provide your critic on the same.	*CO1*	*PO2*	**10**

		c)	Justify why parallel processing is an important aspect in instruction execution cycle.	*CO2*	*PO2*	**04**
			UNIT - II			
	3	a)	A computer has an 8 GByte memory with 64 bit word sizes. Each block of memory stores 16 words. The computer has a direct-mapped cache of 128 blocks. The computer uses word level addressing. What is the address format? If we change the cache to a 4- way set associative cache, what is the new address format?	*CO2*	*PO2*	**04**
		b)	Analyze the following code and determine temporal and spatial locality element. Justify which locality is best. *var sum = 0;* *for (i = 0; i < n; i++){* *for(j=0; j < m ; j++){* *sum += a[i][j];* *}* *}* *return sum;*	*CO2*	*PO2*	**06**
		c)	Draw the I/O subsystem diagram and illustrate the interface functionality of the same.	*CO1*	*PO12*	**10**
			OR			
	4	a)	A computer has a single cache (off-chip) with a 2 ns hit time and a 98% hit rate. Main memory has a 40 ns access time. What is the computer's effective access time? If we add an on-chip cache with a .5 ns hit time and a 94% hit rate, what is the computer's effective access time? How much of a speedup does the on-chip cache give the computer?	*CO2*	*PO2*	**05**
		b)	Consider a system with two processors, P1 and P2, each with its own cache. Both caches use the MSI protocol. Determine the read and write operations for data consistency across multiple caches in a multiprocessor system.	*CO2*	*PO2*	**07**
		c)	Differentiate Temporal, Locality and Sequential localities. Also identify what type of locality is applicable for the following. i. computed values and reuse them for repeated accesses. ii. access a submatrix within a 2D array. iii. iterate over an entire array	*CO2*	*PO1*	**08**
			UNIT - III			
	5	a)	Define speedup in a pipeline. Calculate and compare the speedup ratio for a 6-segment non-pipeline and pipeline execution systems with a clock cycle of 40ns which requires to execute 100 tasks in sequence.	*CO2*	*PO2*	**08**
		b)	Analyze the following vector operations and justify how chime and convey vector processing mechanisms can be applied for the	*CO2*	*PO2*	**07**

			following instruction set? Represent cycles and pipelining stages accordingly. *V1 = Load(A)* *V2 = Load(B)* *V3 = V1 + V2* *V4 = V3 * scalar* *A and B are Array elements.*			
		c)	For the following vector operations provide their corresponding vector machine instructions i. Vector + Vector ii. Vector + Scalar iii. Vector * Vector iv. Scalar * Vector v. Vector - Scalar	*CO2*	*PO2*	**05**
			OR			
	6	a)	Design and explain a non-linear 3-staged pipelined for each sequence flow from the given reservation table. 	*CO2*	*PO2*	**08**
		b)	Write a python program using regular functions to perform matrix transpose using vector striding technique.	*CO2*	*PO2*	**07**
		c)	Provide a python code to convert a scalar operation to a vector operation.	*CO2*	*PO2*	**05**
			UNIT - IV			
	7	a)	With a neat diagram, illustrate the working of GPU architecture	*CO3*	*PO2*	**08**
		b)	Differentiate parallelism in CPU and GPU	*CO3*	*PO3*	**06**
		c)	Justify with a neat diagram why GPU do not implement data cache and control flow units with respect to hardware architecture?	*CO3*	*PO2*	**06**
			OR			
	8	a)	Illustrate memory management in GPU.	*CO3*	*PO2*	**08**
		b)	Illustrate data flow mechanism from CPU to GPU and vice versa. Signify the importance of such architecture.	*CO3*	*PO3*	**06**
		c)	Compare and contrast CPU and GPU	*CO3*	*PO2*	**06**
			UNIT - V			
	9	a)	Write a python code using CUDA concepts by adding two vectors using GPU parallelism.	*CO3*	*PO1*	**10**

		b)	Compare SMIT and SIMD in terms of control flow, data dependency and memory access.	*CO3*	*PO2*	**10**
			OR			
	10	a)	List four applications of CUDA. Also write on CUDA working mechanism.	*CO3*	*PO1*	**10**
		b)	Illustrate how CUDA computations are divided into threads and blocks.	*CO3*	*PO2*	**10**

****