



**D. Y. Patil College of Engineering, Akurdi, Pune 411044**  
**Department of Computer Engineering**

**Prelim Exam**

Date: 29/12/2020

Class : BE Computer  
Academic Year : 2020-21

Div: A & B  
Sem : I

Subject : High Performance Computing  
Exam Date: 31/12/2020

<i>Q. No.</i>	<i>Question Description</i>	<i>Options</i>	<i>Correct Answer</i>	<i>Marks</i>	<i>CO</i>	<i>PO</i>	<i>PSO</i>	<i>BTL</i>
1	Which of the following is the type of parallelism?	a. Bit level parallelism b. Instruction level parallelism c. Loop level parallelism d. All of the above	D	1	1	1,1 2	1	2
2	Which of the parallelism is used by VLIW	a. Bit level parallelism b. Instruction level parallelism c. Loop level parallelism d. Task level Parallelism	B	1	1	1,1 2	1	2
3	Tendency of a software process to access information items whose addresses are near one another known as	a. Spatial Locality b. Temporal locality c. Permanent Locality d. Sequential Locality	a	1	1	1	1	1
4	Parallel Computers are classified based on Flynn's taxonomy which among the following options does not come under this	a. SISD b. SIMD c. MIMD d. SIPD	d	1	1	1,1 2	1	1
5	Which among the following is the popular multistage network	a. Hypercube b. Omega	b	1	1	1	1	2

		<b>c. Gamma</b> <b>d. K-D Mesh</b>						
6	The multicore architecture that consists of dedicated application specific processor cores that would target the issue of running variety of applications to be executed on a computer.	<b>a. Homogeneous core architecture.</b> <b>b. Heterogeneous core architecture.</b> <b>c. Polaris core architecture</b> <b>d. None of the above</b>	<b>b</b>	1	1	<b>1</b>	<b>1</b>	<b>3</b>
7	Decomposition of computation into a small number of large task is	<b>a. Fine grained granularity</b> <b>b. course grained granularity</b> <b>c. coarse grained granularity</b> <b>d. task grained granularity</b>	<b>C</b>	1	2	<b>1</b>	<b>3</b>	<b>1</b>
8	Which among the following is the type of decomposition	<b>a. Data-decomposition</b> <b>b. Hybrid decomposition</b> <b>c. Speculative decomposition</b> <b>d. All of the above</b>	<b>D</b>	1	2	<b>1,1 2</b>	<b>3</b>	<b>2</b>
9	The 15-puzzle problem uses which type of decomposition	<b>a. Data decomposition</b> <b>b. Exploratory decomposition</b> <b>c. Speculative decomposition</b> <b>d. Recursive decomposition</b>	<b>B</b>	1	2	<b>1,4 ,12</b>	<b>3</b>	<b>2</b>

10	An interaction pattern is considered to be _____if it has some structure that can be exploited for efficient implementation	<b>a. Structured interaction</b> <b>b. unstructured interaction</b> <b>c. Regular interaction</b> <b>d. Irregular interaction</b>	<b>C</b>	1	2	<b>1,1 2</b>	<b>3</b>	<b>2</b>
11	The mapping in which is tasks are distributed to processes during execution is called as____	<b>a. Dynamic mapping</b> <b>b. Static mapping</b> <b>c. Pre-execution mapping</b> <b>d. In-process mapping</b>	<b>a</b>	1	2	<b>1</b>	<b>1</b>	<b>1</b>
12	The parallel algorithm model in which mapping of tasks is done dynamically where pointer to tasks is stored in physically shared list/priority queue/hash table/tree is called	<b>a. The data parallel model</b> <b>b. Producer consumer model</b> <b>c. The task graph model</b> <b>d. Work pool model</b>	<b>d</b>	1	2	<b>1,2</b>	<b>1</b>	<b>2</b>
13	The world's first GPU is marketed by NVIDIA in 1999 is	<b>a. GeForce 356</b> <b>b. GeForce 256</b> <b>c. GeForce 3800</b> <b>d. GeForce 956</b>	<b>B</b>	1	6	<b>5</b>	<b>3</b>	<b>1</b>
14	The operation in which data from all processes are combined at a single destination process is	<b>a. All to one reduction</b> <b>b. All to all reduction</b> <b>c. one to all reduction</b> <b>d. None of the above</b>	<b>A</b>	1	3	<b>1</b>	<b>1</b>	<b>2</b>
15	In scatter operation a single node sends a unique message to every node is also called as	<b>a. One-to-one personalized communication</b> <b>b. One-to-all broadcast</b>	<b>C</b>	1	3	<b>1</b>	<b>1</b>	<b>2</b>

		<p>communication</p> <p>c. One-to-all personalized communication</p> <p>d. all-to-all personalized communication</p>						
16	Single port communication node can communicate on all the channels connected to it and provides apparent speedup	<p>a. True</p> <p>b. False</p>	<b>B</b>	1	3	<b>1</b>	<b>1</b>	<b>1</b>
17	Symmetric multiprocessors architecture are sometimes known as	<p>a. <b>Uniform memory access</b></p> <p>b. <b>Static memory access</b></p> <p>c. <b>Variable memory access</b></p> <p>d. <b>All of the above</b></p>	<b>A</b>	1	3	<b>1</b>	<b>1</b>	<b>1</b>
18	Heuristic is way of trying	<p>a. <b>To discover something or an idea embedded in a program</b></p> <p>b. <b>To search and measure how far a node in a search tree seems to be from a goal</b></p> <p>c. <b>To compare two nodes in a search tree to see if one is better than another</b></p> <p>d. <b>All of the mentioned</b></p>	<b>a</b>	1	4	<b>1,2</b>	<b>3</b>	<b>2</b>
19	A * algorithm is based on	<p>a. <b>Breadth-First search</b></p> <p>b. <b>Depth-first Search</b></p> <p>c. <b>Best first search</b></p> <p>d. <b>Hill climbing</b></p>	<b>C</b>	1	5	<b>1,2</b>	<b>1</b>	<b>2</b>

20	Best – First search can be implemented using the following data structure	<b>a. Queue</b> <b>b. Stack</b> <b>c. Priority Queue</b> <b>d. Circular Queue</b>	<b>C</b>	1	5	1,2	1	1
21	_____is a measure of the fraction of time for which a processing element is usefully employed	<b>a. Scalability</b> <b>b. Efficiency</b> <b>c. Speedup</b> <b>d. isoefficiency</b>	<b>B</b>	1	5	1,2	1	2
22	The ____of a parallel system is a measure of its capacity to increase speedup in proportion to the number of processing elements	<b>A. speedup</b> <b>B. Cost</b> <b>C. Efficiency</b> <b>D. Scalability</b>	<b>D</b>	1	3	1,1 2	1	2
23	____helps us determine the best algorithm/architecture combination for a particular problem without explicitly analyzing all possible combinations under all possible co	<b>a. Isoefficiency Metric of scalability</b> <b>b. Efficiency matric of scalability</b> <b>c. Cost metric of scalability</b> <b>d. None of the above</b>	<b>A</b>	1	3	1,3	1	2
24	It is defined as a ratio of the time taken to solve a problem on a single processing element to the time computer with p identical processing elements	<b>a. Total parallel overhead</b> <b>b. Efficiency</b> <b>c. Cost</b> <b>d. speedup</b>	<b>D</b>	1	3	1,1 2	1	1
25	In Practice a speedup greater than p is sometimes observed. It is called as _____	<b>a. scalability effect</b> <b>b. superscalar effect</b>	<b>C</b>	1	3	1,2, 12	1	1

		<b>c. super linearity effect</b> <b>d. speedup effect</b>						
26	Odd-even transposition sort is not cost -optimal, because time product is	<b>a. <math>\theta(n^2)</math></b> <b>b. <math>\theta(n \log n)</math></b> <b>c. <math>O(n^3)</math></b> <b>d. <math>O(n + \log n)</math></b>	<b>A</b>	1	5	1,2,5	3	3
27	The quicksort algorithm, which has an average complexity of	<b>a. <math>O(n^3)</math></b> <b>b. <math>O(n + \log n)</math></b> <b>c. <math>\theta(n \log n)</math></b> <b>d. <math>\theta(n^2)</math></b>	<b>C</b>	1	5	1,2,5	1	3
28	Parallel code executes in many concurrent Device (GPU) threads across multiple parallel processing elements, called	<b>a. Synchronising multiprocessor</b> <b>b. Streaming multiprocessor</b> <b>c. Scalable multiprocessor</b> <b>d. Summative multiprocessor</b>	<b>B</b>	1	6	1,2,12	1	2
29	_____ partitions the vertices among different processes and has each process compute the single-source shortest path for all vertices assigned to it	<b>a. Source parallel formulation</b> <b>b. Single partitioned formulation</b> <b>c. Source partitioned formulation</b> <b>d. Shortest path partitioned formulation</b>	<b>C</b>	1	5	1,2,12	3	2
30	A processor, assigned with a thread block that executes	<b>a. Multithreaded DIMS</b>	<b>B</b>	1	2	1	1	1

	code, which we usually call a	<p><b>processor</b></p> <p><b>b. Multithreaded SIMD processor</b></p> <p><b>c. Multithreaded queue</b></p> <p><b>d. Multithreaded stack</b></p>						
31	Processor of system, which can read/write GPU memory, is known as	<p><b>a. Server</b></p> <p><b>b. Kernel</b></p> <p><b>c. Guest</b></p> <p><b>d. Host</b></p>	<b>D</b>	1	6	<b>1</b>	<b>1</b>	<b>1</b>
32	CUDA stands for	<p><b>a. Compute uniform device architecture</b></p> <p><b>b. Computing universal device architecture</b></p> <p><b>c. Computer unicode device architecture</b></p> <p><b>d. Compute unified device architecture</b></p>	<b>D</b>	1	6	<b>1,2,5</b>	<b>2</b>	<b>1</b>
33	The device that are being used primarily for database, file server and mostly for web application are known as	<p><b>a. Servers</b></p> <p><b>b. Desktops</b></p> <p><b>c. Tablets</b></p> <p><b>d. Supercomputers</b></p>	<b>A</b>	1	1	<b>1</b>	<b>1</b>	<b>1</b>
34	GPU are designed for running a large number of complex tasks	<p><b>a. True</b></p> <p><b>b. False</b></p>	<b>B</b>	1	6	<b>1,2</b>	<b>1</b>	<b>1</b>
35	The parallel algorithm design contains a number of processes where one process may send the identical data to all other processes is called as	<p>a. All to one broadcast</p> <p>b. All to all broadcast</p> <p>c. One to all broadcast</p> <p><b>d. None of these</b></p>	<b>C</b>	1	3	<b>1</b>	<b>1</b>	<b>2</b>
36	The efficient utilization can be done by devising a	<p>a. Recursive doubling</p> <p>b. Recursive</p>		1	3	<b>1</b>	<b>1</b>	<b>1</b>

	broadcasting algorithm with the method known as	c. Scatter and Gather <b>d. None of these</b>	a					
37	The balanced tree is mapped neutrally from the hypercube algorithm for one-to-all broadcast where intermediate are the _____ and each leaf nodes are the _____	a. switching nodes, processing nodes <b>b. processing nodes, switching nodes</b>	a	1	3	<b>1,1</b> <b>2</b>	<b>1</b>	<b>2</b>
38	Finding prefix-sum operation is also called as scan operation	a. True <b>b. False</b>	a	1	3	<b>1,1</b> <b>2</b>	<b>1</b>	<b>1</b>
39	All to all personalized communication is also called as	a. Scan operation b. Total exchange method c. None of these	B	1	3	<b>1,1</b> <b>2</b>	<b>1</b>	<b>2</b>
40	On which network broadcast and reduction operations performed in two steps: 1. Operations along with row 2. Operations along with column	a. Ring b. Hypercube c. Linear array <b>d. Mesh</b>	d	1	3	<b>1,1</b> <b>2</b>	<b>1</b>	<b>2</b>
41	Gather operation is also called as all to one reduction	a. True <b>b. False</b>	b	1	3	<b>1,8</b>	<b>1</b>	<b>1</b>
42	The method which is used in various parallel algorithm like Fourier transform, matrix transpose, some parallel database join operations is called as	a. All-to-all personalized communication b. All-to-all Broadcast c. Total exchange method d. Both a & c	d	1	3	<b>1,1</b> <b>2</b>	<b>1</b>	<b>1</b>
43	Consider a sequence in which numbers are originally arranged <2,4,5,6,1>, then sequence of Prefix sum will be	a. <2,6,11,17,18 > b. <6,15,21,22 > c. None of these	a	1	3	<b>4</b>	<b>2</b>	<b>3</b>
44	Select the parameters on which the parallel runtime of a program depends.	A. nput size B. umber of processors C.	<b>D</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>4</b>



		ommunication parameters of the machine D. ll of the above						
45	The time that elapses from the moment the first processor starts to the moment the last processor finishes execution is called as _____.	A. Serial runtime B. Parallel runtime C. Overhead runtime D. Excess runtime	<b>B</b>	<b>1</b>	<b>4</b>	<b>4</b>	<b>3</b>	<b>4</b>
46	Select how the overhead function ( $T_o$ ) is calculated.	A. $T_o = T_P - T_S$ B. $T_o = p*n T_P - T_S$ C. $T_o = p T_P - T_S$ D. $T_o = T_P - pT_S$	<b>C</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>
47	The parallel time for odd-even sort (efficient parallelization of bubble sort) is 50 seconds. The serial time for bubble sort is 175 seconds. Evaluate the speedup of bubble sort.	A. 3.75 B. 3.5 C. 0.33 D. 0.26	<b>B</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>
48	Consider the problem of adding $n$ numbers by using $n$ processing elements. The serial time taken is $\Theta(n)$ and parallel time is $\Theta(\log n)$ . Evaluate the efficiency.	A. $E = \Theta(n / \log n)$ B. $E = \Theta(n \log n)$ C. $E = \Theta(\log n / n)$ D. $E = \Theta(1 / \log n)$	<b>D</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>
49	What will be the efficiency of cost optimal parallel systems?	A. $E = O(n)$ . B. $E = O(1)$ .	<b>B</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>3</b>

		<p>C. <math>E = O(p)</math>.</p> <p>D. <math>E = O(n \log n)</math>.</p>						
50	Which law states that the maximum speedup of a parallel program is limited by the sequential fraction of the initial sequential program?	<p>A. Amdahl's Law</p> <p>B. Flynn's Law</p> <p>C. Moore's Law</p> <p>D. Van Neumann's Law</p>	<b>A</b>	<b>1</b>	<b>4</b>	<b>3</b>	<b>3</b>	<b>1</b>
51	<p>Arrange the steps for the Matrix-Vector 2-D partitioning</p> <p>i) result vector is computed by performing an all-to-one reduction along the columns.</p> <p>ii) Alignment of the vector <math>x</math> along the principal diagonal of the matrix.</p> <p>iii) Copy the vector elements from each diagonal process to all the processes in the corresponding column using <math>n</math> simultaneous broadcasts among all processors in the column.</p>	<p>A. i, ii, iii</p> <p>B. ii, iii, i</p> <p>C. iii, i, ii</p> <p>D. ii, i, iii</p>	<b>B</b>	<b>1</b>	<b>4</b>	<b>3</b>	<b>3</b>	<b>1</b>
52	<p>Arrange the communication sequence in Matrix-Vector 2-D partitioning:</p> <p>i) all-to-one reduction in each row</p> <p>ii) one-to-all broadcast of each vector element among the <math>n</math> processes of each column</p> <p>iii) one-to-one</p>	<p>A. i, ii, iii</p> <p>B. ii, iii, i</p> <p>C. iii, ii, i</p> <p>D. ii, i, iii</p>	<b>C</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>4</b>

	communication to align the vector along the main diagonal							
53	Parallel time in Rowwise 1-D Partitioning of Matrix-Vector Multiplication where p=n is ____.	A. $\Theta(1)$ B. $\Theta(n \log n)$ C. $\Theta(n^2)$ D. $\Theta(n)$	<b>D</b>	<b>1</b>	<b>4</b>	<b>4</b>	<b>3</b>	<b>4</b>
54	NVIDIA thought that 'unifying theme' of every forms of parallelism is the	<b>a. CDA thread</b> <b>b. PTA thread</b> <b>c. CUDA thread</b> <b>d. CUD thread</b>	<b>c</b>	<b>1</b>	<b>6</b>	<b>1,2, 12</b>	<b>1</b>	<b>2</b>
55	Thread being blocked altogether and being executed in sets of 32 threads, called a	<b>a. Thread block</b> <b>b. 32 thread</b> <b>c. 32 block</b> <b>d. Unit block</b>	<b>a</b>	<b>1</b>	<b>6</b>	<b>1,2, 12</b>	<b>1</b>	<b>2</b>
56	Length of a vector operation in a real program is often	<b>a. Known</b> <b>b. Unknown</b> <b>c. Visible</b> <b>d. Invisible</b>	<b>a</b>	<b>1</b>	<b>6</b>	<b>1,2, 12, 6</b>	<b>1</b>	<b>3</b>
57	A code, known as grid which runs on a GPU consisting of a set of	<b>a. 32 thread</b> <b>b. Unit block</b> <b>c. 32 block</b> <b>d. Thread block</b>	<b>d</b>	<b>1</b>	<b>6</b>	<b>1,1 2,5</b>	<b>1</b>	<b>1</b>
58	NVIDIA unveiled the industrys first DirectX 10 GPU is____	<b>a. GTX 1050</b>	<b>b</b>	<b>1</b>	<b>6</b>	<b>1,1 2,5</b>	<b>1</b>	<b>1</b>

		<b>b. GeForce 8800 GTX</b> <b>c. GeForce GTX 1080</b> <b>d. GTX 1060</b>				<b>1</b>		
59	The number of instructions being executed defines the	<b>a. Instruction count</b> <b>b. Hit time</b> <b>c. Clock rate</b> <b>d. All above</b>	<b>A</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>
60	In CUDA Programming kernel is launch using which pair of brackets?	<b>a. &lt;&lt;&lt;&gt;&gt;&gt;</b> <b>b. {{{}}}</b> <b>c. ((()))</b> <b>d. [[][]]</b>	<b>d</b>	<b>1</b>	<b>6</b>	<b>1,2,12,5</b>	<b>3</b>	<b>2</b>
61	In CUDA programming the transfer of data between host and device special function used is ____	<b>a. Memcopy()</b> <b>b. Memorycpy()</b> <b>c. cudaMemcpy()</b> <b>d. cudaMemorycpy()</b>	<b>c</b>	<b>1</b>	<b>6</b>	<b>1,2,12,5</b>	<b>1</b>	<b>1</b>
62	Streaming multiprocessor in CUDA, divides the thread in a block is called as____	<b>a. WRAP</b> <b>b. Packet</b> <b>c. Grid</b> <b>d. Thread block</b>	<b>a</b>	<b>1</b>	<b>6</b>	<b>1,12,5</b>	<b>1</b>	<b>2</b>
63	Sources of overheads in parallel program are	<b>a. Idling</b> <b>b. Interprocess communication</b> <b>c. Excess computation</b> <b>d. All of the above</b>	<b>d</b>	<b>1</b>	<b>3</b>	<b>1,12,2</b>	<b>1</b>	<b>2</b>
64	What are the sources of overhead in parallel	<b>A. Interprocess interaction</b>	<b>D</b>	<b>1</b>	<b>4</b>	<b>4</b>	<b>3</b>	<b>4</b>

	programs?	B. Idling C. Excess computation D. All of the above						
65	What are the performance metrics of parallel systems?	A. Execution time B. Total parallel overhead C. Speedup D. Efficiency E. All of the above	<b>E</b>	<b>1</b>	<b>4</b>	<b>4</b>	<b>3</b>	<b>4</b>
66	The isoefficiency function determines the ease with which a parallel system can maintain a constant efficiency. True or false?	A. True B. False	<b>A</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>
67	Which matrix-matrix multiplication algorithm uses a 3-D partitioning?	A. Cannon's algorithm B. DNS algorithm C. Both of the above D. None of the above	<b>B</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>
68	A solution representing a parallelism in an algorithm is	<b>A. CDA</b> <b>B. PTA</b> <b>C. CUDA</b> <b>D. CUD</b>	<b>c</b>	<b>1</b>	<b>6</b>	<b>1</b>	<b>1</b>	<b>2</b>
69	Blocking optimization is used to improve temporal locality, for reduce	<b>A. Hit miss</b> <b>B. Misses</b>	<b>B</b>	<b>1</b>	<b>5</b>	<b>1</b>	<b>1</b>	<b>2</b>

		<b>C. Hit rate</b> <b>D. Cache misses</b>						
70	Data are allocated to disks in the RAID at the	<b>A. Block level</b> <b>B. Cache level</b> <b>C. Low level</b> <b>D. High level</b>	<b>A</b>	1	6	<b>1</b>	<b>1</b>	<b>1</b>
71	In CUDA C programming serial code is executed by__and parallel code is executed by__	<b>a. CPU,CPU</b> <b>b. GPU,CPU</b> <b>c. GPU,GPU</b> <b>d. CPU,GPU</b>	<b>d</b>	1	6	<b>1,2,12,5</b>	<b>2</b>	<b>2</b>
72	Kernel function is qualified by the qualifier	<b>a. __local__</b> <b>b. __universal__</b> <b>c. __global__</b> <b>d. A or C</b>	<b>C</b>	1	6	<b>1,3</b>	<b>1</b>	<b>1</b>

(Mrs. D.A. Phalke & Mrs. Neha D. Patil)  
Subject Teacher

(Mrs. Vaishali Kolhe)  
Academic Coordintor

( Dr. Kailash Shaw)  
Dept. NBA Coordinator

(Dr. Vinayak Kottawar)  
HOD Computer