FINAL EXAM IN INF-3201

Exam in	**:**	**INF-3201**
Date	**:**	**December 7, 2011**
Time	**:**	**09:00 – 13:00**
Place	**:**	**Adm.bygget, B154**

Approved remedies :

- English dictionary

- English-Norwegian/Norwegian-English dictionary

The exam contains 3 pages including this cover page

Contact person:

Lars Tiede, mob. 45500550

FACULTY OF SCIENCE AND TECHNOLOGY

University of Tromsø, N-9037 Tromsø, Phone 77 64 40 01, Telefax 77 64 47 65

Please give short and concise answers. State explicitly any assumptions you do.

1) Parallel computers (15%)

a) What is Amdahl's law? If the serial fraction of an application is 20%, what is the maximum speed-up of the application?

b) What is *super-linear* speedup? Explain why search algorithms can achieve super-linear speedup.

c) What is Flynn's classification? Which types of computers do single-core computers, multi-core computers and graphics processing units (GPUs) typically belong to, respectively, according to Flynn's classification?

2) Message-passing computing (10%)

a) What is the difference between ***synchronous*** message passing and ***asynchronous*** message passing? Which one does not require local storage for messages? Explain your answer by providing a send-receive protocol that does not require local storage for messages.

b) What is collective communication in MPI? What is the main difference between MPI_Bcast(), MPI_Scatter() and MPI_Alltoall()?

3) Parallelization techniques (20%)

a) What is the main difference between the partitioning strategy and the divide-and-conquer strategy?

b) Fibonacci numbers are defined as follows

$F(0)=0$; $F(1)=1$;

$F(n) = F(n-1) + F(n-2)$ for $n>1$.

Which strategy, partitioning or divide-and-conquer, is most appropriate to compute F(n) in parallel? Why?

c) Write a function Fib(n) to compute F(n) in parallel using Cilk++. How to obtain the serialization of your function for a sequential regression test?

d) What is a (determinacy) race? Explain whether there are any possible races in your function Fib(n). Which tool in Cilk++ can be used to detect races?

4) Synchronization: Barriers (10%)

a) What is a barrier, and how is it used?

b) Why must simple barrier implementations often use two *distinct* phases, one for entry, and one for exit?

c) Outline an implementation (use pseudo-code) of a barrier with a centralized counter on a message passing system.

5) GPGPU computing with CUDA (15%)

a) What is a “Kernel”?

b) What are “thread blocks”, “warps”, “half warps”, and “grids”? How are they related?

c) Describe the properties of the entities of the CUDA memory model. Both "logical", i.e. how do they relate to each other, how can they be used, and "technical", i.e. access latencies and sizes, sharing and partitioning issues, possible quirks when using. You do not have to give exact numbers; relative characterizations, for example “large” or “very slow”, are sufficient.

d) What is occupancy? What occupancy is at least desired to hide register read-after-write latencies completely? Give a formula or at least reason with words. You do not have to know any numerical values – use variables and symbolic constants where you need them.

6) Data-parallel computations – segmented scan operations (15%)

Let A be a data vector, and F be a flag vector. Values 1 in the flag vector indicate the beginning of a segment.

a) We execute the *distribute* operation on vectors A and F given below. Write correct values to the result vector *Distributed*.

A	=	[0	5	3	8	1	9	2	1]
F	=	[1	0	1	0	0	0	1	0]
Distributed	=	[?	?	?	?	?	?	?	?]

b) In general, how can the *distribute* operation be implemented using a segmented +-scan?

c) How can the +-scan operation for the middle segment of A be parallelized to run concurrently on several processors? How many processors are ideal for that operation (in terms of achievable concurrency)? Illustrate your answer by drawing relevant data structures.

7) Distributed shared memory (15%)

a) What is “distributed shared memory” (DSM)?

b) Outline examples that illustrate when page-invalidate or page-update is the better update strategy for page-based DSMs (at least one for each strategy).

c) What is granularity in a DSM? Discuss how different granularities of shared data affect a DSM.

d) What distinguishes a *weak* memory consistency model from stricter memory consistency models?

Good luck!