A39849

Calculators may be used in this examination provided they are not capable of being used to store alphabetical information other than hexadecimal numbers

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Third Year Undergraduate

06-40089

LH High Performance Computing

Main Summer Examinations 2025

Time allowed: 2 hours

[Answer all questions]

Turn Over

Note

Answer ALL questions. The paper will be marked out of 100.

Question 1 [30 marks]

The following C/C++ function computes an outer product of two length N arrays and stores the resulting $N \times N$ matrix in a length N^2 array using row-major indexing.

```
void outerProd(int N, double* a, double* b, double* matrix) {
    for(int i = 0; i < N; i++) {
        double ai = a[i];
        for(int j = 0; j < N; j++) {
            matrix[i + N*j] = ai*b[j];
        }
    }
}
```

(a) Explain the concepts of memory hierarchy and locality of reference, and how they together impact application performance. Use these ideas to explain why the above code performs poorly, and suggest a small change to improve it. **[10 marks]**

(b) Explain the concept of Single Instruction Multiple Data (SIMD), or vector, processing and explain, with an example, why the inner loop of the function might not be automatically vectorized by a compiler. Propose a further small change that would ensure vectorization, explaining how it would help and how it would impact the use of the function. **[10 marks]**

(c) Imagine N is so large that the CPU running this program cannot fit any of these arrays on a single L1 cache line. Explain how this would impact performance and suggest further improvements to the code. **[10 marks]**

Question 2 [35 marks]

The finite difference method involves approximating derivatives using the Taylor series expansion, which for an arbitrary smooth function f is given by:

$$f(x+\Delta x) = f(x) + \Delta x \frac{\mathrm{d}f(x)}{\mathrm{d}x} + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + \frac{\Delta x^3}{6}\frac{\mathrm{d}^3 f(x)}{\mathrm{d}x^3} + \cdots = \sum_{n=0}^{\infty} \frac{\Delta x^n}{n!}\frac{\mathrm{d}^n f(x)}{\mathrm{d}x^n}$$

(a) Many HPC problems involve solving differential equations, which requires discretization. What is discretization and why is it needed? **[4 marks]**

(b) State the forward and central difference approximations to the first derivative and use the Taylor series to compare them in terms of accuracy **[6 marks]**

(c) The 1D viscous Burgers' equation is given by:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2},$$

where $u = u(x,t)$ is a function of x and t and ν is constant. Derive an explicit finite different scheme to solve this equation approximately and iteratively given an initial condition $u(x,0) = u^0(x)$. Your scheme should be second order accurate in space but can be first order accurate in time: $O(\Delta t + \Delta x^2)$ as Δx and Δt approach zero. You do not have to consider any spatial boundary conditions. **[9 marks]**

(d) Periodic boundary conditions (PBCs) in 1D are sometimes used to compute solutions on a ring-shaped domain. This means requiring that the solution satisfies $u(0,t) = u(L,t)$ for some length $L > 0$ and for all values of t. Explain how you would modify the iterative scheme from the last part in order to enforce PBCs. Assume that $L = N\Delta x$, where N is an integer. **[8 marks]**

(e) Explain how you might come up with an explicit, iterative finite difference scheme to approximately solve the 1D wave equation, which is given by:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}.$$

Assume that you have initial conditions for both the function, $u(x,0) = u^0(x)$, and its time derivative, $\frac{\partial u(x,0)}{\partial t} = v^0(x)$. You do not have to consider any spatial boundary conditions. **[8 marks]**

Question 3 [35 marks]

A parallel program using p distributed-memory compute units is operating on a rectangular three-dimensional (3D) grid. The size of each dimension of the grid is N, but the program's communication pattern in each direction is different. It is nearest-neighbour in the x direction (one step away), next-nearest-neighbour in the y direction (up to two steps away), and next-next-nearest-neighbour in the z direction (up to three steps away).

(a) A simple model of communication time between compute units is a linear equation, that is $t_{comm} = t_s + t_w L$. State what t_{comm}, t_s, t_w, and L represent. **[4 marks]**

(b) Use this communication model to construct three performance models for this program. The models should calculate the execution time of one iteration of the program when the grid is partitioned between compute units using 1D, 2D, and 3D decompositions respectively. You should make the following assumptions: constant computation time for each gridpoint each iteration, gridpoints can be always be equally divided between compute units, no idle time, and no overlap between computation and communication steps (although each unit can send and receive a message at the same time). Finally, choose optimal directions for the decompositions, and define any parameters that you introduce. **[14 marks]**

(c) Use each of these three models (1D, 2D, and 3D) to derive an expression for the parallel speedup of the program in terms of N, p, and any parameters you introduced. What does each model predict for the speedup if p becomes arbitrarily large (approaches infinity) while other parameters remain constant? **[8 marks]**

(d) In each of the three cases (1D, 2D, and 3D), what is the largest value that p can take as a function of N before the assumptions of the model break down. Give your answer in the form $p_{max} = kN^{\alpha}$. Using this result, compare the three decompositions in terms of their scalability to a large number of compute units. **[9 marks]**

This page intentionally left blank.

Do not complete the attendance slip, fill in the front of the answer book or turn over the question paper until you are told to do so

Important Reminders

- Coats/outwear should be placed in the designated area.
- Unauthorised materials (e.g. notes or Tippex) must be placed in the designated area.
- Check that you do not have any unauthorised materials with you (e.g. in your pockets, pencil case).
- Mobile phones and smart watches **must** be switched off and placed in the designated area or under your desk. They must not be left on your person or in your pockets.
- You are not permitted to use a mobile phone as a clock. If you have difficulty seeing a clock, please alert an Invigilator.
- You are not permitted to have writing on your hand, arm or other body part.
- Check that you do not have writing on your hand, arm or other body part – if you do, you must inform an Invigilator immediately
- Alert an Invigilator immediately if you find any unauthorised item upon you during the examination.

Any students found with non-permitted items upon their person during the examination, or who fail to comply with Examination rules may be subject to Student Conduct procedures.