# UNIVERSITY OF LONDON
## IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

## EXAMINATIONS 1998

BEng Honours Degree in Computing Part III
MEng Honours Degree in Information Systems Engineering Part IV
BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
MSc Degree in Advanced Computing
MSc Degree in Computing Science
for Internal Students of the Imperial College of Science, Technology and Medicine

*This paper is also taken for the relevant examinations for the*
*Diploma of Membership of Imperial College*
*Associateship of the City and Guilds of London Institute*
*Associateship of the Royal College of Science*

PAPER 3.25 / I4.22

PARALLEL PROBLEM SOLVING

Friday, May 15th 1998, 10.00 - 12.00

*Answer THREE questions*

For admin. only: paper contains 4 questions

1a The main issues involved in organising a parallel computation comprise *partitioning, communication and mapping.* Discuss how these concerns are addressed in the parallel programming paradigms exemplified by

i) Fortran together with a message passing library such as MPI.

ii) A data-parallel programming language such as High Performance Fortran (HPF)

iii) The Structured Co-ordination Language (SCL) applied to a conventional language such as Fortran.

b An image is represented by a two dimensional array of pixel values. The pixel averaging operation consists of repeatedly replacing each pixel by the average of itself and its four neighbours e.g.

$$p'_{ij} = \frac{p_{ij} + p_{i-1\,j} + p_{i+1\,j} + p_{ij+1} + p_{ij-1}}{5}$$

This operation is repeated until the maximum change in any pixel value from one iteration to the next is smaller than some given value.

i) Discuss strategies for executing this and similar operations efficiently in parallel on a distributed memory machine. What control and data structures are required to support this operation and why?

ii) Outline a program in the Structured Co-ordination Language (SCL) to compute this operation. You may assume the existence of any co-ordination operations you have discussed in i) above but you should explain their intended behaviour

2a "The Structured Co-ordination Language (SCL) combines ideas from co-ordination languages, skeletons and functional programming." Discuss this statement and explain what features from these areas are incorporated into SCL. What is the difference between SCL and a general purpose functional language and what advantages does this bring with regard to parallel implementation?

b Discuss how distributed data structures are represented and manipulated in SCL.

c The following High Performance Fortran (HPF) data distribution directives specify a data distribution of two N*M arrays, A and B, onto P*Q processors, where the rows of B are displaced N/P positions vertically (with wrap around) and the two structures are then partitioned and aligned into blocks of size (N/P)*(M/Q).

```
real dimension(N, M) :: A
real dimension(N, M) :: B
!HPF$ PROCESSORS PROC(P, Q)
!HPF$ DISTRIBUTE(BLOCK, BLOCK) ONTO Proc :: TEMPLATE
!HPF$ ALIGN A(I, J) WITH TEMPLATE(I, J)
!HPF$ ALIGN B((I+ N/P) MOD N),  J) WITH TEMPLATE(I, J)
```

i) Using diagrams or example data structures illustrate the data distribution that would be caused by such directives.

ii) Outline an SCL program that would result in the same data distribution.

*Turn Over……*

3a Describe the possible sources of parallelism inherent in the C4.5 tree induction algorithm. Discuss two possible ways of parallelising the C4.5 tree induction algorithm to take advantage of this inherent parallelism.

b You are given a data set to be used as training data containing 40 million items. Each item has 4 discrete attributes (including one for the class value). The maximal number of different values an attribute can take is 3. When applying a parallel tree induction algorithm for classification, what kind of parallel evaluation strategy would you choose and why?

c An organisation has 100 branches all over London. Each branch accumulates large data sets every hour. Each branch is currently using local workstations to mine its own data using an induction algorithm. The organisation is now planning to purchase a parallel computer for mining the daily accumulated data by collecting the data from all its branches into the large parallel computer via the internet and then using the parallel version of the induction algorithm for classification.

i) What are the major problems of this approach?

ii) Design an alternative method for classifying the entire data set that does not require purchasing a large parallel computer.

4 The following partial differential equation models a derivative, f, on a security, S, over time, t with other parameters constant.

$$\frac{\partial f}{\partial t}+(r-\frac{\sigma^2}{2})\frac{\partial f}{\partial S}+\frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial S^2}-rf=0$$

a Using the following Taylor series expansion approximations discretise the equation into a form suitable for solving on a finite difference grid.

$$\frac{\partial f}{\partial t}\cong\frac{f(S,t+\partial t)-f(S,t)}{\partial t}$$

$$\frac{\partial f}{\partial S}\cong\frac{f(S+\partial S,t+\partial t)-f(S-\partial S,t+\partial t)}{2(\partial S)}$$

$$\frac{\partial^2 f}{\partial S^2}\cong\frac{f(S+\partial S,t+\partial t)+f(S-\partial S,t+\partial t)-2f(S,t+\partial t)}{(\partial S)^2}$$

b Explain how a system of linear equations can be formed from this discretistion. Illustrate how these equations can be solved explicitly, given suitable boundary conditions.

c Describe how the processes of discretisation and numerical solution of partial differential equations may be automated, and how the solving of the linear equations may be parallelised.

*End of paper*