

THE AUSTRALIAN NATIONAL UNIVERSITY
First Semester Examination – June 2002

COMP3320

High Performance Scientific Computing

Study Period: 15 minutes

Time Allowed: 3 hours

Permitted Materials: Calculator, non-programmable

Exam questions total 100 marks.

Answer ALL questions.

*Clarity and conciseness in answers will be highly valued;
marks may be lost for supplying irrelevant information in
answers.*

Question 1 [22 marks]

Provide short and to the point answers to the following questions.

- (a) What is the difference between truncation error and rounding error? [2 marks]

- (b) For what range of x would the following expression be difficult to compute?

$$\frac{1}{1-x} - \frac{1}{1+x}$$

How might it be rearranged so that it can be computed for this problem range? [2 marks]

- (c) Give four features of the Fortran 95 programming language that make it well suited to computational science applications. [2 marks]

- (d) Is the following piece of Fortran 95 code correct?

```
program mystery
  integer, dimension(0:11)::Amat
  integer :: i,j
  Amat=(((i*j,i=3,1,-1),j=-1,2)/)
  print 99,Amat
  99 format("Matrix A",/,(5i5))
end program mystery
```

If so, exactly what is printed out when the program runs? Take care to detail each line and mark each blank space. If the code is incorrect, state why it is incorrect. [2 marks]

- (e) State one defining property of a singular matrix. [1 mark]

- (f) Give mathematical definitions for the 1-norm and 2-norm of the matrix A (denoted as $\|A\|_1$ and $\|A\|_2$ respectively). In general, which is easier to compute? [2 marks]

- (g) What is the difference between elapsed time and CPU time? Under what circumstances are they equal? [2 marks]

- (h) What are performance counters? Give two different reasons for using them. [2 marks]

(i) Explain the term trace-driven compilation.

[1 mark]

(j) In the context of compilers what are basic blocks and what is the intermediate language?

[2 marks]

(k) What does it mean to insert bubbles in to a pipeline?

[1 mark]

(l) What is the difference between a write-back and write-through cache? What is meant by cache coherency?

[3 marks]

Question 2 [20 marks]

You are a technical writer for the prestigious “Perspectives on Computational Science” magazine. The bulk of your readers are practicing computational scientists who, although computer literate, spend the majority of their time working in their own diverse range of application areas. They buy your magazine because they have found it to contain valuable advice on what hardware to purchase, useful information on trends in the computer industry, and technical tips that they can apply to their own particular field. Your editor has asked you to write an article for next months edition on one of the following topics.

- “To cluster or not to cluster, that is the question”. Clusters of PCs have emerged as increasingly important computing platforms. Your article should discuss the merits of “do-it-yourself” clusters compared to conventional shared memory parallel processors. A non-exclusive list of topics you might want to consider include: i) cost, ii) administration, iii) 64-bit, iv) applicable programming models, v) likely applications.
- “Hold your cash, tomorrow’s chips are hotter”. In the last 20 years we have seen dramatic improvements in CPU performance. This article should alert your readers to changes in computer hardware that are likely to impact on the performance of their codes in the next 5 years. A non-exclusive list of topics you might want to consider include: i) MHz, ii) processors per die, iii) pipelines, iv) caches, v) very long instruction word (VLIW) architecture.
- “Perform or perish”. When performance is critical it pays to devote some attention to tuning your code for the computer system you are using. This article should provide your readers with a basic introduction to single processor performance analysis and optimization. A non-exclusive list of topics you might want to consider include: i) code profiling, ii) compiler flags and compiler operation, iii) loop unrolling, iv) memory access, v) ultimate hardware limitations on performance.

15 marks will be awarded based on the topics you discuss bearing in mind the readership of the magazine. The remaining marks will be awarded for the overall style and structure of your article.

Question 3 [14 marks]

Three COMP3320 students have each kindly written a code for you that performs 1-dimensional discrete Fourier transformations. You intend to use one of these codes in your work on signal processing. All three codes have the same basic arguments of:

- n (input) - the number of data points
- x (input) - a vector of length n ($x_1, x_2, x_3, \dots, x_n$) containing the data to be transformed.
- t (input) - a vector of length n ($t_1, t_2, t_3, \dots, t_n$) giving the time at which each corresponding x data value was recorded.
- y (output) - a vector of length n ($y_1, y_2, y_3, \dots, y_n$) containing the resulting Fourier transform.

(a) Give TWO reasons why Fourier transformations might be used in signal processing. [1 mark]

(b) What data types (i.e. integer, real, complex, character etc) would you expect to be used for each of n , t , x and y ? What constraint would you expect there to be on the values of $t_1, t_2, t_3, \dots, t_n$? [3 marks]

(c) For the following TWO input data sets

$$x_A = [1, -1, 1, -1, 1, -1, 1, -1]$$

$$x_B = [4, 0, 3, 6, 2, 9, 6, 5]$$

comment on the general form of the Fourier transform (y_1, y_2, \dots, y_n). Note you are NOT required to evaluate the Fourier transformation, just to comment on interesting features or relations that may be expected to exist between its components. [2 marks]

You test all three codes using input data sets with different n . Within machine precision all three codes give identical results, however, in some cases the code from student#1 aborts. You run each code with each input data set on the same computer and gather the following performance data:

Input Data	Size n	CPU time (secs)		
		student#1	student#2	student#3
a	64	0.04	0.41	5.24
b	96	Abort	0.92	17.69
c	128	0.09	1.64	41.94
d	192	Abort	3.69	141.56
e	256	0.20	6.55	335.54
f	512	0.46	26.21	2684.35
g	783	Abort	61.31	9600.97
h	1024	1.02	104.86	21474.84

You conclude that in each case the student has implemented a fundamentally different algorithm for performing the Fourier transformation.

- (d) Describe what you believe each student has done. Give clear justification for your comments based on the above data. State the probable reason that the code from student#1 aborts.

[6 marks]

- (e) Student#1 suggests that his code can be made totally general by combining it with parts of the code from student#2. He claims that provided that n is NOT a prime number this new code will be faster than the original code from student#2. Moreover, he claims that this has nothing to do with optimization of the code, but results from a reduced operation count. Comment briefly on what you think student#1 plans to do?

[2 marks]

Question 4 [14 marks]

You wish to fit a quadratic of the form

$$p(t) = x_1 + x_2t + x_3t^2$$

to the following four data points

t	-1	0	1	2
y	-1	-4	5	6

- (a) The above problem can be described by an overdetermined system of linear equations involving the coefficients x_1 , x_2 and x_3 . Write down these equations in matrix form.

[1 mark]

- (b) Solve this matrix equation using the normal equation method, Gaussian elimination and either forward or backward substitution. Clearly detail all your working.

[8 marks]

- (c) For the general case of a polynomial of degree N fitted to M data points detail the computational cost of each of the distinct steps you undertook in part (b). Your answer should be expressed as $O(N^i M^j)$ where i and j are rational numbers (i.e. the prefactor is not important).

[3 marks]

- (d) Using the values of x_1 , x_2 and x_3 you determined in part (b) and for $t = -1, 0, 1$ and 2 evaluate the polynomial $p(t)$. What is the relative error between these “fitted values” of y and the actual values of y given above?

[2 marks]

Question 5 [16 marks]

You are given following segment of Fortran 95 code:

```
SUBROUTINE unknown(a,b,c,n)
  INTEGER:: n, i, j, k
  REAL*8:: a(n,n), c(n,n), b(n), temp

  DO k=1,n-1
    DO i=k+1,n
      c(k,i)=a(k,k)*a(k,i)
      temp=a(k,i)/a(k,k)
      DO j=k+1,n
        a(j,i)=a(j,i)+temp*a(j,k)
      END DO
      b(i)=b(i)+temp*b(k)
    END DO
    c(k,i)=c(k,i)*b(i)
  END DO
```

The code is executed on a 250MHz UltraSPARC-II processor. This processor can issue a group of up to 4 instructions per cycle, with:

- up to 2 integer instructions,
- up to 1 load or store instruction,
- up to 2 *different* floating point instructions
- up to 1 branch instruction.

The latency of a double precision load or store instruction is two cycles, while the latency of a floating point multiplication or addition is 3 cycles.

- (a) What is the minimum cycle time required to complete the floating point operations associated with just one iteration of the most inner j loop in the above code, assuming no loop unrolling. You should ignore all integer operations associated with the loop index and just give the cycle time from issuing the first floating point load until completion of the final floating point store. Based on this cycle time what is the best MFLOP rate you will achieve for this loop. Show exactly how you derive your answers.

[5 marks]

- (b) Write, in Fortran 95, a 4-way unrolled version of the inner j loop. What is the minimum cycle time to complete one iteration of this 4-way unrolled loop. Your time should relate only to the main loop not to any preconditioning loops. Based on this cycle time what is the best MFLOP rate you will achieve for this loop.

[5 marks]

(c) What is the theoretical maximum MFLOP rate that can be achieved for the j loop? What level of unrolling is required to achieve this performance? Loop unrolling consumes an important hardware resource that may ultimately limit the maximum level to which you would want to unroll a loop. What is this resource?

[4 marks]

(d) Your mate suggests unrolling the k loop will increase performance. Do you agree and why?

[2 marks]

Question 6 [14 marks]

The following Fortran 95 code was parallelized using OpenMP directives. It computes the value of Pi by numerical integration of the following expression:

$$\pi_{compute} = 4 \int_0^1 \frac{1}{1+x^2} dx$$

The number of processes used is set via an environment variable.

```
program Pi
  implicit none
  integer :: n_div=10000000          ! number of regions that the <0,1>
                                   ! interval is divided into

  integer :: i
  real*8  :: h,sum,Pi_calculated,x

  h = 1.d0/n_div
  sum=0.0d0

!$OMP PARALLEL DEFAULT(NONE), SHARED(h, n_div), PRIVATE(x), REDUCTION(+:sum)

!$OMP DO SCHEDULE(STATIC,5)
  do i = 1,n_div
    x = (i - 0.5d0)*h
    sum = sum + 1.0d0/(1.0+x*x)
  enddo
!$OMP END DO

!$OMP END PARALLEL

  Pi_calculated = 4.0d0*sum*h
  write(*,100) Pi_calculated
100 format(1x,'Computed Pi    ',f18.16)

  stop
end program Pi
```

- (a) Explain in detail the role of EACH directive or clause in the following OpenMP statements:

```
!$OMP PARALLEL DEFAULT(NONE), SHARED(h, n_div), PRIVATE(x), REDUCTION(+:sum)
!$OMP DO SCHEDULE(STATIC,5)
!$OMP END DO
!$OMP END PARALLEL
```

[6 marks]

- (b) Write code to show how you could remove the two OpenMP statements

```
!$OMP DO SCHEDULE(STATIC,5)
!$OMP END DO
```

and achieve the same objective using some or all of the following subroutine and function calls

- OMP_SET_NUM_THREADS(np) - subroutine, argument np sets the number of threads
- OMP_GET_NUM_THREADS() - function, return value represents the current number of threads
- OMP_GET_THREAD_NUM() - function, return value is a number that is unique to the calling thread that ranges from 0 to the number of active threads minus one.

[4 marks]

- (c) Using 2 threads on a dedicated 20 processor shared memory system the above code completes in 0.2 seconds elapsed time. Using 6 threads reduces this time to 0.1 seconds. Make a prediction for the performance you would obtain using 10 threads on the same machine. How would you expect the parallel performance of this code to change if ndiv was increased? You are required to show clearly how you derive your results.

[4 marks]

