AL
tory
on Laboratory

RAL Report
RAL-94-103

Comp

$\alpha_{critical}$ for Parallel Processors

D Moncrieff R E Overill and S Wilson

September 1994

Rutherford Appleton Laboratory Chilton DIDCOT Oxfordshire OX11 0QX

$\alpha_{critical}$ for parallel processors

D. Moncrieff
Supercomputer Computations Research Institute,
Florida State University,
Tallahassee, FL 32306,
U.S.A.

R.E. Overill
Algorithm Design Group,
Department of Computer Science,
King's College London,
Strand,
London WC2R 2LS,
UK

S. Wilson
Rutherford Appleton Laboratory,
Chilton, Oxfordshire OX11 0QX,
UK

Abstract

$\alpha_{critical}$ is defined as the fraction of a computational task that must be executed in parallel for the theoretical peak performance for two parallel processors to be equal. Values of $\alpha_{critical}$ are given for a number of contemporary high performance computing machines. $\tilde{\alpha}_{critical}$, the *practical parallelization break-even fraction*, is defined and discussed.

Keywords: Parallel processing, Amdahl's Law

1 Introduction

Amdahl's Law for parallel processors [1] [2] [3] [4] [5] may be written

$$S(\alpha, p) = ((1-\alpha) + \alpha/p)^{-1} \tag{1}$$

where S is the observed speed-up when a program is executed in a parallel processing environment compared with the uniprocessor performance. α $(0 \leq \alpha \leq 1)$ is the fraction of the computation that can be executed in parallel and p is the number of processing elements. It is assumed that there are no communications or memory latency and no synchronization overheads. It should be noted, however, that Amdahl's Law applies to a single problem of a given size and takes no account of any dependence that α may have on the size of the problem considered[6]. If $r_{\max}$ denotes the theoretical peak performance for a single processing element, then the peak rate of execution for a homogeneous system of p processors and a given α is

$$R = r_{\max} S(\alpha, p) \tag{2}$$

Given two computing machines with $r_{\max,i}$, p_i, $i = 1, 2$, it can be asked which machine will deliver the higher rate of execution for a computational task when a fraction α can be carried out in parallel. One machine (machine 1) might, for example, consist of a relatively small number of powerful processors as exemplified by the CRAY C-90 computer. The other machine (machine 2) might consist of a larger number of less powerful processing elements, for example, the CRAY T3D computer.

2 $\alpha_{critical}$ for parallel processors

For machine i, the *peak* rate of execution for a given value of α is

$$R_i(\alpha, p_i) = r_{\max,i} S_i(\alpha, p_i) = r_{\max,i}((1-\alpha) + \alpha/p_i)^{-1} \tag{3}$$

The difference between the theoretical performance levels of the machines 1 and 2 is

$$\Delta_{12}(\alpha) = R_1(\alpha, p_1) - R_2(\alpha, p_2) \tag{4}$$

The two machines will have equal theoretical peak performance if $\Delta_{12}(\alpha) = 0$, a condition that will be satisfied provided that the curves defined by equation (3) with $i = 1, 2$, intersect. This will be the case provided that

$$R_1(0, p_1) > R_2(0, p_2) \tag{5}$$

and

$$R_1(1, p_1) < R_2(1, p_2) . \tag{6}$$

$\Delta_{12}(\alpha) = 0$ can be solved for α to give

$$\alpha_{critical} = \left[1 - \left\{\left(\frac{1}{p_1 r_{\max,1}} - \frac{1}{p_2 r_{\max,2}}\right) / \left(\frac{1}{r_{\max,1}} - \frac{1}{r_{\max,2}}\right)\right\}\right]^{-1}, \tag{7}$$

with

$$0 \leq \alpha_{critical} \leq 1 \tag{8}$$

$\alpha_{critical}$ is the fraction of the code which has to be carried out in parallel in order for the two machines to achieve equal theoretical peak rates of execution. When $\alpha < \alpha_{critical}$ the theoretical peak rate of execution is greater on machine 1 and when $\alpha > \alpha_{critical}$ machine 2 delivers the greater theoretical peak rate of execution. $\alpha_{critical}$ is the *theoretical parallelization break-even fraction.*

By way of example, consider a fully configured CRAY C-90 computer with sixteen processors each of which can deliver a theoretical peak vector performance of $\sim 1000\ Mflop/s$: $p_1 = 16$ and $r_{\max,1} = 952\ Mflop/s$. Such a machine might be compared with a 2048 processor CRAY T3D computer for which $p_2 = 2048$ and $r_{\max,2} = 150\ Mflop/s$. Substituting these values into equation (5) gives $\alpha_{critical} = 0.9890$, so that over 98.90% of the computation must be executed in parallel for the CRAY T3D computer to yield a higher theoretical peak performance than the CRAY C-90 machine. Gregory [7] has discussed the intersection of the curve defined by Amdahl's Law for the CRAY C-90 with the corresponding curves for the Intel Paragon and the Thinking Machines Corporation CM-5 massively parallel processors.

As a second example, consider a fully configured CRAY Y-MP computer which has eight processors each of which can deliver a theoretical maximum vector performance of 333 $Mflop/s$; thus $p_1 = 8$ and $r_{\max,1} = 333\ Mflop/s$. Let us compare such a machine with a 256 processor CRAY T3D computer for which we have $p_2 = 256$ and $r_{\max,2} = 150\ Mflop/s$. These values when substituted in equation (7) give $\alpha_{critical} = 0.9129$, so that over 91.29% of the computation must be capable of being performed in parallel for the use of the CRAY T3D to lead to a higher peak execution rate than a fully configured CRAY Y-MP computer.

In this paper, we determine values of $\alpha_{critical}$ for various pairs of contemporary machines. Table 1 summaries the basic characteristics of the machines considered. In Table 2, values of $\alpha_{critical}$ are given for various pairs of machines. Entries are only given when conditions (5) and (6) are satisfied.

Now, suppose that the *sustained* performances of the individual processors on machines 1 and 2 are $\tilde{r}_1$ and $\tilde{r}_2$, respectively. Then $\tilde{\alpha}_{critical}$, the *actual parallelization break-even fraction*, is a function of $\tilde{r}_1$ and $\tilde{r}_2$

$$\tilde{\alpha}_{critical}(\tilde{r}_1, \tilde{r}_2) = \left[1 - \left\{\left(\frac{1}{p_1 \tilde{r}_1} - \frac{1}{p_2 \tilde{r}_2}\right) / \left(\frac{1}{\tilde{r}_1} - \frac{1}{\tilde{r}_2}\right)\right\}\right]^{-1}. \tag{9}$$

Putting

$$\tilde{r}_i = \beta_i r_{\max,i}, \quad i = 1, 2, \quad 0 \leq \beta_i \leq 1 \tag{10}$$

then $\tilde{\alpha}_{critical}$ is a function of β_1 and β_2. For machine i with multiple vector processors, Hockney and Jesshope [8] have shown that β_i represents the degradation from the theoretical peak performance of a machine due to vector start-up, task synchronization, and data communication:

$$\beta_i = \left[\left(1 + \frac{n_{\frac{1}{2},i}}{n}\right) \left(1 + \frac{s_{\frac{1}{2},i}}{s}\right) \left(1 + \frac{f_{\frac{1}{2},i}}{f}\right) \right]^{-1} \qquad (11)$$

where $n_{\frac{1}{2},i}$ is the half-performance vector length, $s_{\frac{1}{2},i}$ is the half-performance task granularity, and $f_{\frac{1}{2},i}$ is the half-performance computational intensity for machine i; these three parameters are respectively functions of the architecture, the operating system, and the hardware of the machine i. n is the average vector length, s is the average task granularity, and f is the average computational intensity of the application in question. Values of $\tilde{\alpha}_{critical}$ for the CRAY C-90 (machine 1) and CRAY T3D/2048 (machine 2) are displayed in Table 3. Again, entries are only given when conditions (5) and (6) are satisfied. $\tilde{\alpha}_{critical} = \alpha_{critical}$ when $\beta_1 = \beta_2$, $\tilde{\alpha}_{critical} > \alpha_{critical}$ when $\beta_1 > \beta_2$, and $\tilde{\alpha}_{critical} < \alpha_{critical}$ when $\beta_1 < \beta_2$.

3 Summary

$\alpha_{critical}$ encapsulates in a single parameter the difference in the relative theoretical peak performance of parallel machines. The values of $\alpha_{critical}$ in Table 2 lend support to the view [3] that the limits of expectation for massively parallel processors may have been somewhat over-optimistic when these machines are compared directly with parallel vector processors. This conclusion is not significantly altered when the effects of overheads due to vector start-up, synchronization, and communication are taken into account as Table 3 demonstrates.

Acknowledgement.
D.M. acknowledge the support of the US Department of Energy through contract number DE-FC05-85ER2500000.

Table 1.
Some characteristics of machines considered

Machine	p	$r_{\max}(Mflop/s)$
Fujitsu VP550	222	1650
NEC SX-3R/44	4	6400
CRAY C-90	16	952
CRAY Y-MP	8	333
CRAY T3D/2048	2048	150
CRAY T3D/256	256	150
TMC CM-5E	1024	160
Intel Paragon XP/S	1840	75
KSR2	256	80
IBM SP-1	256	125
IBM SP-2	2560	264
Meiko CS-2	64	200

Table 2.

Values of $\alpha_{critical}$ for various pairs of machines

	Fujitsu VP550	NEC SX-3R/44	CRAY C-90	CRAY Y-MP
CRAY T3D/2048	–	0.9945	0.9890	0.9078
CRAY T3D/256	–	0.9980	0.9930	0.9129
TMC CM-5E	–	0.9946	0.9887	0.8979
Intel Paragon XP/S	–	0.9976	0.9953	0.9656
KSR2	–	–	0.9985	0.9668
IBM SP-1	–	0.9990	0.9951	0.9356
IBM SP-2	0.9996	0.9898	0.9771	0.6773
Meiko CS-2	–	–	–	0.8704

Table 3.
Values of $\tilde{\alpha}_{critical}$ for the CRAY C-90 (machine 1) and CRAY T3D/2048 (machine 2).

		β_1				
		0.1	0.3	0.5	0.7	0.9
	0.1	0.9890	0.9971	0.9985	0.9991	–
	0.3	0.9478	0.9890	0.9940	0.9960	–
β_2	0.5	0.8132	0.9789	0.9890	0.9927	–
	0.7	–	0.9657	0.9832	0.9890	–
	0.9	–	–	–	–	–

References

[1] G.M. Amdahl, 1967, Validity of the single processor approach to achieving large scale computing capabilities, *Proc. AFIPS Joint Spring Computer Conf.*, 483-485

[2] W.H. Ware, 1972, The ultimate computer, *IEEE Spectrum* **9**, 84-91

[3] G.M. Amdahl, 1988, Limits of expectation, *Intern. J. Supercomput. Applns.* **2**, 88-94

[4] J.J. Hack, 1989, On the promise of general purpose parallel computing, *Parallel Computing*, **10**, 261-275

[5] B.W. Davies, 1990, Supercomputing - A Forward Look, in *Supercomputational Science*, edited by R.G. Evans and S. Wilson, Plenum, New York, pp. 333-343.

[6] J.L. Gustafson, 1988, Comms. ACM **31**, 532-533

[7] P. Gregory, 1992, Will MPP Always Be Specialized?, *Supercomputing Review* **5**, (3) 28-31

[8] R.W. Hockney and C.R. Jesshope, 1988, *Parallel Computers 2*, Adam Hilger, *pp.* 101-116