FILOMAT (Niš) 9:3 (1995), 839–849
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

A SET OF AXIOMS FOR EVALUATING THE MULTIPROCESSOR PERFORMANCES

I. Ž. Milovanović, E. I. Milovanović,
M. D. Mihajlović and M. K. Stojčev

ABSTRACT. When designing a parallel computer it is very important that it has the predicted performances. The challenge for a computer designer is to discover the minimum organization and equipement necessary to achieve given level of performance. So, the developing analythical model for characterizing and understanding the parallel system performances is of a crucial interest. In order to avoid erroneous conclusions about the behaviour of parallel system a severe mathematical formulations should be involved. In this paper we give a survey of axioms that were proposed in the literature in order to introduce a scientific approach in studing the parallel system performances. Further we shall propose a modified and reduced set of axioms based on discrete mathematical apparatus.

1. Introduction

From the very beginning of digital computer development, the designers always storve to increase the speed of operations. There are number of possible ways to achieve this. An obvious approach is to improve the technology implemented in the realization of the computer components. There is of course a natural limitation in technology development: no signal can propagate faster than the speed of the light. Another way for increasing the speed of computation is by performing as many operations as possible simultaneosly, concurrently, in parallel, using parallel computers [8].

A parallel computer is one that consists of a collection of processing units, or processors, that cooperate to solve a problem by working simultaneously on different parts of that problem. The number of processors used can range from a few tens to several millions. As a result, the time required to solve a problem by a traditional uniprocessor computer is significantly reduced.

This approach is attractive for a number of reasons [1]. First, for many computational problems, the natural solution is a parallel one. Second, the cost and size of computer components have declined so sharply in recent years that parallel computers with a large number of processors have become feasible. And, third, it is possible in parallel processing to select the parallel architecture that is best suited to solve the problem or class of problems under consideration. Indeed, architects of parallel computers have freedom to decide how many processors are to be used, how powerful these should be, what interconnection network links them to one another, whether they share a common memory, to what extent their operations are to be carried out synchronously, and host of other issues. This wide range of choices has been reflected by many theoretical models of parallel computation proposed as well as by several parallel computers that were actually built. Since parallel computers are composed of multiple processors, interconnected to each other, and sharing the use of memory, input–output peripherials and other resources, estimating the performances of these systems is really complex. The fact that the same system behaves differently when solving various problems makes the performance evaluation even more difficult. Different problems have different possibilities for parallelization. Some problems can't be parallelized at all.

When designing parallel computer it is very important that the system has the predicted properties. It is also very important to design the algorithm that exploits both parallelism inherent in the problem and that available on the computer. The challenge for a computer designer is to discover the minimum organization and equipement necessary to achieve a given level of performance. By performance we mean the manner in which, or the efficiency with which, a computer system meats its goal. So, the developing analytical models for characterizing and understanding the parallel system performances is of a crucial interest. But, attempts to express some measure of performance as a explicit function of certain parameters were not successful always. Moreover, omitting one of the parameters leads to erroneous conclusions about the behaviour of parallel system. Thus, for example, in 1967 Amadahl (see [7]) made the observation that if s is the serial fraction in an algorithm, then its speedup is bounded by $1/s$, no matter how many processors are used. For example, if there are only 5% of the algorithm that can't be parallelized, then maximal speedup that can be achieved is 20, no matter how many processors are used. This statement, now popularly known as Amadahl's law, has been used by Amadahl and others to argue against the usefulness of large scale parallel computers. Fortunately, Amadahl was wrong. He missed the fact that the serial fraction, s, is a function of problem size, m. Moreover, for most scientific and technical applications it has

property that $\lim_{m\to\infty} s(m) \to 0$ [4].

The above and other similar examples imply that in studing the performances of a parallel system, a severe mathematical formulations should be involved. Therefore the following should be developed:

- Explicit mathematical formulas that characterize the performances of parallel system
- Axioms for basic parameters.

The key for scientific approach in studing the performances lies in solving the above problems.

Some common performance measures of parallel algorithm running on a parallel computer are the execution time, the speedup, the efficiency, the scalability, etc.

2. Definitions and assumptions

In this section, we introduce some terminology used in the rest of the paper.

We assume the system of n identical processors interconnected in some way for the purpose of passing data and control information between the processors. Communication between the processors can be achieved via common memory modules or by message passing. Each processor is supplied by some amount of local memory. By a parallel system we mean a combination of a parallel algorithm and a parallel architecture comprising of identical processing units.

Definition 2.1. *The degree of parallelism* of a numerical algorithm is the number of operations in the algorithm that can be done in parallel.

Note that the degree of parallelism is independent of the number of processors in the system; it is an intristic measure of the parallelism in the algorithm.

Definition 2.2. *The average degree of parallelism* of an algorithm is the total number of operations in the algorithm divided by the number of stages. A stage is comprised of the operations that can be performed in parallel.

Consider a program for which execution time on a single processor is equal to $T(1)$. When this program runs on a multiprocessor, the execution time can be divided into two components:

- component with running time T_s that must be run sequentialy;
- component with sequential running time T_p that can be subdivided into parallel components running on different processors.

Note that

$$T(1) = T_s + T_p \,. \tag{1}$$

Since a small number of problems posses ideal intristic parallelism, T_s is greater than zero.

Assume that $T(n)$ is execution time when program is running on n–processor system. The $T(n)$ involves the following components:

- serial execution time T_s,
- parallel execution time, equal to T_p/n if the parallelizable part of the program can be partitioned into n parallel components of equal running time, and
- synchronization and communication overhead $T_0(n)$.

According to the previous, we have the following definition.

Definition 2.3. *The execution time of an algorithm* running on n–processor system is

$$T(n) = T_s + \frac{T_p}{n} + T_0(n) \,. \tag{2}$$

Definition 2.4. *The speedup* of parallel system is defined as

$$S(n) = \frac{T(1)}{T(n)} \,. \tag{3}$$

The parallel algorithm may not be the best algorithm on a single processor, so for $T(1)$ we take the execution time on a single processor of the fastest serial algorithm.

Definition 2.5. *The efficiency* of the parallel system is

$$E(n) = \frac{S(n)}{n} = \frac{T(1)}{nT(n)} \,. \tag{4}$$

Definition 2.6. *Parallel cost penalty* is defined as

$$C(n) = nT(n) \,. \tag{5}$$

Definition 2.7. *Relative parallel cost penalty* is

$$R(n) = \frac{nT(n) - T(1)}{n-1}. \tag{6}$$

Definition 2.8. *The gain factor* of a parallel system is

$$G(n) = \frac{T(1) - T(n)}{T(1)} = 1 - \frac{1}{S(n)}. \tag{7}$$

It is not difficult to see from the above definitions, that the execution time is the primary measure of a parallel system performance which is used as a basis for estimating other characteristics of a system. So, it was natural to establish the set of axioms for this metric.

3. The set of axioms

In the text that follows we are going to give the survey of axioms that were proposed in literature in order to introduce a scientific approach in studing the parallel system performances. Further, we shall propose a modified set of axioms based on discrete mathematical apparatus.

As we have already mentioned, the execution time is the most important measure of parallel system performance. Its component $T_0(n)$ represents the influence of communication and synchronization between processors on execution time. The value of $T_0(n)$ directly affects the performance of the whole system. So, the analytical methods for characterizing and understanding this measure were developed. In [2] Flatt and Kennedy introduced the following axioms for $T_0(n)$:

F.1 $T_0(n)$ is continuous and twice differentiable in respect to n,
F.2 $T_0(1) = 0$,
F.3 $T_0'(n) > 0$ for all $n \geq 1$, hence $T_0(n)$ is nonnegative,
F.4 $nT_0''(n) + 2T_0'(n) > 0$ for all $n \geq 1$,
F.5 There exists $n_1 \geq 1$ such that $T_0(n_1) = T(1)$.

On the basis of the involved axioms, the authors have investigated the impact of synchronization and communication overhead on the performance of parallel systems. They have established upper bounds on the power of parallel processing in the presence of synchronization and communication overheads.

The pioniers work of Flatt and Kennedy has motivated researchers to investigate the following:

a) Is the set of axioms **F.1**–**F.5** the minimal one, or it can be reduced, and, can some conditions be weaker?
b) What is physical and/or geometrical meaning of **F.1**–**F.5**?
c) Why $T_0(n)$ and other measures are considered as real functions, if they are defined on the set of natural numbers, $\mathbb{N}$?

In [6] the problems a) and b) were considered. The axiom **F.5** is rejected as too strong, and instead of it the condition

$$\lim_{n\to\infty} T_0(n) = +\infty \tag{8}$$

is tested.

As a basic value in [6] the function $D(n) = nT_0(n)$, instead of $T_0(n)$, is taken. This enables author to introduce the following more geometricaly intuitive axioms:

D.1 $D(1) = 0$,
D.2 $D(n) \geq 0$,
D.3 $D(n)$ is strictly convex and differentiable.

The author has proved that any $T_0(n)$ satisfying **F.1** to **F.4** also satisfies **D.1** to **D.3**. The question c) was not addressed in this paper.

In [5] the problems a) and c) were addressed. Namely, the values $T_0(n)$ and $T(n)$ were considered as members of sequences $\{T_0(n)\}$ and $\{T(n)\}$, respectively. This enables authors to reduce the set of axioms, defined by Flatt and kennedy, from five to the following three:

P.1 $T_0(1) = 0$, $T_0(2) \geq 0$,
P.2 $(n+2)\Delta^2 T_0(n) + 2\Delta T_0(n) > 0$, for $n \geq 1$.
P.3 There exists n_1 such that $T_0(n_1) = T(1)$.

It was shown that performance evaluation can be carried out very efficiently using discrete mathematical apparatus.

Let us note that it is natural to use the discrete mathematical apparatus, since the number of processors in the system is an integer value. Besides, by utilizing this apparatus the condition **F.1** (that the function is continuous and twice differentiable) becomes needless. Also, the axiom **F.2** is expressed as natural and elemental condition $T_0(2) \geq 0$. The physical and/or geometrical meaning for **P.2** and **P.3** can't be given.

Inspired by the papers [6] and [5] and having in mind questions a), b) and c), we propose in this paper a new set of axioms for sequence $\{D(n)\}_{n\in\mathbb{N}}$, $D(n) = nT_0(n)$, as follows:

A.1 $D(1) = 0$,
A.2 $D(2) \geq 0$,
A.3 $\Delta^2 D(n) > 0$, for $n \geq 1$.

Usage of discrete apparatus enables us to propose somewhat weaker conditions for $D(n)$ compared with **D.2** and **D.3** from [6]. Namely, instead of $D(n) \geq 0$ we take a condition $D(2) \geq 0$ and for sequence $\{D(n)\}$ we assume to be convex instead of function $D(n)$ being strictly convex and differentiable. Further, instead of axiom **P.3** from [5] we shall take a weaker condition (8) as in [6].

Now, we shall prove the following result.

Theorem 3.1. *The set of axioms* **A.1**–**A.3** *is equivalent with* **P.1**–**P.2**.

Proof. Statement of the Theorem 3.1 directly follows from the equalities $D(1) = T_0(1)$, $D(2) = 2T_0(2)$ and $\Delta^2 D(n) = (n+2)\Delta^2 T_0(n) + 2\Delta T_0(n)$. □

The assumption of axiom **P.3** is not involved in **A.1**–**A.3**. Therefore, we are going to prove the main result for $T(n)$ from [5] using **A.1**–**A.3** and under assumption that (8) is satisfied. But, first, we shall prove two auxiliary results.

Lemma 3.1. *The sequence* $\{D(n)\}_{n\in\mathbb{N}}$, *is positive and monotone increasing.*

Proof. According to **A.3** it follows that

$$\sum_{k=1}^{n} k\Delta D(k+1) > \sum_{k=1}^{n} k\Delta D(k)\,.$$

From the above inequality it follows that $\dfrac{D(n+2)}{n+1} > \dfrac{D(n+1)}{n}$, i.e.

$$\frac{D(n+2)}{n+1} > \frac{D(n+1)}{n} > \cdots > \frac{D(2)}{1}\,,$$

and according to **A.2** we have $D(2) \geq 0$, i.e. the sequence $\{D(n)\}$ is monotone increasing. □

Lemma 3.2. *The sequence* $\{T_0(n)\}_{n\in\mathbb{N}}$ *is monotone increasing.*

Proof. According to inequality $\Delta^2 D(k) > 0$, i.e. $\Delta D(k) > \Delta D(k-1)$, we have that

$$(k+1)\Delta T_0(k) > (k-1)\Delta T_0(k-1)\,.$$

above inequality it follows

$$\sum_{k=2}^{n} k(k+1)\Delta T_0(k) > \sum_{k=2}^{n}(k-1)k\Delta T_0(k)\,,$$

i.e.

$$n(n+1)\Delta T_0(n) > 1\cdot 2\Delta T_0(1) = 2T_0(2) = D(2) \geq 0\,,$$

which establishes the result. □

We now present the main result for $T(n)$.

Theorem 3.2. *If equality (8) is satisfied, then there exists an unique value, $n \geq n_0$, for which the sequence $\{T(n)\}$ reaches the minimum, i.e. the inequality*

$$T(n_0) \leq T(n)$$

for all $n \geq 1$, is valid. When the inequality $D(2) \leq T_p$ is valid, then n_0 is an unique solution to

$$(n_0-1)n_0\Delta T_0(n_0-1) - T_p \leq 0$$
$$n_0(n_0+1)\Delta T_0(n_0) - T_p \geq 0\,. \tag{10}$$

Proof. Let $D(2) \leq T_p$, i.e. $2T_0(2) \leq T_p$. According to (2) the equality

$$\Delta T(n) = \Delta T_0(n) - \frac{1}{n(n+1)}T_p = \frac{1}{n(n+1)}\left(n(n+1)\Delta T_0(n) - T_p\right)$$

is valid. From (10), for $n = 1$ we obtain the inequality

$$\Delta T(1) = \frac{1}{2}\left(D(2) - T_p\right) \leq 0\,. \tag{11}$$

Now, it is necessary to prove that the sequence $\{T(n)\}$ is not decreasing for all $n \geq 1$, but there exists n for which $\Delta T(n) \geq 0$. Assume the opposite, i.e. that for all $n \geq 1$ the inequality $\Delta T(n) \leq 0$ is valid. Then according to (10) we obtain

$$\Delta T_0(n) = \Delta T(n) + \frac{1}{n(n+1)}T_p \leq \frac{1}{n(n+1)}T_p\,.$$

According to (11) we have

$$\sum_{k=1}^{n-1}\Delta T_0(k) \leq \sum_{k=1}^{n-1}\frac{1}{k(k+1)}T_p\,,$$

i.e.

$$T_0(n) \leq T_p \left(1 - \frac{1}{n}\right).$$

From the last inequality it follows that

$$\lim_{n \to +\infty} T_0(n) \leq T_p$$

which is in contradiction to the assumption that (8) is valid. Consequently, we conclude that the assumption $\Delta T(n) \leq 0$ for all $n \geq 1$, is not correct. Namely, there are values for n for which $\Delta T(n) \geq 0$, i.e. there is at least one value $n = n_0$ for which the inequalities

$$\Delta T(n_0 - 1) \leq 0\,, \qquad \Delta T(n_0) \geq 0 \tag{12}$$

and (9) are valid.

Now, it is necessary to prove that n_0 is unique. Using **A.3** we obtain

$$\Delta\left(n(n+1)\Delta T_0(n)\right) = (n+1)\Delta^2 D(n) > 0\,. \tag{13}$$

From (13) it can be concluded that the sequence $\{n(n+1)\Delta T_0(n)\}$ is monotone increasing. Accordingly, there exist the unique value $n = n_0$ such that inequalities (9) and (12) are valid.

Now, we are going to prove that for $n = n_0$ the sequence $\{T(n)\}$ reaches a minimum. To prove this it is enough to show that $\Delta^2 T(n_0 - 1) > 0$ and that the sequence $\{T(n)\}$ is monotone increasing for $n \geq n_0$.

From (10) we obtain

$$\Delta^2 T(n) = \Delta^2 T_0(n) + \frac{2}{n(n+1)(n+2)} T_p\,. \tag{14}$$

According to (10) and (14) it follows that

$$\Delta^2 T(n) = \frac{1}{n+2}\left(\Delta^2 D(n)\right) - \frac{2}{n+2}\Delta T_0(n)\,. \tag{15}$$

Substituting $n = n_0 - 1$ in (15) and using inequality (12), the inequality

$$\Delta^2 T(n_0 - 1) \geq \frac{1}{n_0 + 1}\left(\Delta^2 D(n_0 - 1)\right) > 0$$

is obtained.

Since the sequence $\{n(n+1)\Delta T_0(n)\}$ is monotone increasing, and according to equality $n(n+1)\Delta T(n) = n(n+1)\Delta T_0(n) - T_p$, it follows that the sequence $\{n(n+1)\Delta T(n)\}$ is monotone increasing, also. On the other hand, since $n_0(n_0+1)\Delta T_0(n_0) - T_p \geq 0$ it follows that $n_0(n_0+1)\Delta T(n_0) \geq 0$. Now, according to inequality $n(n+1)\Delta T(n) \geq n_0(n_0+1)T(n_0) \geq 0$, we obtain that $\Delta T(n) \geq 0$, for all $n \geq n_0$.

If we assume that $D(2) \geq T_p$, then from (10) we obtain $\Delta T(1) = \frac{1}{2}(D(2) - T_p) \geq 0$. Further, since $\Delta T(n) \geq \Delta T(1) > 0$ for all $n \geq 1$, we conclude that in this case $n_0 = 1$. □

Remark. Theorem 3.2 have been proved in [5] under conditions **P.1** to **P.3**.

According to the results proved in Theorem 3.2 we are going to prove the following results for sequences $\{S(n)\}$ and $\{C(n)\}$, $n \in \mathbb{N}$.

Theorem 3.3. *Let the inequality (8) be satisfied. The sequence $\{S(n)\}$ has an unique maximum at $n_0 \geq 1$. Also, if $D(2) < T_p$ then*

$$\frac{T(1)}{\Delta C(n_0)} \leq S(n_0) \leq \frac{T(1)}{\Delta C(n_0 - 1)}. \tag{16}$$

Proof. In theorem 3.2 we have proved that, under certain conditions, the sequence $\{T(n)\}$, $n \in \mathbb{N}$, has the unique minimum at $n_0 \geq 1$. The following is also valid

$$\Delta T(n_0 - 1) \leq 0, \qquad \Delta T(n_0) \geq 0. \tag{17}$$

From (3) we have that

$$\Delta S(n) = T(1)\left(-\frac{\Delta T(n)}{T(n)\,T(n+1)}\right). \tag{18}$$

Combining (17) and (18) we obtain

$$\Delta S(n_0 - 1) \geq 0, \qquad \Delta S(n_0) \leq 0. \tag{19}$$

This means that the sequence $\{S(n)\}$, $n \in \mathbb{N}$, has the unique maximum at $n_0 \geq 1$.

Since $C(n) = nT(n)$, i.e. $C(n) = \dfrac{nT(1)}{S(n)}$, the following is also valid

$$\Delta C(n) = T(1)\,\frac{S(n) - n\,\Delta S(n)}{S(n)\,S(n+1)},$$

and

$$\Delta C(n) = T(1)\,\frac{S(n+1) - (n+1)S(n)}{S(n)\,S(n+1)}. \tag{21}$$

By substituting n with $n_0 - 1$ in (20) and n with n_0 in (21), the right and left parts of inequality (16) are obtained, respectively. □

Similarly as in [2], [5], some other properties of sequences defined by (2) to (7) can be proved. For the sake of illustration, we give some properties that directly follows from axioms **A.1** to **A.3**.

Theorem 3.4. *The sequence* $\{C(n)\}$, $n \in \mathbb{N}$, *is monotone increasing and convex.*

Theorem 3.5. *The sequence* $\{E(n)\}$, $n \in \mathbb{N}$ *is monotone decreasing, convex, and has a property* $E(1) = 1$.

Theorem 3.6. *The sequence* $\{R(n)\}$, $n \in \mathbb{N}$, *is monotone increasing for* $n \geq 2$.

Theorem 3.7. *If the equality (8) is valid, then the sequence* $\{G(n)\}$, $n \in \mathbb{N}$, *has an unique maximum at* $n_0 \geq 1$.

References

[1] S.G. Akl, *The Design and Analysis of Parallel Algorithms.*, Prentice–Hall Inc., New Jersey, 1989.

[2] H.P. Flatt and K. Kennedy, *Performance of Parallel Processors*, Parallel Comput. **12** (1989), 1–20.

[3] G. Golub and J.M. Ortega, *Scientific Computing: An Introduction with Parallel Computing*, Academic Press, Inc., 1993.

[4] J.L. Gustafsson, *The Scale–Sized Model: A Revision of Amadahl's Law*, Proc. 3rd Conf. on Supercomputing, Boston '88, 1988.

[5] I.Ž. Milovanović, E.I. Milovanović, M.K. Stojčev, *Discrete Timing Model for Parallel Processing*, Facta Universitatis (Niš) Ser. Math. Inform. **7** (1992), 123–144.

[6] D. Müller–Wichards, *Problem Size Scaling in the Presence of Parallel Overhead*, Parallel Comput. **17** (1991), 1361–1376.

[7] J.M. Ortega, *Introduction to Parallel and Vector Solution of Linear Systems*, Plenum Press, New York and London, 1988.

[8] D. Tabak, *Multiprocessors*, Prentice–Hall Inc., New Jersey, 1990.

Faculty of Electronic Engineering, Beogradska 14, P.O. Box 73, 18000 Niš, Serbia.