

Optimization Scheme Based on Parallel Computing Technology

Xiulai Li, Chaofan Chen, Yali Luo, and Mingrui Chen^(✉)

Hainan University, Haikou, Hainan, China
mrchen@hainu.edu.cn

Abstract. Parallel computing is a high performance technology to solve problems, in order to improve computing efficiency, we use the processor to concurrent execute several parts divided from one problem. Based on the current issues in parallel computing area, both the data processing repetition rate and the parallel computing time depend on the time of the last thread in the task completing. This paper was written to take an overview of the existing parallel computing techniques and structures, and propose a solution of adding an advanced thread or advanced processor to make up the deficiency in parallel computing area.

Keywords: Parallel structure · Parallel computing · Parallel model · Parallel technology

1 Introduction

From the day of the birth of the computer, people continue to redouble efforts to improve the speed of the computer, and has achieved very significant results. However, this effort will not be long before the termination of the limit of the physical device. One of the common characteristics of people in the effort to develop a new generation of computers is the use of parallel technology. Increase in the same time interval the number of operations technology called parallel processing technology; design for parallel processing computer called parallel computers; to solve the problem in parallel computer called parallel computing; in parallel computer implementation of problem solving algorithm called parallel algorithm [1].

Traditionally, the general software design is a serial calculation:

- (1) The software runs on a computer with only one CPU;
- (2) The problem is decomposed into discrete sequence of instructions;
- (3) The instruction is executed by one by one;
- (4) At anytime CPU up to only one instruction at run time. The operational principle of CPU is described as Fig. 1.

In the simplest case, parallel computing is to use a number of computing resources to solve the problem.

- (1) The purpose of using multi-core CPU to run;
- (2) The purpose of the problem is decomposed into discrete parts can be solved at the same time [2];

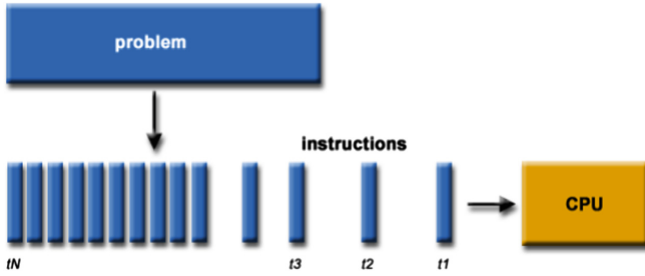


Fig. 1. The operational principle of CPU

- (3) The purpose of each part is subdivided into a series of instructions;
- (4) In each part of the instruction can be executed simultaneously in different CPU [3]. The operational principle of multi-CPU is described as Fig. 2.

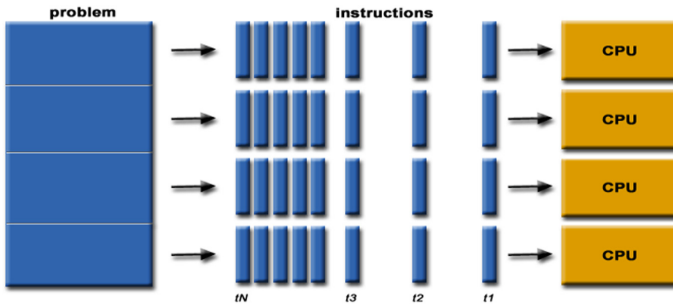


Fig. 2. The operational principle of multi-CPU

A wide range of parallel computing needs, but to sum up there are three types of applications: Compute-Intensive applications, such as large-scale scientific and (Data-Intensive); data intensive applications, such as numerical library, data warehouse, data mining and visualization; network intensive applications, such as collaborative work, remote control and remote medical diagnosis etc. [4].

Parallel computing, said simply that the computation is made in parallel computer, it is often said that the calculation and high performance, super computing is a synonym for any high performance computing and super computing cannot do without parallel technology [5].

2 Parallel Computing Architecture

Since the parallel computing technology since the middle of 60 s, the parallel processing has experienced from the array machine (SIMD), the vector processor, the shared memory vector machine (SMP), massively parallel processing, distributed storage system (MPP) to the workstation (COW) process [6].

Parallel architecture is the basis of parallel computing, and the design mechanism of various parallel programs are also different. It can be roughly divided into the following five categories.

2.1 SIMD

Array processor (SIMD) is a duplicate set processing unit to carry out the provisions of the same instruction operations on their assigned data in a single control unit under control by means of an interconnected array is operation level parallel computer SIMD [7]. The SIMD type parallel computer has played an important role in the development of parallel computer, but due to the development of processor technology since 90s, for science and engineering calculation of the SIMD type parallel machine has basically quit the stage of history. The system of SIMD is described as Fig. 3.

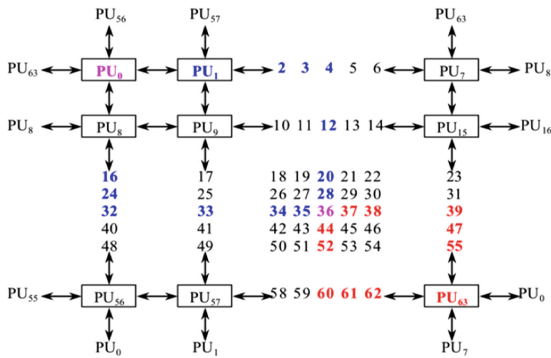


Fig. 3. The system of SIMD

2.2 Vector Machine

Vector Machine can perform high-speed processing of vector operation with a special vector registers and vector flow components, except scalar registers and scalar functions [8]. The system of vector machine is described as Fig. 4.

2.3 SMP

Shared memory processor systems share a central memory, in general there are specialized multi machine synchronous communication components, can support the development of data parallel or control [9]. But the processor number is too much, the processor to the central memory channel will become a bottleneck, limiting the development of the parallel machine, which is one of the main reasons for large-scale distributed memory parallel machine developed. The system of SMP is described as Fig. 5.

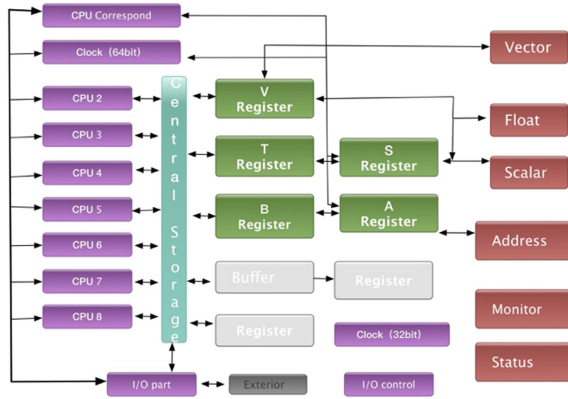


Fig. 4. The system of vector machine

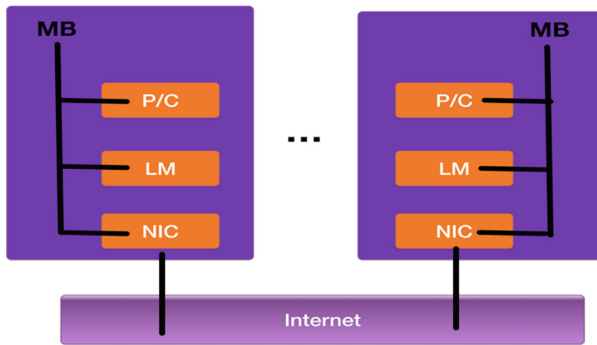


Fig. 5. The system of SMP

2.4 MPP

Distributed memory multiprocessor system which is composed of many parallel nodes, each node has its own processor and memory nodes connected to the interconnection network, parallel development support data also support the control of parallel development [10]. The system of MPP is described as Fig. 6.

2.5 COW

The workstation cluster of workstations (COW) is a collection of all computer nodes interconnected by high performance networks or local area networks [11]. Typically, each node is a SMP server, a workstation or a PC machine, which can be isomorphic or heterogeneous. The number of computers in general is a few to dozens, support for control of parallel and data parallel. Each node has a complete operating system, network software and user interface, can be used as a control node and computing nodes, that is equal between nodes. The cluster system's performance in recent years is

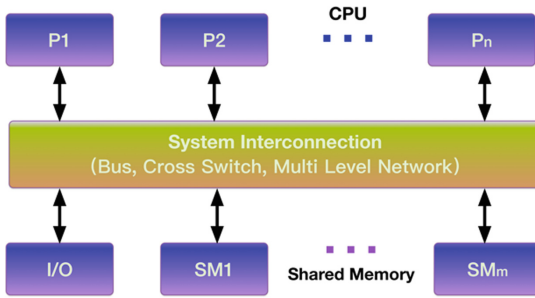


Fig. 6. The system of SMP

striking, because of its excellent performance, good flexibility and parallel processing ability, in addition to widely as a research topic, application development in various industries is also very fast. The system of COW is described as Fig. 7.

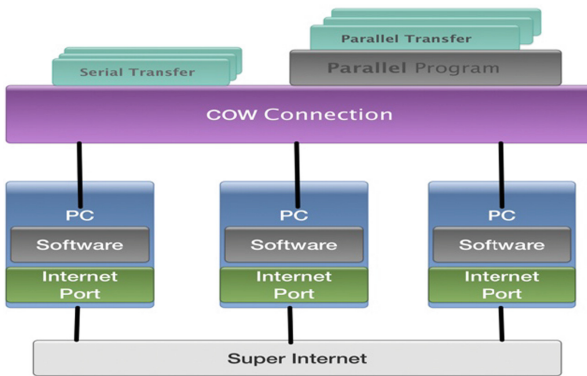


Fig. 7. The system of COW

3 Theoretical Model of Parallel Computing Technology

Parallel computing is the process of solving the problem of computing resources at the same time, it is an effective method to improve the computing speed and processing power of computer system [12]. Its basic idea is to use multiple processors to solve the same problem, the problem is decomposed into several parts, and each part is calculated by an independent processor. The parallel computing system can be either a specially designed super computer with multiple processors or a cluster of independent computers which are interconnected in a certain way. Through the parallel computing cluster to complete the data processing, and then return the results to the user.

The theoretical model of structure, the problem will be resolved is divided into N, N computing resources for the N runway, the problem is solved, and a huge problem can also be multiple computing resources to solve the basic model, as follows. In an

ideal situation, the time consumed by parallel computing is the formula, that is, each independent computing resource completes the task at the same time, the consumption time is the time to solve the problem. The ideal model of parallel computing is described as Fig. 8.

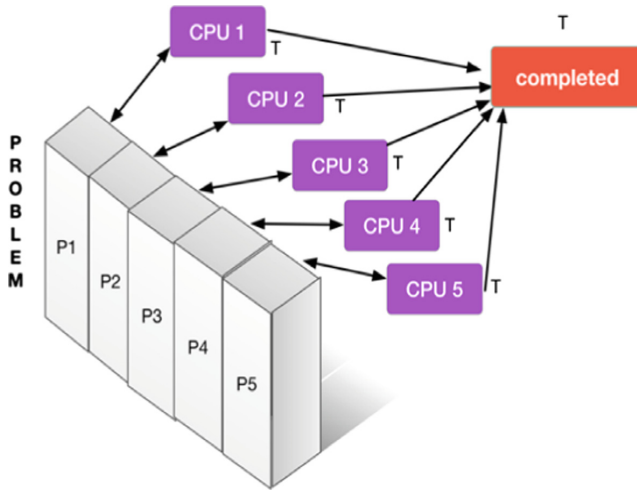


Fig. 8. The ideal model of parallel computing

According to the above parallel computing technology model, it can be known that the time consumed by the parallel computation is the slowest problem modules. The actual calculation may appear in many situations. First of all, the module partition problem, we can not guarantee that every module of the size of the problem is the same, assuming that dealing with computing resources ability is equal, this will lead to the time of computing resources to receive the largest part module significantly longer than other computing resources. It affects the efficiency of parallel computing computing. Assuming that the problem can be evenly divided into N module, if there is a single computing resource because of memory overflow or computational problems, this part module is stopped or delayed, resulting in increase of parallel computing time or can not complete the task. The unreasonable partition of problem model is described as Fig. 9. The abnormal CPU model is described as Fig. 10.

Based on the parallel computing cluster model, the problem to be solved by the main control machine is divided into N problem module, and then assigned to the N computer. In an ideal case, the size of each part module is the same as the computing power of each computer, and the ideal processing time for parallel computing is the time for a single computer to deal with the part module [13]. Assuming that the processing capacity of each computer is the same, but one problem of all is too large, the time consumed by the parallel computing is the time to solve the biggest problem. Assuming the master machine assigned to each computer of the same size, solve the problem in the process, if a computer or abnormal downtime, which leads to the

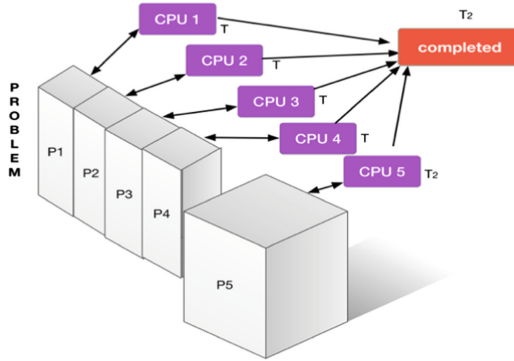


Fig. 9. The unreasonable partition of problem model

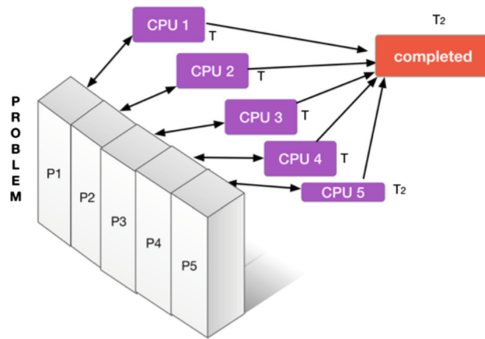


Fig. 10. The abnormal CPU model

problem of processing time is lengthened obviously or in the problem can not be solved in parallel computing.

4 Parallel Computing Technology Optimization

For the problems mentioned in the third chapter, there are many similar problems in the process of the actual parallel computation. Part problem segmentation is not reasonable, resulting in a single independent processor consumes too long, which greatly reduces the efficiency of parallel computing. The process of parallel computing, due to its single processor, the processing speed is relatively slow or midway accident downtime, so the calculation of the time was pro-longed or parallel because some calculation results did not reach a lead to the parallel computing can not be completed. The practical problem, parallel computing an obvious disadvantage is the repeated calculation, by part module problem segmentation in a lot of data and calculation methods are the same, the calculations have been repeated on different computers, it will reduce the efficiency of calculation greatly [14].

In real life, there are many examples about parallel, we can learn from the life of the solution to solve the problems encountered in parallel computing. For example, there is a pile of goods need to be transported from A place to B place, so we prepare a lot of goods vehicles, trucks are loaded cargo weight is not the same, the speed of different trucks carrying is not the same. When the last truck arrives at B place, the task is completed. In the process of transportation, if a truck is loaded with heavy goods, then its speed will be very slow. If one of the vehicles due to their own reasons for slower speed or the middle of the problem then arrives at the B place of time will be late or stop in the road which leads to the task can not be completed. In this situation, we can arrange a fast large trucks, to deal with similar problems occurred in the handling process, to ensure that the task is completed in high efficiency.

Parallel computing technology based on the practical problems, the optimization scheme proposed in this paper, the parallel computation with one or more advanced processor, the processor computing power was significantly faster than that of other processors. Parallel computing process, the main control computer if the layout of a computer to detect the problem is too large, the main control computer will arrange the task of the computer to the advanced computer to continue processing. In the course of parallel computing, if there is a problem with a single computer, the master computer will give the task of the problem computer to the advanced computer. Parallel computing in the process of marking method to calculate more than a certain period of time, the method to replant advanced computer, this time after the node calculation by computer, reduce repeated calculation times so as to improve the efficiency of parallel computing. The super CPU model is described as Figs. 11 and 12.

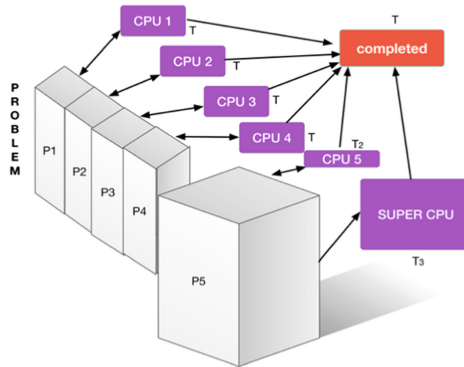


Fig. 11. The super CPU model

Algorithm steps:

- (1) Equal cutting problem $P, P(0), P(1), \dots, P(s), \dots, P(n-1)$
- (2) The task is assigned to each CPU, $C(0), C(1), \dots, C(s), \dots, C(n-1)$
- (3) Begin execution, Execute(P)
- (4) The control center to monitor the task, If it find a large task $P(s)$ to the Super CPU processing, cycle monitoring processing, until each subtask is almost equal.

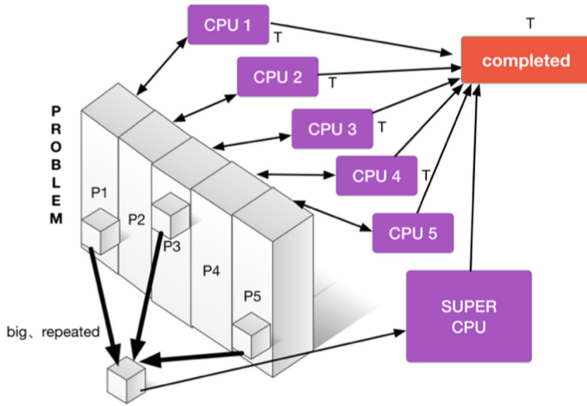


Fig. 12. The super CPU model

- (5) The control center real-time monitoring task, Monitor (task), if found in the task $P(s)$ is too large or abnormal CPU during execution, $(P(s))/(C(s)) \gg \bar{t}$, the task is assigned to the Super CPU processing
- (6) The main control center searches for duplicate parts in the subproblem, $P'(0) = P'(1) = \dots = P'(s)$, which is processed by the Super CPU and returns the value to each CPU module
- (7) Repeat steps 5 and 6, and the priority $V5 > V6$
- (8) Until the last process $P(s)$, handled by the Super CPU
- (9) Each sub-problem is resolved, the main control center integrated sub-questions solve, the task is completed

5 Conclusions

For the parallel computing technology, this paper proposes a solution to the problem about parallel computing from a new direction, and improves the original model. The scheme is suitable for most parallel computing technologies, especially for large data parallel computing technology, which can solve more problems in the future.

References

1. Chen, G.L., Sun, G.Z., Zhang, Y.Q., et al.: Study on parallel computing. *J. Comput. Sci. Tech.* **21**(5), 665–673 (2006)
2. Isard, M., Yu, Y., Birrell, A., et al.: Dryad: distributed data-parallel programs from sequential building blocks. Technical report, Microsoft Research Technical Report, Microsoft Corporation (2006)
3. Asanovic, K., Bodik, R., James, J., et al.: The landscape of parallel computing research: a view from Berkeley. Technical report, Electrical Engineering and Computer Sciences, University of California, Berkeley (2006)

4. Mattson, T.G., Sanders, B.A., Massingill, B.L.: *Patterns for Parallel Programming*. Prentice Hall, New Jersey (2005)
5. Rajkumar, B., Chee, S.Y., Srikumar, V.: Market-oriented cloud computing: vision, hype, and reality for delivering IT services as computing utilities. In: *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications*, 25–27 September 2008, Dalian, pp. 15–22. IEEE CS Press, Los Alamitos (2008)
6. Sun, X.H.: Scalable computing in the multicore era. In: *Proceedings of the Inaugural Symposium on Parallel Algorithms, Architectures and Programming*, 16–18 September 2008, pp. 1–18. University of Science and Technology of China Press, Hefei (2008)
7. Furtak, T., Amaral, J.N., Niewiadomski, R.: Using SIMD registers and instructions to enable instruction-level parallelism in sorting algorithms. In: *Proceeding of the 19th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)* (2007)
8. Pardo, M., Sberveglieri, G.: Classification of electronic nose data with support vector machines. *Sens. Actuator B: Chem.* **107**, 730–737 (2005)
9. Roig, C., Ripoll, A., Senar, M., Guirado, F., et al.: A new model for static mapping of parallel applications with task and data parallelism. In: *Proceeding of the International Parallel and Distributed Processing Symposium*, pp. 78–85 (2002)
10. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Commun. ACM* **51**(1), 107–113 (2005)
11. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. In: *Sixth Symposium on Operating System Design and Implementation*, 6–8 December 2004, San Francisco, CA, pp. 10–23. USENIX Association, Berkeley (2004)
12. Komatitsch, D., Goddeke, D., Erlebacher, G.: Modeling the propagation of elastic waves using spectral elements on a cluster of 192 CPUs. *Comput. Sci. Res. Dev.* **25**(1–2), 75–82 (2010)
13. Grama, A.Y., Gupta, A., Kumar, V.: Isoefficiency: measuring the scalability of parallel algorithms and architectures. *IEEE Parallel Distrib. Technol.* **1**(3), 12–21 (1993)
14. Ino, F., Fujimoto, N., Hagihara, K.: LogGPS: a parallel computational model for synchronization analysis. In: *Proceedings of the 2001 ACM SIGPLAN Symposium on Principles and Practices of Parallel Programming, PPOPP 2001*, Snowbird, Utah, USA, pp. 133–142. ACM (2001)