GRAPE-4: A TERAFLOPS MACHINE FOR N-BODY SIMULATIONS

MAKOTO TAIJI
Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan
E-mail : taiji@ism.ac.jp

Abstract. We have developed a massively parallel special-purpose computer system for N-body simulations, GRAPE-4 (GRAvity-PipE 4). The GRAPE-4 system is designed for high-accuracy simulations of dense stellar systems. The GRAPE-4 calculates gravitational forces, their derivatives in time and potential energies. It has a hardware for prediction of positions and velocities, which is used for the individual timestep scheme. We integrated 1692 chips of 640 megaflops performance to achieve the peak speed of 1.08 teraflops.

1. Introduction

Gravitational N-body simulation is one of the most important method for the study of star clusters. In N-body simulations, almost all computational time is spent in calculating gravitational forces, since the number of interactions between particles is proportional to the square of the number of particles. Thus, it requires huge computing resources beyond teraflops performance to simulate the post-collapse evolution of real clusters with $10^4 \sim 10^5$ particles. We have developed special purpose computer systems GRAPE (GRAvity PipE) to accelerate for N-body simulations of globular clusters, galaxies, cluster of galaxies, and the universe(Sugimoto *et al.* , 1990; Ebisuzaki *et al.* , 1993). In this paper we describe GRAPE-4, a teraflops massively-parallel special-purpose computer system for gravitational N-body simulations (Taiji *et al.* , 1994). GRAPE-4 is suitable for high-accuracy N-body simulations with the hierarchical timestep scheme.

GRAPE calculates only gravitational forces. The host workstation, which is connected to GRAPE, integrates orbits of particles. In the large N-body simulations, the force calculation, which costs $O(N^2)$ in time, dominates computational time. On the other hand, both the calculation in a host and the communication between a host and a GRAPE system cost $O(N)$. Therefore, commercial workstations can satisfy the requirements for the communication speed as well as for the calculation speed, although the calculation speed of GRAPE exceeds teraflops. The GRAPE systems can be also used to accelerate clever algorithms like the tree algorithm (Makino, 1991) or the Particle-Particle Particle-Mesh (PPPM) method (Brieu *et al.* , 1995).

2. What GRAPE-4 calculates

GRAPE-4 calculates forces, force derivatives in time, and potential energies from the positions, the velocities, and the masses. Force derivatives in time are necessary in fourth-order Hermite scheme, which is more simple and accurate than schemes based on the Newton interpolation (Makino and Aarseth, 1992). Here, the positions and the velocities at a block step are calculated from those at local time for particles using third order predictors. In the hierarchical timestep scheme, we have to evaluate predictors of *all* particles, while we only calculate forces and correctors of particles which share the block step. Therefore, predictor calculations become too expensive to be performed by the host computer as N is increased. Since we have to send the predicted coordinates of *all* particles, the communication between the host and GRAPE-4 is also increased. To solve these problems, we added a hardwired pipeline for predictor calculations to GRAPE-4.

J. Andersen (ed.), Highlights of Astronomy, Volume 11B, 600–602.
© 1998 *IAU. Printed in the Netherlands.*

https://doi.org/10.1017/S1539299600018244 Published online by Cambridge University Press

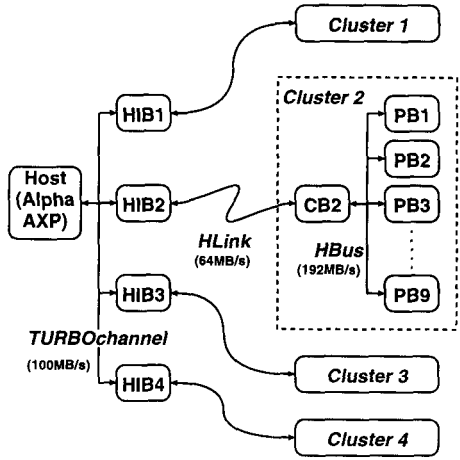

Figure 1. The block diagram of the GRAPE-4 system. HIB: host interface board, CB: controller board, PB: processor board.

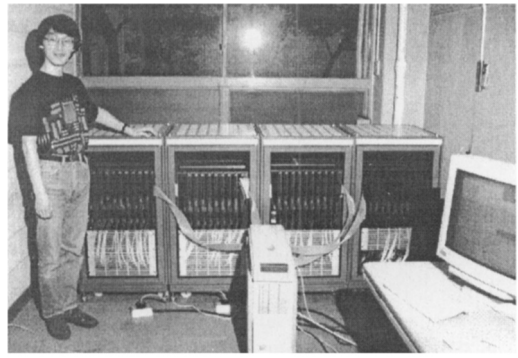

Figure 2. Photograph of the GRAPE-4 system

3. System Architecture

Figure 1 shows the block diagram of the GRAPE-4 system. It consists of a host computer, host interface boards (HIB), controller boards (CB), and processor boards (PB). This system has a two-level structure. It has four clusters, and each cluster has one CB and nine PBs. Each PB has 47 HARP (Hermite AcceleratoR Pipe) chips which calculates forces and their derivatives with the performance of 640 megaflops. The HARP chip has virtually two pipelines by time-shared operation. The system has 1692 HARP chips and the peak performance of 1.08 teraflops. We use Digital Equipment Corporation (DEC) Alpha AXP 3000 series workstation with a TURBOchannel bus as a host computer. Since it consumes only 10 kW of power, no heavy cooling system is necessary. It cost about 1.2 million dollars to build the hardware of GRAPE-4. Figure 2 shows a photograph of the GRAPE-4 system.

4. Why GRAPE is so fast?

Why the GRAPE systems could a-chive such a high performance at a low cost? There are two important differences in architecture compared with commercial MPP systems. The first one is

https://doi.org/10.1017/S1539299600018244 Published online by Cambridge University Press

the very-long-pipeline architecture, which leads high-performance LSI chips. The second one is the broadcast memory architecture.

First, we explain the advantage of the very-long-pipeline architecture. Nowadays we can pack $> 10^7$ transistors into one silicon LSI. Since a floating point multiplier and adder need less than $\sim 10^{5.5}$ transistors even in double precision, one LSI can have a few tens or hundreds of arithmetic units. It is very difficult to utilize so many arithmetic units efficiently. The very-long-pipeline architecture is one of answers to solve the problem. In this architecture, the whole force calculation are done in one or a few cycles using many arithmetic units. The HARP chip calculate one interaction, which usually requires $50 \sim 60$ operations, in three cycles. In other words, it performs about 20 operations per cycle. Thus its performance reaches 640 megaflops at 32 MHz. This is very high performance considering the fact that it is made by 1.0 μm technology in early 90s.

Next, we explain about the broadcast memory architecture. The most serious problem in recent computer systems is the memory bottleneck. The speed of processors increases quite rapidly but that of memories increases rather slowly. Therefore, it becomes difficult to achieve the enough bandwidth. Fortunately, for some applications like classical particle simulations we can overcome this difficulty by the broadcast memory architecture.

In the case of GRAPE-4, multiple pipelines on a processor board calculate forces on different particles in parallel. The coordinates and masses of particles are supplied from the predictor (memory) unit to the pipelines. Since these pipelines calculate forces on different particles, we can use the same coordinates for calculation of all these forces. Therefore, all the inputs of pipelines can share the common output from the predictor unit. Thus, the actual bandwidth between the predictor unit and the pipelines is only 0.2 Gbyte/sec, but virtually it reaches 20 Gbyte/sec since the data are broadcasted to all 94 parallel pipelines. Such a memory architecture is not used in commercial general-purpose computers, since it can be applied only for some specific applications like classical force calculations.

5. Future prospects

The GRAPE architecture will become more important in future. General-purpose computers cannot utilize efficiently the increasing amount of transistors. On the other hand, GRAPE can use all of them to increase arithmetic units. The new GRAPE-6 project (Makino, in this volume) will achieve performance far beyond those of general-purpose parallel processors.

It is also possible to develop a semi-special-purpose computer for particle simulations. Since the most important advantage of the GRAPE systems is the broadcasting memory architecture, a parallel computer with this architecture will be very useful for particle simulations. Such a computer is also useful for other calculations like Gaussian elimination or Genome data mining.

ACKNOWLEDGMENTS

This work was partially supported by the Grant-in-aid for Specially Promoted Research (04102002) of the Ministry of Education, Science, and Culture.

References

Brieu, P. P., Summers, F. J., and Ostriker, J. P. 1995. *Astrophys. J.*, in press.
Ebisuzaki, T., Makino, J., Fukushige, T., Taiji, M., Sugimoto, D., Ito, T., and Okumura, S. K. (1993) *Publ. Astron. Soc. Japan*, **45**, 269.
Makino, J. (1991) *Publ. Astron. Soc. Japan*, **43**, 621–638.
Makino, J., and Aarseth, S. J. (1992) *Publ. Astron. Soc. Japan*, **44**, 141.
Sugimoto, D., Chikada, Y., Makino, J., Ito, T., Ebisuzaki, T., and Umemura, M. (1990) *Nature*, **345**, 33.
Taiji, M., Makino, J., Fugushige, T., Ebisuzaki, T., and Sugimoto, D. (1996) *Pages 141–150 of: Dynamical Evolution of Star Clusters, Proceedings of IAU Symposium 174.*

https://doi.org/10.1017/S1539299600018244 Published online by Cambridge University Press