PAPIA (Parallel Protein Information Analysis) System and MolTreC Parallel Molecular Dynamics Simulator Running on a Compact 8-node Linux PC Cluster

Yutaka Akiyama [1] akiyama@rwcp.or.jp
Tamotsu Noguchi [1] noguchi@rwcp.or.jp
Kentaro Onizuka [12] onizuka@rwcp.or.jp
Makoto Ando [13] ando@rwcp.or.jp

[1] Parallel Application Laboratory, Real World Computing Partnership (RWCP), Tsukuba Mitsui Bldg., 1-6-1 Takezono, Tsukuba 305-0032, Japan
[2] Matsushita Research Institute Tokyo, Higashimita 3-10-1, Tama-ku, Kawasaki 214-8501, Japan
[3] NKK Corporation, 1-1-2 Marunouchi, Chiyoda-ku, Tokyo 100-0005, Japan

1 Introduction

We have successfully ported our parallel application programs, the PAPIA system [1] and the MolTreC no cut-off molecular dynamics simulator [2] onto a new compact 8-node Linux PC cluster.

The PAPIA (Parallel Protein Information Analysis) system [1] is an integrated parallel computing system dedicated to typical search/analysis problems required in protein research, such as structure similarity search, sequence homology search, and multiple sequence alignment. The system was implemented on a 64-node PC cluster (PAPIA Cluster) and we have been operating a WWW-based free service since April 1998 (`http://www.rwcp.or.jp/papia/`). Now the PAPIA system also runs on a compact and inexpensive 8-node cluster which we will demonstrate at GIW99 (Fig. 1, Table 1).

Figure 1: New 8-node PAPIA mini Cluster (left), and 64-node PAPIA Cluster (right).

MolTreC (Molecular Dynamics by Tree Code) [2] is our original parallelized MD program which efficiently performs no cut-off accurate calculation of Coulomb potentials using the Barnes-Hut tree code algorithm and the multi time-step method. MolTreC also runs on the PAPIA and now on the PAPIA mini cluster. With the SCore multi-user parallel environment developed by RWCP [3], we can even run the PAPIA system and MolTreC (or any other application) simultaneously on a cluster.

2 Performance of PAPIA mini Cluster

Fig. 2 shows the effectiveness of parallel computation on our PC clusters: the 8-node PAPIA mini Cluster and the 64-node PAPIA Cluster. The left graph shows the execution time of the parallel multiple alignment program provided in PAPIA WWW, for aligning 64 kinase sequences. The right graph shows the execution time for 100 simulation steps of MolTreC, on the well-known PROWAT

Table 1: Hardware Specifications of Two PC Clusters.

	PAPIA mini Cluster	PAPIA Cluster [1]
No. of Nodes	8 nodes	64 nodes (+2 monitor nodes)
Processor	Dual Pentium II, 450MHz (SpecInt = 16.9)	Pentium Pro, 200MHz (SpecInt = 8.2)
Memory	512MB/node	256MB/node
Disk	6.3GB/node	4.1GB/node
Network	Myrinet 1.28Gbit/sec. 100-BaseT	Myrinet 1.28Gbit/sec. 100-BaseT
Physical Size	H635×W770×D635 mm	H1600×W1600×D800 mm

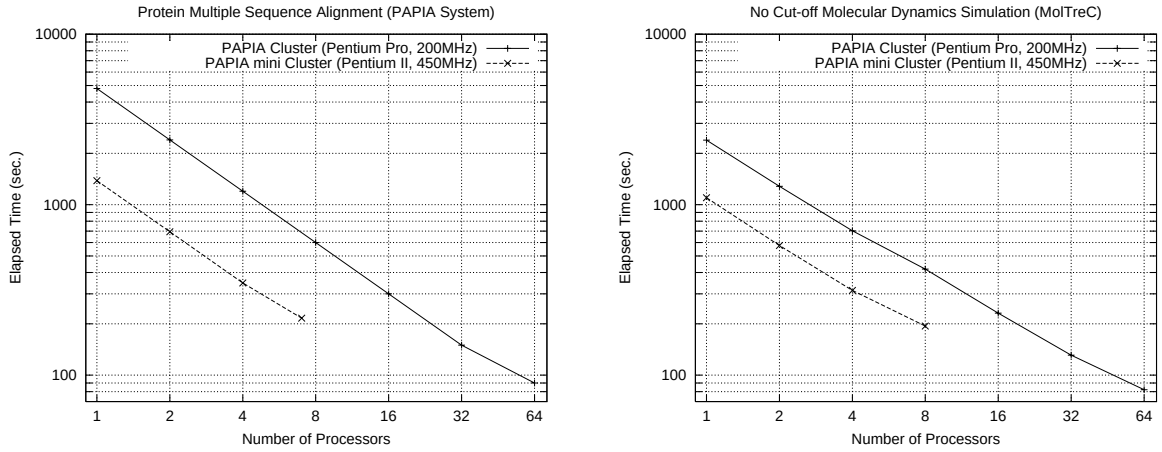

Figure 2: Comparison of Execution Time between the PAPIA mini and the PAPIA Cluster: Protein Multiple Sequence Alignment, 64 kinase sequences with 80 amino acids. (left); and MolTreC No Cut-off MD Simulation, PROWAT (plastocyanin) 11585 atoms, for 100 steps. (right); log-log scale.

(plastocyanin in water) benchmark with 11585 atoms. In both graphs, the horizontal axis shows the number of worker processors and the vertical axis shows the execution time in seconds (log-log scale).

For the multiple alignment (left), the PAPIA mini cluster is more than 3-times faster than the old cluster, while the CPU itself (see SpecInt value) is only twice as fast. This comes from the improved bus/memory performance of the Pentium II. In the no cut-off molecular dynamics (right), the performance gap of the two clusters is smaller because of a fixed cost, heavy communication load.

The PC cluster approach is an easy and quick way to fully utilize the power of the latest microprocessors. PC clusters will become a powerful but inexpensive tool for bioinformatics researchers.

References

[1] Akiyama, Y., Onizuka, K., Noguchi, T., and Ando, M., Parallel protein information analysis (PAPIA) system running on a 64-node PC cluster, *Proc. the 9th Genome Informatics Workshop*, 131–140, 1998.

[2] Akiyama, Y., Onizuka, K., Noguchi, T. and Ando, M., Biological- and chemical- parallel applications on a PC cluster, *Lecture Notes in Computer Science*, Springer-Verlag, 1615:220–233, 1999.

[3] Ishikawa, Y., Tezuka, H., Hori, A., Sumimoto, S., Takahasi, T., O'Carroll, F., and Harada, H., RWC PC cluster II and score cluster system software – high performance Linux cluster, *Proc. of the 5th Annual Linux Expo*, 55–62, 1999.