Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

A Commodity Computing Cluster

P. J. Teuben, M. G. Wolfire, M. W. Pound, L. G. Mundy

Astronomy Dept, University of Maryland, College Park

Abstract. We have assembled a cluster of Intel-Pentium based PCs running Linux to compute a large set of Photodissociation Region (PDR) and Dust Continuum models.

For various reasons the cluster is heterogeneous, currently ranging from a single Pentium-II 333 MHz to dual Pentium-III 450 MHz CPU machines. Although this will be sufficient for our "embarrassingly parallelizable problem" it may present some challenges for as yet unplanned future use. In addition the cluster was used to construct a MIRIAD benchmark, and compared to equivalent Ultra-Sparc based workstations.

Currently the cluster consists of 8 machines, 14 CPUs, 50GB of disk-space, and a total peak speed of 5.83 GHz, or about 1.5 Gflops. The total cost of this cluster has been about $12,000, including all cabling, networking equipment, rack, and a CD-R backup system.

The URL for this project is `http://dustem.astro.umd.edu`.

1. Introduction

Commodity PC hardware components with the virtually freely available Linux operating system now provide a viable alternative compute engine (cf. Beowulf clusters) to the traditional supercomputer and desktop arrangements which most astronomy departments were using up until recently.

Approximately 10^5 PDR and Dust Continuum models are needed (Pound et al. 2000), each of them requiring about 20 Gflop's. A typical 10 node Pentium-II cluster can currently reach about 1 Gflops, and would provide us with a compute engine that could produce these models in a few months time. Each model generates a small amount (about 15kB) of data, which is copied back to the server, a Solaris machine on which the web server is running providing Java clients with a visual interface to the data (Pound et al. 2000).

2. Hardware

For a variety of reasons (ramp up, using Moore's law to our advantage, funding cycles in a multi-year project) we decided to build up the cluster over several months. This of course means that one winds up with a non-homogeneous cluster, which may be a handicap for codes that require more symmetric CPUs. In addition, maintenance can be complicated because of different components (video cards, ethernet cards, etc.). Our first two machines (a single CPU Pii-

400 and Pii-333) arrived in Spring 1998, and later that year three dual Pii-400s arrived. In Spring 1999 we acquired three more dual Piii-450s, and then finally assembled the 8 boxes/14 node cluster in a rack.

The machines have been networked together with fast (100 Mbit) ethernet and a simple hub, since we did not need more sophisticated I/O other than transferring the final results to a central database. We also keep the cluster on a private network, not only for security reasons, but also to limit public IP usage. In a more serious Beowulf-type cluster, switches with Gigabit networking would be employed to provided faster interprocess communication.

3. Software

There are a variety of ways to software "glue" individual workstations together (all methods are not limited to Linux PCs). These will take care of issues like load-balancing, shared memory across machines etc.

- Beowulf: a CESDIS/NASA project, started with 16 i486 nodes in 1994 (Sterling & Becker),
- MOSIX: Linux kernel adaptations to improve clustered load-balancing, memory usage etc.,
- PVM (Parallel Virtual machine), MPI (Message Passing Interface), MPI-ii, LAM (Local Area Multicomputer), DIPC (Distributed Inter-Process Communication) BSP (Bulk-Synchronous Parallel), LINDA,
- good compilers gcc/g++/g77, but excellent commercial alternatives available (Absoft, NAG, Paralogic, PGI),
- scheduling software (EASY, DQS, Condor),
- Visualization: Promising new technology from the gaming market! (TNT2, Vodoo2: OpenGL/GLX, MESA).

4. Benchmarks: MirStones

The MIRIAD package (Sault et al. 1995) did not have a standard benchmark (cf. DDT in AIPS) or baseline test. A new benchmark was thus devised (MirStones) to compare different compilers, different Linux distributions, various types of hardware and other operating systems under similar hardware configurations (Teuben 2000) Version 1 of this benchmark tests basic radio interferometric data computing and manipulation (mapping and deconvolution) and produces a modest 350MB of data. A typical benchmark takes 3-8 minutes on a modern workstation, and a MirStone is normalized to unity when the benchmark takes 5 minutes.

Here are the basic ingredients to the benchmark, in MIRIAD terminology:

- `uvgen`: Generate 3 model visibility datasets, representing different array configurations (9.2MB each)
- `uvcat`: Catenate different visibility datasets (final 26.4 MB)
- `invert`: Map (32 channels) making (map 131.6MB, beam 4.1MB), basically a Fourier transform operation
- `clean`: Deconvolution (Hogbom/Clark/Steer CLEAN) (clean 32.9MB)
- `restor`: Restore clean components into a clean map (cmap 131.6MB)

The following table and figure summarizes and compares a few other popular benchmarks with *MirStones*:

Table 1. Sparc/Solaris and Intel/Linux benchmarks

Machine	Freq MHz	SPEC fp95	SPEC int95	A_m^{93} AipsMark	MirStone
Pmmx	166		5.6	3.5	0.28
Pmmx	266				0.31
Pii	333			10.4	0.63
Pii	400	12.4	15.3	14.5	0.94
Piii	450				1.02
Piii	500	14.7	20.6	17.9	1.12
Piii-X	550			21.6	1.32
AMD-K7	900				1.81
Sparc-4/50 IPX	40			1.0	0.042
Sparc10/30	30				
Sparc10/40	40	1.00	1.00		
		1.38	1.13		
Ultra-1	170	9.06	6.26	6.3	0.32
Ultra-2	200	-	-		0.67
Ultra-10	300	12.9	12.1	8.8	0.81
Ultra-10	360	19.9	15.2		1.19
Ultra-10	440	22.7	17.9		

Acknowledgments. This project was supported in part by a NASA ADP grant NAG5-6750 and by a NASA-AMES Cooperative agreement NCC2-1058.

References

Pound, M.W. et al. 2000, this volume, 630

DDT, (Dirty Dozen Test), Aips memo[1] 85

Sault R. J., Teuben, P. J., & Wright, M. C. H. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 433

Teuben, P. J. 2000, BIMA memo[2] (in preparation)

Rudolph, A. & Teuben, P. J. 1991, BIMA memo 11.

[1] `http://www.cv.nrao.edu/aips/ddt.html`

[2] `http://bima.astro.umd.edu/memo`

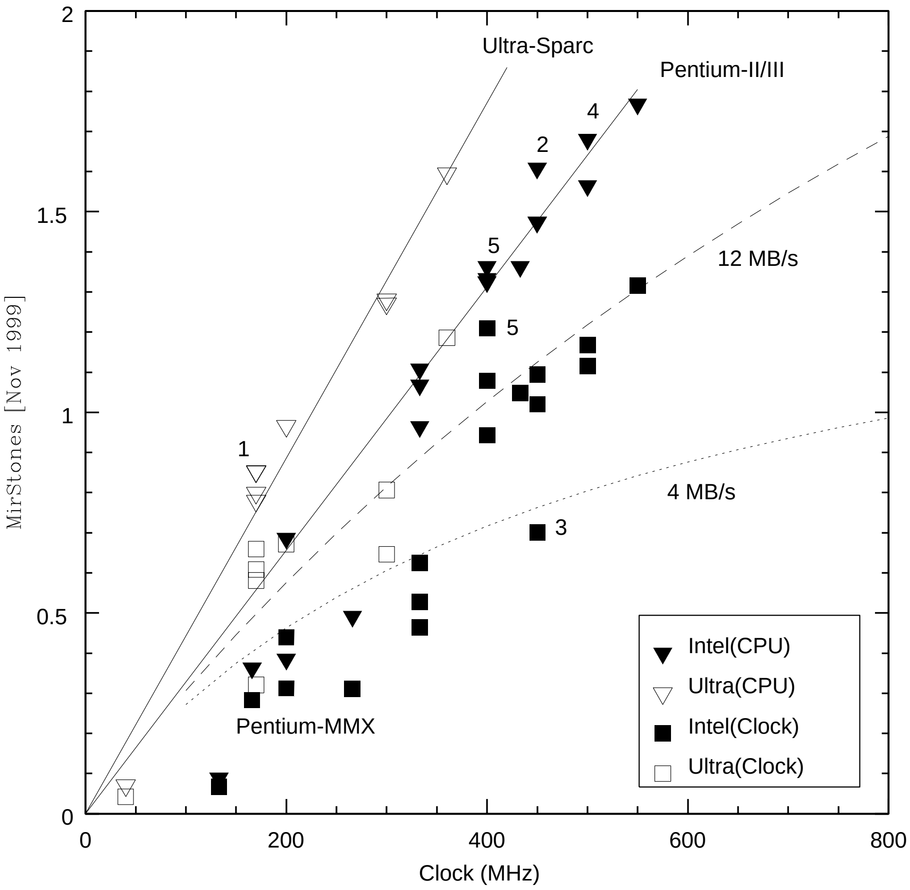

Figure 1. MirStones: square symbols represent the actual MirStone, a wall-clock time based measurement (1 represents 5 minutes elapsed time), whereas the triangles measure pure CPU ignoring any system and other I/O needed. The solid lines are linear estimates of the average Sparc and Pentium CPU performance, whereas the dashed and dotted curves are expected MirStone curves for an effective disk I/O performance of 12 and 4 MB/s resp. Various deviations from the norm should be noted: (1) Sparc optimized compiler; (2) Pentium optimized compiler; (3) non-UDMA disk; (4) Athlon K7; and (5) SCSI/Xeon.