Mem. S.A.It. Suppl. Vol. 13, 18

Memorie della
Supplementi

Report on the INAF-CINECA agreement

G. Bodo

Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Torino, Via Osservatorio 20, I-10025 Pino Torinese, Italy, e-mail: bodo@oato.inaf.it

Abstract. I will discuss what are the computational resources available to the italian astronomica community through the agreement between INAF and CINECA and how they have been used. I will try to characterize the astronomical computational community and what are his needs.

Key words. Numerical Simulations, High Performance Computing

1. Introduction

Numerical simulations have become essential for interpreting and understanding astrophysical phenomena. A far from exhaustive list of problems in which they are one of the fundamental tools of investigations includes the dynamics of large scale cosmological structures, the formation and evolution of galaxies, accretion processes, high energy phenomena, the dynamics of the interstellar medium, star formation, supernovae, solar physics and planetary formation. The need for simulations derives from the complexity and high nonlinearity of the processes involved in the study of the physical structure of cosmic objects.

The multiplicity of the physical processes together with the huge dynamic range of spatial and temporal scales of astrophysical systems leads to large requirements in term of computational resources, and High Performance Computing (HPC) tools have become essential. This makes clear that, for keeping the pace with the international competition, the availability of HPC resources to the italian astronomical community is fundamental. As an answer to this, in 1997 the "Consorzio Nazionale per l'Astronomia e l'Astrofisica" (CNAA) stipulated an agreement with the CINECA supercomputing center for giving access to CINECA HPC resources. With that agreement 100000 CPU hours were dedicated to the national astronomical community. The groups that were already active in the field of HPC benefitted from the stable availability of computational resources, with the possibility of maintaining and developing their competences in the field, in addition a growth of the computational community has been also observed.

In 2002 the agreement was renewed by INAF, the last agreement ended in 2007 and a new one has been stipulated for the period 2008-2010. In this report I will discuss the last agreement that covered the years 2006-2007, making also comparisons with the previous period (see the report presented by the management committee for the agreement 2002-2005; Capuzzo Dolcetta et al. (2004)). In particular I will discuss what were the resources available and how they have been used.

Send offprint requests to: G. Bodo

Table 1. Evolution of resources

Year	CPU hours	Peak perf.
1997-2001	100 000	1
2001-2005	300 000	15
2006-2007	800 000	60
2008-2010	1 720 000	204

2. The resources

With the INAF-CINECA agreement "High Performance computing and archiving of simulation results in astrophysics" for the years 2006-2007, 800000 CPU hours were reserved on CINECA HPC systems for the astronomical community. We can start by examining the evolution of the available resources from 1997 (the date of the first agreement) given in Table 1. The Table includes also the new agreement (2008-2010). The second column gives the number of allocated CPU hours and the third column gives a relative measure of the resources taking into account the increase of the processor power. The data in this last column make use of the peak performance of the systems derived from the Top 500 lists (http://www.top500.org).

Of course the increase observed in Table 1 reflects the general increase in computing power, however the increase of our resources is somewhat lower than the general increase. Fig. 1 (http://www.top500.org) shows the world evolution of the performances of HPC systems. The top curve (green) shows the integrated performance of the 500 most powerful systems, the middle (red) curve shows the performance of the top system and the lowest curve (violet) shows the performance of the 500th system. All the curves give an increase of a factor 1000 from 1997, while our resources, as shown in Table 1 have increased by a factor 200. From another perspective, if we look at the position of the top italian system in the Top 500 lists (http://www.top500.org) we observe a drop from the 30th rank in 2002 to the 225th rank in June 2006. Afterwards we have a rise to 47th rank in 2007. In the new list for June 2008, the first italian sustem ocupies the 90th rank.

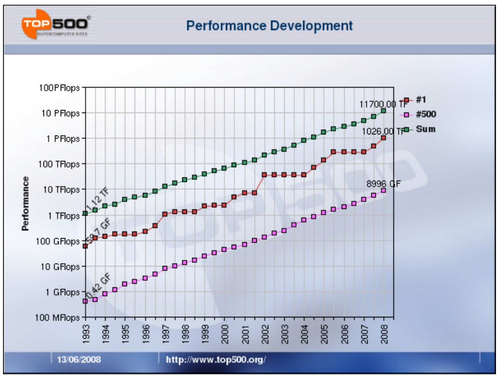

Fig. 1. Worldwide evolution of HPC systems.

I will now examine in detail the resources available through the agreement and how they have been allocated. The main systems available in the period considered were an IBM SP5 with 512 processor elements, 1TB RAM and a peak performance of 5 Tflops and an IBM Cluster of 1024 processor elements, 1 TB RAM and a peak performance of 6 Tflops. The computational resources have been allocated to the projects by the scientific committee appointed by INAF. There are three kinds of projects: key projects, standard projects and development projects. The objective of key projects is to provide a substantial allocation for a small number of high-impact applications. Standard projects are intended for applications with a lower request of resources, finally development projects are for development and testing of numerical codes. The definition of the three project types in term of the minima and/or maxima of the resources that can be requested has been changed following the evolution of the available computational power. For the period that we are examining, development projects had to require less than 5000 CPU hours and less than 8 processors, standard projects had to require less than 20000 CPU hours and less than 64 processors, key projects needed to require at least 100000 CPU hours and a use of 64-128 processors. Standard projects have been allocated with two calls per year and key projects with one call per year, development projects, instead, could be presented at any time.

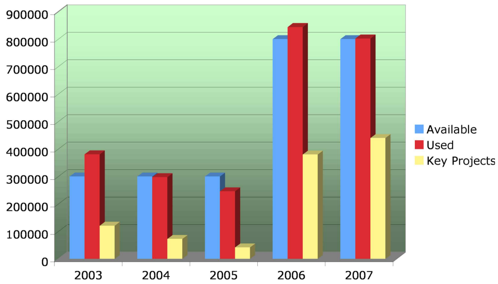

Fig. 2. Temporal distribution of the usage of computational resources

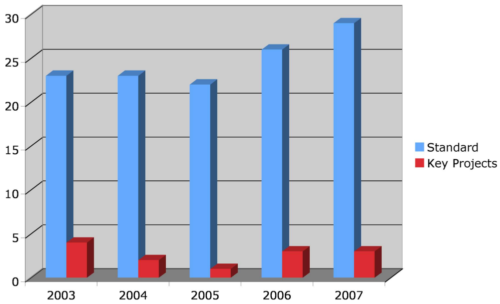

Fig. 3. Temporal distribution of the number of projects.

3. The use of the resources

In Fig. 2, I show for the last five years the time allocated, the time effectively used and the time used in the key projects. We see an increase of the fraction of time allocated to key projects, that has reached 50% of the available time.

In Fig. 3, I show the number of projects for which resources have been allocated, from the figure we can see a slight increase in the last two years. In the period considered, the total number of people involved has been about 80, with 41 different PI's. There have been three key projects both in 2006 and in 2007, involving six different groups. Table 2 gives data on the evolution of the size of key projects. In the second column it is shown the average size in CPU hours of the projects and in the third column it is shown the maximum size in CPU hours. Note that in 2004 and 2005 there has been only one key project. The average size of the projects in term of data produced has been of several TBytes.

Table 2. Size of key projects

Year	Average CPU hours	Maximum CPU hours
2003	30 000	40 000
2004	40 000	40 000
2005	40 000	40 000
2006	125 000	150 000
2007	150 000	200 000

Table 3. Size of standard projects

Size CPU hours	Fraction
0 - 4 000	0.00
4 000 - 8 000	0.18
8 000 - 12 000	0.26
12 000 - 16 000	0.18
16 000 - 20 000	0.22

I can give some indication on the trends considering the partial data on the allocations for 2008. The fraction of time allocated to key projects is 57%, the average size is 333000 CPU hours and the maximum size is 400000 CPU hours (which is the maximum size allowed for 2008).

In Table 3, I examine the distribuzion of size (in CPU hours) of standard projects. The second column gives the percentage of projects with size given by the first column. We can see a relatively flat distribution and projects of different sizes are present.

The data discussed above point out that on one side several groups need substantial amounts of computational resources and top level HPC systems, on the other side, however, there are projects that require smaller amounts of resources.

We can now move to examine the kind of astrophysical problems that have been considered. The numerical codes used by the majority of projects belong either to the class of gravitational N-Body codes or to the class of codes for astrophysical plasmas in the fluid or magnetofluid approximation. In Table 4, I show the number of projects that belongs to these two

Table 4. Distribution of code types.

Code Type	N. projects
N-Body	38
MHD	34
Other	13

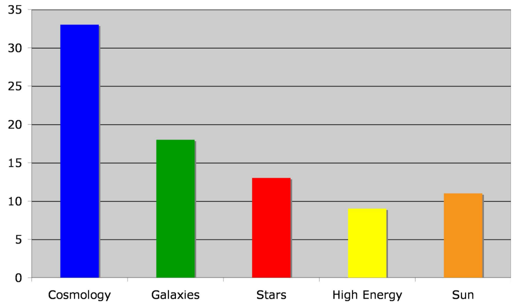

Fig. 4. Distribution of the number of project per scientific area.

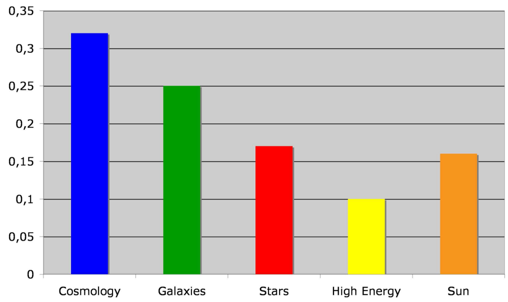

Fig. 5. Distribution of the fraction of the total CPU time per scientific area.

classes, a smaller fraction of projects makes use of other kinds of numerical codes.

Figs. 4 and 5 show respectively the number of projects and the fraction of computing times for different classes of astrophysical problems considered. Both figures show a predominance of cosmology projects and a relatively even distribution among other topics.

Finally Table 5 gives the distribution of projects for different sites, showing a wide geographical distribution of computational groups belonging either to the Observatories or to the Universities or being mixed groups.

Table 5. Geographical distribution of computational groups.

Site	N. projects
TO	7 + 1kp
MI	6 + 1kp
BO	14
PD	8
FI	11 + 1kp
PA	9 + 2kp
CT	5
CA	4
RM	9
TS	5 + 1kp
NA	3
PI	1

4. Conclusions

In this report, in addition to a discussion on the available resources, I have tried to characterize the computational community, how big it is and what are his needs. For concluding I would like to point out that for carrying out a computational project, in addition to the computing time, there are several other important needs. Simulations produce very large amount of data, so there is a problem of tranferring, storing, visualizing and analysing those data. In addition we have to consider the development of the complex numerical codes used for the calculations. Part of the codes that have been used at CINECA are public codes, that, however have sometimes been modified for suiting them to the user needs. Another part is made by locally developed codes. I would like to emphasize that during the last years the italian astronomical computational community has developed a big expertise in numerical schemes and codes both in the field of gravitational N-Body and in the field of plasma simulations (Orlando et al. 2003; Londrillo & Del Zanna 2004; Mignone & Bodo 2005; Del Zanna et al. 2007; Mignone et al. 2007; Maio et al. 2007).

Finally, let me mention the EU project PRACE (Partnership for Advanced Computing in Europe), that started this year and has the goal of laying the foundation for a future European supercomputer infrastructure. It is essential for the computational community that the italian system stays at pace with these european developments in the field of HPC.

Acknowledgements. I would like to thank the components of the scientific committee: V. Antonuccio, G. Bono, A. Chieffi, C. Gheller, G. Murante, F. Reale and G. Tormen.

References

Capuzzo Dolcetta, R., Antonuccio, V., Erbacci, G. & Gheller, C. 2004, Rapporto sulla Convenzione INAF-CINECA

Londrillo, P., & del Zanna, L. 2004, Journal of Computational Physics, 195, 17

Maio, U., Dolag, K., Ciardi, B., & Tornatore, L. 2007, MNRAS, 379, 963

Mignone, A., & Bodo, G. 2005, MNRAS, 364, 126

Mignone, A., Bodo, G., Massaglia, S., Matsakos, T., Tesileanu, O., Zanni, C., & Ferrari, A. 2007, ApJS, 170, 228

Del Zanna, L., Zanotti, O., Bucciantini, N., & Londrillo, P. 2007, A&A, 473, 11

Orlando, S., Peres, G., Reale, F., Rosner, R., Plewa, T., & Siegel, A. 2003, Memorie della Società Astronomica Italiana Supplement, 1, 45