ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2010, *55*(11-12), 1009-1015

Dedicated to the memory of
Professor Ioan Silaghi-Dumitrescu (1950 – 2009)

HIGH PERFORMANCE COMPUTING INFRASTRUCTURE FOR MATERIALS SCIENCE

Viorel CHIHAIA,[a]* Emil SLUSANSCHI,[b] Alexandru HERISANU,[b] Petru PALADE[c] and Nicolae TAPUS[b]

[a] Institute of Physical Chemistry "Ilie Murgulescu" of the Roumanian Academy, Splaiul Independentei 202, Sector 6, 060021 Bucharest, Roumania; E-mail: vchihaia@gmail.com
[b] Computer Science and Engineering Department, University Politehnica of Bucharest; Splaiul Independentei 313, Sector 6, 060042 Bucharest, Roumania; E-mail: emil.slusanschi@cs.pub.ro, heri@cti.pub.ro, nicolae.tapus@cs.pub.ro
[c] National Institute of Materials Physics, Atomistilor str. 105 bis, PO Box MG. 7, 077125 Magurele, Bucharest, Roumania; E-mail: palade@infim.ro

Received March 30, 2010

Computer Material Science is a branch of material sciences that uses theoretical concepts and models from chemistry, physics, geology, mineralogy, as well as biology and includes them into software applications in order to calculate the structure and the properties of molecules, gases, liquids, soft and condensed matter. This paper presents the estimated performances of the HPC-ICF cluster of the Institute of Physical Chemistry "Ilie Murgulescu" of the Romanian Academy and the NCIT-UPB Cluster developed by the Computer Science Department of the University Politehnica of Bucharest, with the run in the parallel mode on these systems of the **HPL – High Performance Linpack Benchmark** as well as of the electronic structure codes GAMESS and ONETEP. The results of the study show that both clusters break the Teraflop barrier, thus making available a valuable tool for the Romanian scientific community in Computing Materials Science.

INTRODUCTION

The computers help the scientists in the organization, the management and the rationalization of the obtained data from experimental measurements or from theoretical estimations. Furthermore, the modern computer infrastructure, hardware and software, may support them to evaluate various physical-chemical parameters based on the equation of states and models of a given material system. This field has become very consistent in the last fifteen years and has got into a mature investigation instrument in the material investigation called Computer Material Science (CMS).

CMS has an interdisciplinary character as it uses various concepts and methods from various fundamental sciences (mathematics, physics, chemistry, geology, biology, crystallography, mineralogy) and computer science (programming technique, languages, operating systems, and computers networking). Nowadays, CMS plays an important role for all aspects of material science, especially concerning the dynamic domains like surface science, nanoscience, electronics, biology, and drug design. It establishes a direct relationship between structure and properties, delivering theoretical solutions that may guide to synthesize new advanced materials, with new properties. This field has evolved in such way that it can determine with high accuracy the values of different properties with microscopic origin, making possible the analysis of the phenomena and the

* Corresponding author: vchihaia@icf.ro

description of the mechanisms from material science. CMS allows the access to some parameters unavailable experimentally (energy density, electronic density, density of states). It facilitates the access to pressure and temperature domains inaccessible by experiment and permits the analysis of macroscopic parameters function of the contribution of different regions and/or components of the investigated system.

CMS is the scientific field which requires the most computational power: the bigger computational resources allow the use of higher level approaches in order to achieve a greater accuracy, as well as the study of the phenomena which occur at larger scales of time and of space. The computational effort can be reduced by carefully choosing of theoretical models and algorithms. It requires the usage of good computational environment: computational speed, large amount of memory, efficient numerical libraries, parallel running. By using more powerful computers, developing efficient algorithms and using other programming methods the computational effort can be reduced. Significant experience in developing algorithms (adapted to a parallel programming) and programs has been gained. The entire capacity of the modern computers with parallel architecture can be used. The parallel programming is focused on partitioning the problems in jobs for each processor and synchronizing the jobs.

The access of the Roumanian researchers to powerful computers was restricted before '90s due to the embargo imposed by the western countries. Today, the access to supercomputers is financially limited because of their very high prices. Fortunately, there are lower cost solutions such the developing of the so-called computer cluster (a group of usually identical personal computers) or of the blade systems that is a more energetic efficient solution. Recently, the Teraflop-capable hardware computing systems[1,2] become a reality in the academic and research institutions from Romania.

Prof. Dr. Ioan Silaghi-Dumitrescu (project manager – Babes-Bolyai University, Cluj Napoca) and Dr. Ovidiu Nemes (project co-coordinator – Technical University, Cluj Napoca) created the Center for Molecular Modeling and Computational Quantum Chemistry.[3] The center is based on a cluster composed of 70 IBM blade servers with XEON processors, each server having two quad-core processors, 8 GB of RAM and 2x146 GB HDD 10 K in two racks that also contain a storage system (of 2.1 TB) and a communication server. The link between the two racks is through UTP cable at 1 MBps, and the communication between the two partner-institutions is made through optical fibre.[4] The main objective of this project was the building of a cluster/grid calculation platform, in order to use it in modeling on the molecular, meso- and macro-scopic level of complex materials.

In parallel, within the framework of the project ASSG - Virtual Atomic Scale Simulation Group in Material Science[5] a network of three High Performance Computing Centers dedicated to CMS has been developed as integrated platform of three clusters based on IBM Blade nodes IBM: 64 at the HPC-ICF Cluster[6] developed at Institute of Physical Chemistry "Ilie Murgulescu", Roumanian Academy (project manager: Dr. Viorel Chihaia); 32 at the NCIT-UPB Center[7] at Faculty of Automatic Control and Computers, Polytechnic University of Bucharest (project co-coordinator: Prof. Dr. Ing. Nicolae Tapus) and 8 at the National Institute of Materials Physics, Magurele (project co-coordinator: Dr. Petru Palade). The general objective of the project is the development of the CMS facilities made available for the scientific community, considering the existing infrastructure (computer networks, software for simulation and visualization, documentation) and the experience of the group members. ASSG is designed as an open group for new members and it will try to play an important role in gathering together the scientists who work using similar methods and algorithms in different fields and to exhibit to the Roumanian scientific community the abilities and the advantages of the atomic-scale simulations.

In this work we will present the results obtained by a tuned version of the HPL benchmark[8] on the HPC-ICF and NCIT-UPB clusters. Additionally, we show first successful tests of the powerful electronic structure codes GAMESS[9] and ONETEP,[10] which were executed in parallel mode on these platforms.

RESULTS

1. ASSG infrastructure

The clusters HPC-ICF and NCIT-UPB are built of identical solutions and components, based on IBM Blade Centre H21 chassis. The clusters are thus composed of 64 and 32 blade systems, respectively. Each HS21 blade contains two Intel's

Xeon E5405 Quad-Core chips running at 2 GHz, and has 16 Gb of main memory at its disposal and 146 Gb HDD. Fig. 1 depicts the topology of the NCIT-UPB cluster. Each Blade Centre has two I/O Cisco network modules with 8 (2*4) external gigabit ports connected in a backbone 4948 Cisco Gigabit Ethernet switch. All blades have dual Ethernet cards connected through different Cisco/DLink switches. The performance of the node interconnect relies heavily on the location of each node inside a chassis, and thus when communicating with nodes outside a Blade Centre H21 frame, a significant penalty is paid. All external interfaces where connected and bonded, thus allowing a theoretical uplink speed of 2x4 Gbps among all Blade Centre frames. However, the theoretical 8 Gbps throughput was not reached because each blade would have to explicitly split its own traffic through both interfaces. This meant that all applications running on these clusters, would have to be designed to take into account the physical layout of the network, a fact which is of course not feasible. However, a number of MPI implementations, including the OpenMPI used in our clusters, are able to make use of multiple interface cards. All these configuration changes helped us to attack each of the network bottlenecks we encountered. Trying several solutions, we managed to scale and use the available network bandwidth until it was no longer the prevailing bottleneck. The current bottleneck is again the memory capacity and the intrinsic latency of the Gigabit Connections.

During extensive network tests, it was shown that the use of 9000 bit Jumbo Frames instead of the standard 1500 MTU was also able to further improve link throughput. Thus, using Iperf[11] a further 50 Mbps bandwidth increase per blade node was achieved. Tests also showed that both the operating system and the interconnecting switches must use Jumbo frames in order to obtain this gain.

The Sun Grid Engine[12] software provides policy-based workload management and dynamic provisioning of application workloads, and we use it to automatically distribute out MPI tasks across our cluster.

2. The HPL benchmark

The HPL benchmark[8] is a software package that solves a random dense linear system in double precision (64 bit) arithmetic on distributed-memory computers. The algorithm used by HPL can be summarized by a two-dimensional block-cyclic data distribution, a right-looking variant of the LU factorization with row partial pivoting featuring multiple look-ahead depths, a recursive panel factorization with pivot search and column broadcast combined, various virtual panel broadcast topologies, bandwidth reducing swap-broadcast algorithm, and a backward substitution with look-ahead of depth one.

The HPL package provides a testing and timing program to quantify the accuracy of the obtained solution as well as the time it took to compute it. The best performance achievable by this software on a certain system depends on a large variety of factors. Nonetheless, with some restrictive assumptions on the interconnection network, the algorithm and its attached implementation are scalable in the sense that their parallel efficiency is mostly maintained constant with respect to the per processor memory usage. The HPL software package requires the availability of an implementation of the Message Passing Interface MPI (at least Version 1.1 compliant). An implementation of either the Basic Linear Algebra Subprograms BLAS is also needed. It must be noted that machine-specific as well as generic implementations of MPI and BLAS are available for a large variety of systems.

On the HPC-ICF cluster, the HPL benchmark has been performed on 63 of the 64 Blade Servers. The theoretical peak performance of the cluster thus stands at around 4.2 TFlops. After a session of tuning of the HPL benchmark through the parameters offered to the user, we obtained a performance of 1.42 TFlops for a matrix size of 280000, using 504 MPI processes with a block-size of 250, and a mesh distribution of 8x63. The benchmark took 10304 seconds to complete on the ICF cluster. On the other hand, the tests on the NCIT-UPB cluster have been performed on 28 of the 32 Blade Servers available in the cluster, with a theoretical peak performance of around 2 TFlops. The initial out-of-the-box performance was of 720 GFlops and subsequent tuning of the simulation and networking infrastructure and the operating system kernel parameters brought the obtained performance to 1 TFlops for a matrix size of 230000, using 224 MPI processes, a block-size of 250 and a mesh distribution of 8x28. On the NCIT-UPB cluster the benchmark took 8086 seconds to complete.

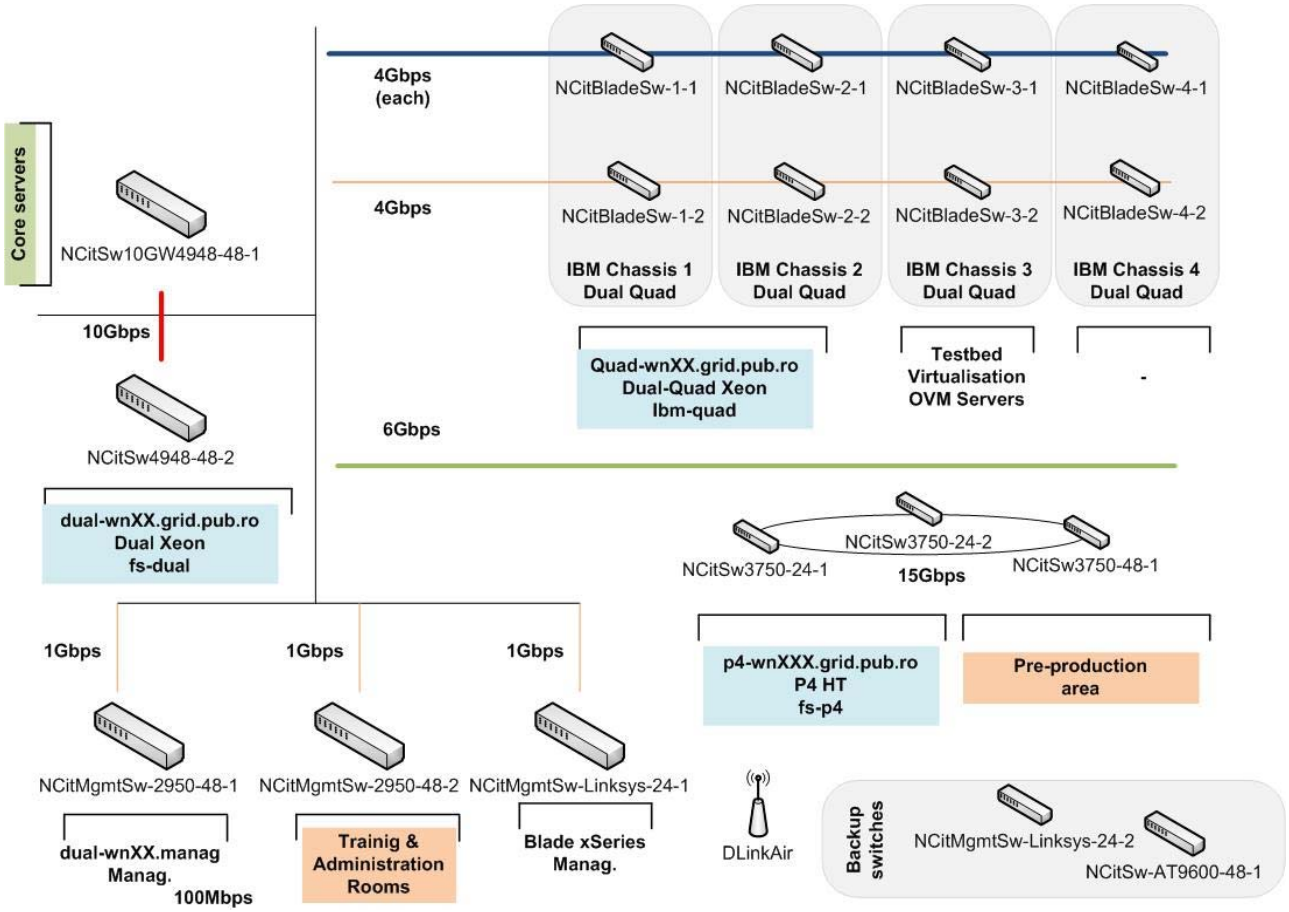

Fig. 1 – Cluster topology and worker node distribution of the NCIT-UPB cluster.

There are a number of parameters for the HPL benchmark that can significantly influence the performance obtained. We will briefly explain just a few of these, in order to show how we tuned the benchmark to the hardware changes operated in our cluster, to reach 1003 Gflops from the 720 GFlops of the base run. During the benchmarking runs a number of measurements were carried out in order to monitor its progress. Figure 2 depicts the traffic observed on the cluster switches during these tests.

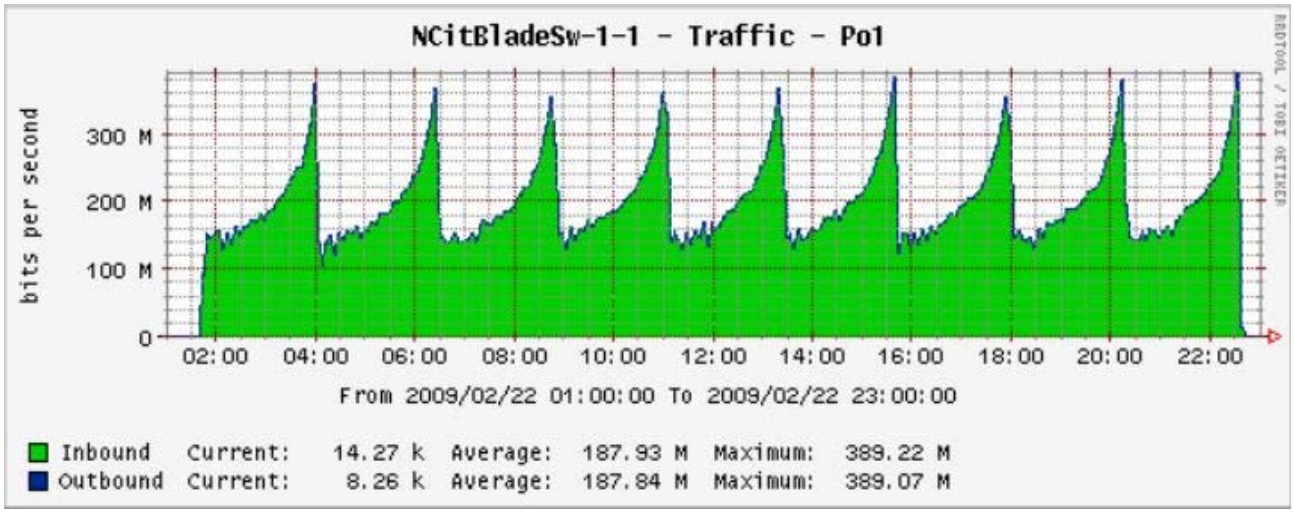

Fig. 2 – Inbound and Outbound traffic on IBM HS21 clusters during HPL tests.

3. The quantum chemistry electronic structure package GAMESS

The package GAMESS (General Atomic and Molecular Electronic Structure System),[9] is a many-electron wave-function software that includes a large variety of Self-Consisting-Field or post-SCF electronic structure techniques based on Hartree Fock, Density Functional and Electron Correlation Theories for ground or excited states. It performs geometry and transition structure optimizations, reaction path searchers, vibrational frequencies calculations, Molecular Dynamics and Monte Carlo simulations. It may be coupled with some molecular mechanics codes to do combined quantum molecular–molecular mechanics calculations for very large atomic systems. The code is very popular within the quantum chemistry community, being used for the characterization of various electronic properties of small molecules, large atomic/molecular clusters or macromolecules at different levels of calculation accuracy.

Table 1

GAMESS Speedup and Efficiency for standard tests

Test No.	Serial Runtime (s)	Speedup			Efficiency		
		2 procs	4 procs	8 procs	2 procs	4 procs	8 procs
12	9.3	1.860	3.207	5.167	0.930	0.802	0.646
22	1.4	1.750	2.800	4.667	0.875	0.700	0.583
30	4.2	1.910	3.500	6.000	0.955	0.875	0.750
31	4.1	1.780	2.733	3.417	0.891	0.683	0.427
33	6.8	1.744	2.833	3.778	0.872	0.708	0.472
34	0.7	1.750	2.333	3.500	0.875	0.583	0.438
35	3.5	1.458	1.944	2.188	0.729	0.486	0.273
37	0.4	1.333	0.800	0.667	0.667	0.200	0.083
38	20.2	1.519	2.061	2.525	0.759	0.515	0.316
40	3.5	1.167	1.029	0.921	0.583	0.257	0.115
41	11.6	1.933	3.625	6.444	0.967	0.906	0.806

In order to verify the correctness and to establish the scalability of the GAMESS installation on our clusters, we conducted a series of standard tests from the package itself. In Table 1 we depict the obtained speedup and efficiency of the most significant tests in terms of performance. Depending on the specific nature of the problem, one can notice speed-ups of up to 6.4 for 8 processors, which is a very good scalability for a package of such complexity. The runtime of the considered systems also varies, from a couple of seconds to about 20s. The performance of the parallel calculations varies accordingly with the size of the atomic system, for small systems scalability being poor.

From preliminary experiments with more complex systems, the GAMESS package scale rather well, when remaining within the confines of a single Blade system. The challenge for us now is to scale the package to more than one computing node. This task is as much dependent on the computing system itself, as it is on the mapping of the GAMESS package to the exact hardware architecture on which it runs.

4. The solid-state linear DFT software ONETEP

The code ONETEP (Order-N Electronic Total Energy Package)[10] is a parallel code designed for accurate electronic structure calculations for solids and molecules within the frame of Density Functional Theory. It performs geometry and

transition state calculations as well as Molecular Dynamics simulations for very large systems (more than 500 atoms), with a computational effort linear versus the size of the system.

It uses a basis of non-orthogonal generalized Wannier functions (NGWF)[13] developed in series of periodic cardinal sine functions, which can be treated as in the usual basis of plane-waves techniques. ONETEP is included as a module in the Materials Studio package[14] but it may work as a stand-alone program, also. Here we used the variant of the code included in the package Materials Studio 5.0.

Table 2
ONETEP Speedup and Efficiency for the MgO system

Number of Processors	Runtime (s)	Speedup	Efficiency
1	11190	-	-
2	6246	1.792	0.896
3	4541	2.464	0.821
4	3590	3.117	0.779
5	3674	3.046	0.609
6	3334	3.356	0.559
7	3332	3.358	0.480
8	3115	3.592	0.449

The scalability of the package ONETEP was tested on the cluster HPC-ICF for a supercell 3x3x3 of the magnesium oxide, which consists in 216 atoms, using the exchange-correlation functional PW91 and a number of 864 NGWFs.

As can be seen in Table 2, the ONETEP package scales to a maximum of 3.5 for 8 cores, with the best efficiencies being obtained for up to four processor cores, for the given simulation. One also observes a drop in performance from four to five processors, and also a marginal improvement from six to seven processors. These issues are most likely due to the load imbalance incurred by the poor mapping of the problem to the available processing processes. That is why a significant improvement can still be observed when using all the cores on a blade, namely eight. However, further tuning is required until this code will be able to scale properly.

CONCLUSIONS

In this work we present the potential of the new computer clusters HPC-ICF and NCIT-UPB, which are now available to Roumanian scientific community. The HPC benchmark performed on the two clusters shows that with appropriate tuning of the parallel environment the performances of both clusters break the barrier of 1 Teraflops. Furthermore, we studied the scalability the two electronic structure codes GAMESS and ONETEP. We showed what steps must be considered to reach the optimal performance of the computing infrastructure. The availability of state-of-the-art hex-core or even oct-core blade systems, may help in the near future with the increased number of cores which seem to be necessary when one desires a balanced run.

In the near future, we will use these machines to design, develop, test and use a multitude of distributed applications[1] in the domain of physical chemistry, as well as other various scientific and engineering fields such as earth-sciences, meteorology, aero and space sciences, together with partners from other research institutes in Romania and abroad.

Acknoledgements: The authors gratefully acknowledge the financial support provided by the National Authority for Scientific Research, Bucharest, Roumania within the Grant Capacities Project 84 CpI/13.09.2007. The authors thank the anonymous referees of the project proposal, who understood the need and the importance of creating such a computational infrastructure for materials science.

REFERENCES

1. "Sourcebook of Parallel Computing", Eds. J. Dongarra et al., Morgan Kaufmann Publishers Inc., San Francisco, 2003.

2. A. Grama, A. Gupta, G. Karypis and V. Kumar: "An Introduction to Parallel Computing", Addison Wesley, 2003.
3. Capacities project **130/14.09.2007**, financed by National Authority for Scientific Research, Bucharest, Roumania.
4. See more technical informations on the web page http://chem.ubbcluj.ro/pagini/anorganica/isi/capacitati.
5. Capacities Project **84 CpI/13.09.2007** - National Authority for Scientific Research, Bucharest, Roumania.
6. High Performance Computer Cluster at Institute of Physical Chemistry, HPC-ICF Cluster Website: http://www.hpc-icf.ro.
7. The Roumanian National Center for Information Technology, University Politehnica of Bucharest, NCIT-UPB Cluster Website: http://cluster.grid.pub.ro/index.php/ncit-cluster.
8. High Performance Linpack Benchmark, Website: http://www.netlib.org/benchmark/hpl.
9. M.W.Schmidt, K.K.Baldridge, J.A.Boatz, S.T.Elbert, M.S.Gordon, J.H.Jensen, S.Koseki, N.Matsunaga, K.A.Nguyen, S.Su, T.L.Windus, M.Dupuis and J.A.Montgomery, J. Comput. Chem., **1993**, 14, 1347; Website: http://www.msg.chem.iastate.edu/gamess.
10. C.-K. Skylaris, P. D. Haynes, A.A. Mostofi and M. C. Payne, J. Chem. Phys., **2005**, 122, 084119; Website: http://www2.tcm.phy.cam.ac.uk/onetep.
11. Iperf - Measuring End-to-End Bandwidth with Iperf Using Web100, Stanford University, **2008**; Website: http://sourceforge.net/projects/iperf/.
12. Sun Grid Engine Website: http://www.sun.com/software/sge.
13. C.-K. Skylaris, A.A. Mostofi, P.D. Haynes, O. Diéguez, and M.C. Payne, Phys. Rev. B, **2002**, 66, 035119.
14. Accelrys Software Inc, Website: http://accelrys.com/products/materials-studio.