Virt&I-Comm.3.2012.26

PACKAGES AND TOOLS FOR THE VIRTUAL CHEMISTRY, MOLECULAR AND MATERIALS SCIENCE AND TECHNOLOGIES RESEARCH COMMUNITY

Alessandro Costantini[1], Daniele Cesini[2], Antonio Laganà[3]

[1] INFN Perugia, Via A. Pascoli 06123, Perugia, Italy
[2] INFN CNAF, Viale Berti Pichat 40127, Bologna, Italy
[3] Department of Chemistry, University of Perugia, Perugia, Italy

ABSTRACT

The present paper outlines the planning lines of the EGI CMMST Virtual Team (VT) project aimed at investigating services and organisational possibilites for the creation of a sustainable Virtual Research Community of the Chemistry, Molecular & Materials Science and Technology computational community.

The paper examines the tools offered to the community to the end of building a solid and highly collaborative distributed environment.

1 INTRODUCTION

Tanks to the progress made during the EGEE [1] and the EGI-inspire [2] European projects, the Chemistry, Molecular and Materials Science and Technologies (CMMST) community has set up a proposal aimed at establishing a CMMST Virtual Team (VT) [3] of the European Grid Infrastructure (EGI) [4] devoted to grounding the assemblage of a homonimous Virtual Research Community. The VT investigates how to exploit the capabilities of the existing EGI tools in building distributed workflows and "workflows of workflows" from various software packages especially for GEMS (Grid Empowered Molecular Simulator) [5, 6]. The mentioned packages deal with electronic structure computations, quantum and classical molecular dynamics computation, statistical averaging over unobserved variables, distributed database and knowledge repositories relying on the expertise of the VRC members.

2 TECHNOLOGIES AND SERVICES OFFERED BY THE CMMST-VRC THROUGH EGI

In this section a prospect of technical and non technical aspects related to the applications and services that can be provided by the VRC to the CMMST users are given. Applications and services are the common patrimony of the VRC and are also the ground on which the VRC supports its members and grows in quality and becomes able to develop higher complexity activities.

In particular, attention has been focused to the software packages ported into the Grid environment and currently in use by the CMMST members and to the capabilities of the existing EGI and Grid-related tools.

ISSN: 2279-8773

2.1 Provided Computational Chemistry applications

Leveraging on the Molecular and Materials sciences research competences of the CMMST community members, the technology of distributed computing has been exploited to the end of tackling multi-scale problems starting from a rigorous treatment of atomic and molecular structures and processes with the aim of enabling real-like simulations of a wide range of problems. These real-like problems concern research, innovation and development in various fields ranging from chemical engineering to biochemistry, chemometrics, omic-sciences, forensic chemistry, medicinal chemistry, food chemistry, energy production and storage, new materials, space technologies, etc. To this end, in a joint endeavour with other international Research Laboratories, various programs devoted to

- the ab initio calculation of the electronic structure of molecular systems and the assemblage of ab initio based potential energy surfaces,
- the integration of quantum and/or classical equations of motion,
  the additional statistical and higher scale treatments necessary to work out the value of measurable quantities

have been developed and/or gathered together to form a Grid Empowered Molecular Simulator (GEMS) [5,6] devoted to the assemblage of realistic innovation after being ported on the computing Grid. A set of in-house developed and third party programs (a full list of chemistry and solid state physics software is given in Ref. [7]) has been implemented on the computing Grid to be used as a cooperative computational engine for real world high-level of complexity applications (aimed at supporting research, innovation and development in various fields ranging from pure science to innovative technologies) of strong economic and social impact through a service oriented approach. Such cooperative endeavour is based on

1) the combined expertise and effort of experts in the field of the molecular science
2) the adoption (with appropriate adaptations) of high level ICT instruments.

A list of ab initio packages (Hartree-Fock (HF), post Hartree-Fock, density functional theory (DFT), Molecular Mechanics (MM) or semi-empirical quantum chemistry type, both open source and commercial software), as those given in Ref. [7], are suitable for insertion into GEMS. Most of them are large packages, often containing several separate programs, and have been developed over many years. The specificity of CMMST is that of adding to the just mentioned popular packages several other ones designed or available for the other 2 blocks of GEMS.

The ported packages are briefly described in the followings and information on literature reference and reference VRC members (RM) are given:

ABC [8] RM (D. Skouteris, L. Pacifici, E. Garcia, S. Rampino)
ABC is a time independent atom-diatom quantum reactive scattering program using a coupled-channel hyperspherical coordinate method to solve the Schrodinger equation for the motion of the three nuclei (A, B, and C) on a single Born-Oppenheimer potential energy surface.

FLUSS-MCTDH [9,10]
FLUSS-MCTDH is a pair o programs carrying out a multiconfigurational time-dependent Hartree (MCTDH) calculation of thermally averaged quantum dynamics properties of multidimensional systems based on a modified Lanczos iterative diagonalization of the thermal flux operator.

NB-MCTDH [9,10]
NB-MCTDH is a multiconfigurational time-dependent Hartree (MCTDH) program for calculating bound states of a generalized N-Body system (including Non Born Oppenheimer cases).

RWAVEPR [11]
RWAVEPR is a time dependent atom diatom quantum reactive scattering program using Jacobi coordinates to integrate rigorously the three-dimensional time-dependent Schroedinger equation by propagating wave packets.

GAMESS-US [12]

ISSN: 2279-8773

GAMESS-US is an ab initio electronic structure molecular quantum chemistry package that calculates potential energy values for moderately large molecular systems using direct and in parallel techniques on appropriate hardware.

SC-IVR [13]
SC-IVR is a Semi-classical (SC) initial value representation (IVR) program based on the outcome of a classical trajectory code used to calculate the thermal rate coefficients for the gas-phase reactions.

GAUSSIAN [14]
GAUSSIAN is an ab initio electronic structure molecular package to calculate potential energy values for moderately large molecular ems performed using direct techniques, or in parallel on appropriate hardware.

DIFFREALWAVE [15]
DIFFREALWAVE: A parallel real wavepacket code for the quantum mechanical calculation of reactive state-to-state differential cross sections in atom plus diatom collisions.

Also in the classical dynamics, contrary to what occurs for the ab initio section, use is made of some general purpose packages for the integration of quantum and/or classical equations of motion. In particular, the most used packages by the CMMST community are:

VENUS96 [16]
VENUS96 is a program developed and maintained by W.L.Hase (QCPE-671). It calculates the trajectory for two reactants (atoms or molecules) by integrating the Hamilton equation in cartesian coordinates. VENUS96 is linked to the Semi-Classical (SC) Initial Value Representation (IVR) program SC-IVR [13] in order to perform semiclassical calculations of the thermal rate coefficient for gasphase reactions.

DL_POLY [17]
DL_POLY is the most used code in the CMMST VRC for the integration of the classical equation of motion of Molecular dynamics calculations. It is a general purpose package of subroutines, programs and data designed to facilitate MD simulations. DL POLY is continually developed at Daresbury Laboratory by W. Smith and I.T. Todorov under the auspices of EPSRC and NERC in support of CCP5. It can be used to simulate a wide variety of molecular systems including simple liquids, ionic liquids and solids, small polar and non-polar molecular systems, bio- and synthetic polymers, ionic polymers and glasses solutions, simple metals and alloys.

GROMACS[18]
GROMACS is a versatile package to perform molecular dynamics by integrating the Newtonian equations of motion for systems with hundreds to millions of particles. It has been used in a large number of case studies and a complete workflow aimed at exploit the interoperability between a local cluster platform (HPC capable) and a Grid platform (mainly HTC capable) has been developed. In the workflow, the coupling among the various jobs is taken care using a link (a semaphore) that de_nes the dependency job chain.

NAMD [19]
NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems and it.. Ported on the Grid environment using OpenMPI parallel libraries, a Direct Acyclic Graph (DAG) has been implemented to run the code in a semi-atuomatic way and facilitate the used on carrying out his/her calculations. The ported NAMD-based system has been used to study the behavior of a lipidic bi-layer in a water box.

Other applications planned to be ported in the Grid environment but not yet integrated in the AppDB are:

CRYSTAL [20]
CRYSTAL is a commercial quantum chemistry ab initio program written by V.R. Saunders, R. Dovesi, C. Roetti, R. Orlando, C.M. Zicovich-Wilson, N.M. Harrison, K. Doll, B. Civalleri, I.J. Bush, Ph. DArco, and M. Lunell from the Theoretical Chemistry of the University of Torino (IT) and the

ISSN: 2279-8773

Computational Materials Science Group at the Daresbury Laboratory near Warrington in Cheshire, England and recognized within the computational chemistry community as a powerful tool for carrying out molecular and material simulations for scientific and technological applications.

APH3D [21]
APH3D is a time independent atom diatom quantum reactive scattering program using a coupled-channel hyperspherical coordinate method to solve the Schrodinger equation for the motion of the three nuclei (A, B, and C) on a single Born-Oppenheimer potential energy surface.

GFIT4C [22]
GFIT4C is a routine devoted to the fit of the Potential Energy surface of four body molecular systems.

POMULT [23]
POMULT is a fortran program supporting the spectroscopy studies of large molecules by locating periodic orbits and equilibrium points in Hamiltonian systems based on 2-point boundary value solvers which use multiple shooting algorithms.

FORWCONV [24]
FORWCONV is a routine devoted to the forward convolution of the Newton Diagrams of a distribution of Newton diagrams to simulate the composition of a gas phase molecular beam.

QUANTUM ESPRESSO [25]
QUANTM ESPRESSO is an integrated suite of Open-Source computer codes for electronic-structure calculations and materials modeling at the nanoscale based on density-functional theory, plane waves, and pseudopotentials.

2.2 Provided Grid tools and services

Taking advantage of the activities of CMMST members, several CMMST applications have been ported in the Grid environment and dedicated user support activities are provided.

The use of different applications and technologies often requires advanced skills not always available to the end-user. For such reason a set of tools have been developed to govern complex ensembles of data, models and programs of an increasing number of applications and to offer a unified user friendly way of composing related tools.

Technologies listed in this document are those provided by EGI and connected initiatives and projects aimed at supporting the various communities operating on the European Grid for their work. In particular, the following technologies have been adopted to be offered to the community members:

AppDB [26]
The EGI Applications Database (AppDB, (http://*appdb.egi.eu/*)) is a central service that stores and provides to the public, information about: tailor-made software tools for scientists and developers to use, the programmers and the scientists who developed them, and the publications derived from the registered software items. All software filled in the AppDB is ready to be used on the European Grid Infrastructure. AppDB is now also able to act as a repository of the registered software.

gUSE WS-PGRADE [27]
Is an open source DCI gateway framework that enables users the convenient and easy access to Grid and Cloud infrastructures and supports the development, visualization, configuration and submission of distributed applications executed on the computational resources. gUSE WS-PGRADE has the structure of a workflow enabling application developers to define their investigations by means of a graphical environment and to generate out of the user-specified description the Grid scripts and commands allowing the execution of the various computational tasks on the distributed computing platforms. For the above mentioned reasons gUSE WS-PGRADE has been used as a workflow engine in many works by developing a set of reusable application-specific workflows.

GC3PIE [28]
Is a library of Python classes for running large job campaigns (high throughput) on diverse batch-oriented execution environments (such ARC) providing facilities for implementing command-line

ISSN: 2279-8773

driver scripts in the form of Python object classes whose behaviour can be customized by overriding specified object methods. The adoption of the framework GC3Pie in many computational campaigns allowed the users to define event-related dependencies between different applications and execute them simultaneously on a large-scale distributed computing infrastructure. Thanks to its the programmatic approach to workflows, the entire execution schema is assembled at runtime and steps can be added and removed dynamically as the program progresses, adapting to the outcome of individual computations.

GriF [29]
Is a Service Oriented Architecture (SOA) Collaborative Framework designed to facilitate the use of the DCIs by non specialists aimed at optimizing the selection of different computing elements for running single and parameter study applications. It consists on a set of java modules aimed at submit and monitor jobs in different computing platforms such as Grid or HPC. The framework has been equipped with a Quality of User (QoU) - Quality of Service (QoS) evaluation module called GCRES based on the monitoring of the user activities. Both GriF and GCRES have been developed within the activities of the CMMST community.

IGI Portal [30]
Is a science gateway developed within the activities of the Italian Grid Initiative (IGI), which is operating the Italian Grid infrastructure. Already used by other communities, the web-based portal provides to the final user several important services such as job submission, workflow definition and data management for both Grid and Cloud environments. Moreover, it implements a robust Authentication and Authorization (AA) mechanism to provide a secure computational environment and, at the same time, hiding the complexity of the X.509 digital certificates on which most of the DCI resources rely.

InSilico Lab [31]
Is an application portal designed to support insilico experiments by easily running computational chemistry software on grids. Unlike manual job submission or grid portals, InSilicoLab enables to run computations on grids without technical knowledge of how to operate it.

MyGAMESS [32]
Is a Framework for Integration of Chemistry applications with the Swiss Grid Portal aimed at providing dynamic access to applications services. Users will profit for example by a considerable reduction of time-consuming data format conversions and/or other code restructuring, once the capabilities of different applications are integrated.

G-LOREP [33]
Grid LOs Repository is a manager of a federation of distributed repositories of research based learning objects. It relies on a central database and allows registered users to create and/or download learning contents for research based education activities. G-LOREP makes also the content automatically shareable among the federation servers and enables the creation of dependency management software attachment as well as fault tolerant submission of the simulations from any federated site.

There are other technologies suited to be used by the CMMST community but their usability is currently under the study of selected computational chemistry experts (which belong to the community) supported by the respective technology developers.
Such technologies are here briefly described:

KEPLER [34]
Is an open source, scientific workflow application. Using Kepler graphical interfaces and components, scientists can create and execute scientific workflows which allow accessing scientific data and executing complex analyses on them. Kepler has been recently adopted in a recent work [35] where a prototypical workflow has been designed and tested for the calculation of the time independent quantum probabilities of the H+H2 benchmark reaction. For this purpose a set of quantum mechanical codes, both belonging to the set of the aforementioned computational applications, have been used.

SHIVA SIMULATION PLATFORM [36]

ISSN: 2279-8773

Is a project providing a multi-systems workflow execution platform and an interoperability solution. The platform currently supports seven workflow systems: Askalon, Kepler, LONI pipeline, MOTEUR, Taverna, Triana and WS-PGRADE.

3 DATA FORMATS AND HPC-HTC PLATFORMS INTEGRATION

Even if the synergistic model adopted by the CMMST community in principle overcomes the limitations of both the grant and the purely opportunistic computing models, much work has still to be done in order to effectively bridge HPC and HTC machines in order to overcome those highly unsatisfactory situations in which neither HPC nor HTC are completely fit to meet the requests of complex CMMST applications. Moreover, also on the resource providers side (and not only on the user one) there are good reasons for coordinating the use of HPC and HTC e-infrastructures to the end of interoperating large computational applications. This in fact allows an optimization of the usage of both HTC and HPC computing resources because it is not infrequent the case in which a user utilizes HPC platforms not as such but as a bunch of loosely coupled processors underutilizing their fast dedicated network. At the same time HTC users may utilize massively distributed HTC platforms to solve tightly coupled computational tasks ending up by wasting a large amount of time in transferring data on the net. A coordination of the two types of platforms to interoperate via a single workflow (or workflow of workflows) and properly manage the various components on the most appropriate hardware, would instead allow a clever composition of complex applications optimizing the use of the various computing resources and providing the users with the best level of performance.

3.1 The CMMST activity

An advantage of such distributed model consists in the possibility for the user to choose the platform better fitted for his/her applications and for the computer centres to qualify the usage of theirs resources. This provides the CMMST members with the possibility of combining different pieces of software to the end of assembling complex realistic simulators (like GEMS) and undertake more ambitious research projects. Such possibility of building workflows of shared programs has, for example, stimulated the setting of proper (de facto) standards of data in quantum chemistry and quantum dynamics [37]. This has prompted the development of tools (like the framework GriF) enabling the redirecting of computer applications to run on the best suited sites (including HPC machines) by properly redirecting the jobs to the most appropriate architecture. This enhances cooperative compute capabilities by opening the perspective of combining different complementary know how into single (higher level of complexity) realistic applications and paves the way to apply for more ambitious research grants to be shared within the community.

Moreover the CMMST community wishes to collaborate in interoperability EGI activities as, for example, the EUDAT and PRACE pilot ones [38] aimed at bringing together different infrastructures and user communities by focusing their interest in the data sharing facilities.

Such activity is expected to provide a set of tools and recommendations that the CMMST community can use to integrate its CMMST workflows.

4 CONCLUSIONS

In the present document the advantages that a VRC status would offer to the members of the CMMST community in terms of an advanced shared usage of hardware and software that (thanks to the expertise offered the various groups) can be integrated to address complex problems and computing simulations of high social impact has been discussed. The list of packages already available to the CMMST community has been discussed, some of the most powerful tools made available by EGI

ISSN: 2279-8773

have been presented, the exploitation of the GriF framework to select the computing resources most suited for the planned calculations are illustrated and the names of some of its members that are candidates to maintain them on the Grid and support the users are given.

REFERENCES

[1] EGEE website: http://www.egee.eu

[2] EGI website-projects: http://www.egi.eu/projects/egi-inspire/

[3] Virtual Team website: https://wiki.egi.eu/wiki/Virtual_team

[4] EGI website: http://www.egi.eu

[5] A. Costantini, O. Gervasi, C. Manuali, N. Faginas Lago, S. Rampino, A. Laganà: COMPCHEM: progress towards GEMS a Grid Empowered Molecular Simulator and beyond, Journal of Grid Computing, 8(4), 571-586 (2010)

[6] S. Rampino, A. Monari, S. Evangelisti, E. Rossi, K. Ruud, A. Laganà: A priori modeling of chemical reactions on a grid-based virtual laboratory. Cracow 09 Grid Workshop, 164-171 (2010) ISBN978-83-61433-01-09

[7] List of quantum chemistry and solid state software:
http://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid_state_physics_software

[8] Skouteris, D., Castillo, J.F., Manolopulos D.E.: ABC: a quantum reactive scattering program. Comp. Phys. Comm. 133 (2000) 128–135

[9] U. Manthe: Direct calculations of reaction rates in Reaction and Molecular Dynamics. A. Laganà and A. Riganelli Eds, Springer Verlag Berlin (ISBN 3-540-41202-6) (2000) p. 130

[10] Beck, M., Jakle, A., Worth, G., Meyer, H. D.: The multiconfiguration timedependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets. Phys. Rep. 324 (2000) 1–05

[11] D. Skouteris, L. Pacifici, A. Laganà, Time dependent wavepacket calculations for the $N(^4S) + N_2(^1\Sigma^+_g)$ system on a LEPS surface: inelastic and reactive probabilities, Mol. Phys. (ISSN:0026-8976) 102(21-22), 2237-2248 (2004).

[12] GAMESS-US see http://www.msg.ameslab.gov/gamess/ 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, J.A.Montgomery J. Comput. Chem., 14, 1347 (1993).

[13] M. Ceotto, S. Atahan, S. Shin, First principles semiclassical initial value representation molecular dynamics, Phys. Chem. Chem. Phys. 11, 3861-3867 (2009)

[14] GAUSSIAN website: http://www.gaussian.com/

[15] Hankel, M., Smith, Sean C., Gray, S. K. and Balint-Kurti, G. G. (2008) DIFFREALWAVE: A parallel real wavepacket code for the quantum mechanical calculation of reactive state-to-state differential cross sections in atom plus diatom collisions. Computer Physics Communications, 179 8: 569-578.

[16] VENUS96 Chemical Dynamics Software and Simulation System website: https://cdssim.chem.ttu.edu/nav/htmlpages/licensemenu.jsp. Last seen May 2013

[17] Smith W, Forester T. R. (1996): DL POLY2: a general purpose parallel molecular dynamics simulation package. Journal of Molecular Graphics, Vol. 14 (3), 136-141

[18] Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E. (2008) `GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation', J. Chem. Theor. Comp. Vol 4, pp.435{447

[19] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 26:1781-

ISSN: 2279-8773

1802, 2005.

[20] R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders, and C. M. Zicovich-Wilson, Z. Kristallogr. 220, 571 (2005).

[21] Laganà, A., Crocchianti, S., Ochoa de Aspuru, G., Gargano, R., Parker, G.A.: Parallel time independent quantum calculations of atom diatom reactivity. LNCS 1041 (1995) 361-370

[22] A. Aguado, C. Tablero, and M. Paniagua, Comput. Phys. Comm. 134, 97 (2001).

[23] S. C. Farantos, POMULT: A Program for Computing Periodic Orbits in Hamiltonian Systems Based on Multiple Shooting Algorithms, Comp. Phys. Comm., 1998, 108, 240-258

[24] N. Balucani, P. Casavecchia, L. Banares, FJ Aoiz, T. Gonzalez-Lezana, P. Honvault and J.M. Launay, J. Phys. Chem. A, 2006, 110, 817.

[25] QUANTUM ESPRESSO website: http://www.quantum-espresso.org/

[26] AppDB website: http://appdb.egi.eu/. Last seen May 2013

[27] Sipos, G., Kacsuk, P.: Multi-Grid, Multi-UserWorkflows in the P-GRADE Portal. Journal of Grid Computing, 3 (2005) 221-238

[28] GC3Pie website: http://code.google.com/p/gc3pie/. Last seen May 2013

[29] C. Manuali, A. Laganà GRIF: A New Collaborative Framework for a Web Service Approach to Grid Empowered Calculations Future Generation of Computer Systems, 27(3), 315-318 (2011)

[30] IGI portal website: https://portal.italiangrid.it/. Last seen May 2013

[31] In Silico Lab website: http://insilicolab.grid.cyfronet.pl/. Last seen May 2013

[32] MyGAMESS - TO SET

[33] S. Tasso, S. Pallottelli, M. Ferroni, R. Bastianini, A. Laganà, Taxonomy management in a Federation of Distributed Repositories: a chemistry use case Lecture Notes Computer Science 7333, 358-370 (2012)

[34] Kepler website: https://kepler-project.org/. Last seen May 2013

[35] Marcin Plóciennik, Michal Owsiak, Tomasz Zok, Bartek Palak, Antonio Gómez-Iglesias, Francisco Castejón, Marcos López-Caniego, Isabel Campos Plasencia, Alessandro Costantini, Dimitriy Yadykin, Per Strand: Application Scenarios Using Serpens Suite for Kepler Scientific Workflow System. Procedia Computer Science, Volume 9, 2012, Pages 1604-1613

[36] SHIWA project website: http://www.shiwa-worklfow.eu/project. Last seen May 2013

[37] E. Rossi, S. Evangelisti, A. Laganà, A. Monari, S. Rampino, M. Verdicchio, K. Baldridge, G.L. Bendazzoli, S. Borini, R. Cimiraglia, C. Angeli, P. Kallay, H.P. Lüthi, K. Ruud, J. Sanchez-Marin, A. Scemama, P. Szalay, A. Tajti, Code Interoperability and Standard Data Formats in Quantum Chemistry and Quantum Dynamics: the Q5/D5cost Data Model submitted to the J. Comp. Chem.

[38] EGI_EUDAT_PRACE pilot project:

https://wiki.egi.eu/wiki/EGI_EUDAT_PRACE_collaboration

ISSN: 2279-8773