IMP Science Gateway: from the Portal to the Hub of Virtual Experimental Labs in Materials Science

Yuri Gordienko*, Lev Bekenev, Olexandra Baskova, Olexander Gatsenko, Elena Zasimchuk
G.V.Kurdyumov Institute for Metal Physics, National Academy of Sciences
Kyiv, Ukraine
Email: *gord@imp.kiev.ua

Sergii Stirenko
High Performance Computing Center, National Technical University of Ukraine "KPI'', Kyiv, Ukraine
Email: sergii.stirenko@hpcc.kpi.ua

Abstract—The user communities outside of computer science itself (for example, in materials science, physics, chemistry, biology, nanotechnologies, etc.) would like to access different software components and various distributed computing infrastructures (DCIs) (like grids, clouds, clusters) in a such transparent way, that they do not need to learn their peculiarities. In this context, "science gateway" (SG) ideology means a user-friendly intuitive interface between scientists (or scientific communities) and different software components + various DCIs, where researchers can focus on their scientific goals and less on peculiarities of software/DCI. "IMP Science Gateway Portal" (http://scigate.imp.kiev.ua) for complex workflow management and integration of distributed computing resources (like clusters, service grids, desktop grids, clouds) is presented. It is created on the basis of WS-PGRADE and gUSE technologies, where WS-PGRADE is designed for science workflow operation and gUSE - for smooth integration of available resources for parallel and distributed computing in various heterogeneous distributed computing infrastructures (DCI). The typical scientific workflows with possible scenarios of its preparation and usage are presented. Several typical use cases for these science applications (scientific workflows) are considered for molecular dynamics (MD) simulations of complex behavior of various nanostructures (nanointertion of graphene layers, defect system relaxation in metal nanocrystals, thermal stability of boron nitride nanotubes, etc.). The user experience is analyzed in the context of its practical applications for MD simulations in materials science, physics and nanotechnologies with available heterogeneous DCIs. The most crucial advantages are as follows: powerful and flexible workflow construction tools, which allow users to apply LEGO-style for creation of complex tunable chains with de facto standard components (like R, LAMMPS, AtomEye, ffmpeg) and customized components (like debyer, Pizza, etc.); smooth access to heterogeneous DCI and software; division of user roles (administrators: operators of portals; power users: principal scientists; end users: scientists, students); user friendly and intuitively understandable GUI (with additional modules, ad hoc changes, etc.); extensive manuals and permanently working fora and helpdesks; short learning curve for usual scientists without extensive knowledge in computer science. Some non-critical drawbacks of the approach include: scarce "stdout" and "stderr" information, tacit context help for beginners, some unavoidable upgrade issues, etc. In conclusion, the "science gateway" approach - workflow manager (like WS-PGRADE) + DCI resources manager (like gUSE) - gives opportunity to use the SG portal (like "IMP Science Gateway Portal") in a very promising way, namely, as a hub of various virtual experimental labs (different software components + various requirements to resources) in the context of its practical MD applications in materials science, physics, chemistry, biology, and nanotechnologies.

Keywords—Distributed computing infrastructure (DCI), grid computing, cluster, service grid, desktop grid, science gateway, computational physics, molecular dynamics, materials science, physics, nanotechnologies.

I. INTRODUCTION

Recently research boom in nanotechnologies has started to search and investigate new nanoscale materials, like nanocrystals, graphene, carbon nanotubes (CNT), boron nitride nanotubes (BNNT), and their complexes, because of their unique properties. Molecular dynamics (MD) simulations of nanoscale structures, processes, and properties are very promising in the wide range of physical parameters, because of necessity to investigate wide range of parameters (so-called “parameter sweeping”). MD simulations with such “parameter sweeping” can be carried out in the available heterogeneous distributed computing infrastructure (DCI) with desktop grids, service grids, clusters, supercomputers, cloud resources. The recent advances in computing algorithms and infrastructures, especially in development of DCIs, allow us to solve these tasks efficiently without expensive scaling-up, especially by means of the science gateway (SG) technology on the basis of WS-PGRADE platform for workflow management and gUSE technology for integration of DCI-resources [1]. Here we demonstrate the capabilities of these technologies realized for MD simulations and further data post-processing on the basis of LAMMPS package for MD simulation [2], and other packages like R for statistical analysis [3], Pizza.py Toolkit for visualisation of atomic configurations [5], debyer for simulation of X-ray diffraction (XRD) and neutron diffraction (ND) analysis (https://code.google.com/p/debyer/), etc. Several typical workflows were created for simulation of several physical processes with various demands for the computing resources: tension of metal nanocrystals under different physical conditions, tension of ensemble of metal nanocrystals under the same conditions, manipulations with complex nanostructures like indentation of graphene membranes, thermal stability of BNNT, etc.

II. BACKGROUND AND RELATED WORKS

Scientists from user communities outside of computer science itself (for example, in materials science, physics,
chemistry, biology, nanotechnologies, etc.) would like to access various DCIs (grids, clouds, clusters) and concentrate on their scientific applications themselves in a such transparent way, that they do not need to learn the peculiarities of these DCIs. In general, “science gateway” (SG) implies a user-friendly intuitive interface between scientists (or scientific communities), various software (standard and/or proprietary) and heterogenous DCIs. Usually, SGs are not specialized for a certain scientific area and scientists from many different fields can use them. By means of SGs researchers can focus on their scientific goals and less on assembling the required DCI components. The most important aspects of SG are: a simplified intuitive graphical user interface (GUI) that is highly tailored to the needs of the given scientific community; a smooth access to national and international computing and storage resources; collaborative tools for sharing scientific data on national and international scale. Several SG technologies and frameworks are known (like ASKALON [6], MOTEUR [7], WS-PGRADE/gUSE/DCI-bridge [1]), which use different enabling components and technologies: web application containers (Tomcat, Glassfish, etc.), portal or web application frameworks (Liferay, Spring, Drupal, etc.), database management systems (MySQL, etc.), workflow management systems [8]. Here usage of WS-PGRADE/gUSE/DCI-bridge is described with some examples in physics, materials science, nanotechnologies. The architecture of science gateway ideology on the basis of gUSE+WS-PGRADE+DCI-Bridge technologies (https://guse.sztaki.hu) is described in details elsewhere [1]. In short, WS-PGRADE portal technology is a web based front end of the gUSE infrastructure. gUSE (grid and cloud User Support Environment) is a permanently improving open source SG framework developed by Laboratory of Parallel and Distributed Systems (MTA-SZTAKI, Budapest, Hungary), that provides the convenient and easy access to DCI. It gives the generic purpose, workflow-oriented GUI to create and manage workflows on various DCIs. WS-PGRADE supports development and submission of distributed applications executed on the computational resources of DCIs. The resources of DCIs (including clusters, service grids, desktop grids, clouds) are connected to the gUSE by a single point back end, the DCI-Bridge [1].

III. IMP SCIENCE GATEWAY INTERNALS

“IMP Science Gateway Portal” [10] (Fig. 1) is based on the WS-PGRADE/gUSE/DCI-bridge framework [1], which is the complex system with many features. It distinguishes different user communities: administrators, power-users and end-users. The administrators operate portal itself. The power-users can develop new workflow applications by means of various workflow features (graph, abstract workflow, template, application). The workflows and their components are stored in built-in repository and published by the application developers, which can be shared among end-users. It provides scientific workflow management for both application developers and end-users, and supports various DCIs including clusters, service grids, desktop grids and cloud.

A. Workflow Management

“IMP Science Gateway Portal” leverages easy and simple GUI utilities for power-users (application developers) to create (easily in LEGO-style manner) the complex workflows (Fig. 2a) from separate simple components, which can be standard or proprietary executables, scripts in various languages, and command-line arguments (Fig. 2b). Application developers can design and create complex workflows with taking into account various runtimes and resources needed for each component (Fig. 2c), and output data type and size (Fig. 2d). End-users can submit, monitor, save, upload, and retrieve various workflows with one or many jobs, and obtain intermediate or final results (Fig. 2f).

The whole process of the complex workflow creation and management can be demonstrated by the MD simulation of tension of metal nanocrystals (Fig. 2a) with the following steps: (1) to register/sign in “IMP Science Gateway Portal”; (2) to create a graph scheme for the basic components of the workflow (yellow bricks with small input-output ports in Fig. 2b); (3) to create a workflow from the graph; (4) to configure the workflow (links between yellow bricks in Fig. 2a); (5) to run the workflow; (6) to monitor the current status of execution. The complex workflow (Fig. 2e) for this application contains several independent components (LAMMPS package, Pizza.py Toolkit, AtomEye package, debayer, R-package, FFmpeg package), which can be combined in various sets for the purposes of the end users [11]. The end-users can check the intermediate results: after MD simulation by LAMMPS package (tables in the red rectangles in the center of Fig. 2f); converted by Pizza.py Toolkit script package for post-processing (table in the red rectangle in the bottom of Fig. 2f); visualization of atomic positions obtained by AtomEye package (images in the top part of Fig. 2f); final results of R-package (plots in the green rectangles at the left and right sides of Fig. 2f); final post-processed video of visualizations of atomic positions obtained by FFmpeg package (video in the violet rectangle in the top part of Fig. 2f).

B. Integration of Computing Resources

To run jobs gUSE allows to compose various resources and execute workflows on various DCIs: clusters (PBS, LSF),
cluster grids (ARC, gLite, GT2, GT4, GT5, UNICORE), supercomputers (e.g. via UNICORE), desktop grids (BOINC) and clouds (via CloudBroker Platform and GAE). At the moment “IMP Science Gateway Portal” uses local resources (server side), several clusters (PBS) with different configuration and policies, tests desktop grids (BOINC) and cluster grids (ARC), and plans to use supercomputers and clouds. In this context, the different requirements for the various components were taken into account as to their possible location, runtime, output file size (see Table I).

![Diagram](image1)

**TABLE I. THE COMPONENTS OF THE GENERAL WORKFLOW (FIG. 2)**

| Component    | Location       | Runtime   | Output Format (Size) |
|--------------|----------------|-----------|----------------------|
| LAMMPS       | Cluster, DCI   | 1-10 days | Text (1-10 GB)       |
| R-package    | Server (Cluster, DCI) | 1-10 min | Plot (< 1 MB)        |
| Pizza-script | Server (Cluster, DCI) | 1-10 min | Text (1-10 GB)       |
| AtomEye      | Cluster, DCI   | < 1 hour  | 3D Image (< 1 MB)    |
| FFmpeg       | Cluster, DCI   | 1-10 min  | 3D Video (< 10 MB)   |

C. User Management

The development and execution stages are separated between users with different background and responsibility (power-users and end-users) in the next two main user groups: (1) “end users” — the scientists without extensive knowledge of computer science with limited power and permissions (like common scientists in materials science, physics, chemistry, etc.), usually they need only restricted manipulation rights and obtain the scientific applications in the ready state to configure and submit them with minimal efforts; (2) “power users” (“application developers”) — the actual designers of the scientific processes and related workflows for simulations, usually they need to design and create the application (workflow) for comfortable work of the end-users.

IV. USE CASES OF PRACTICAL APPLICATIONS

Several practical applications in materials science are presented below, where the different workflows were designed by WS-PGRADE workflow manager and used for computations in DCI by gUSE technology.

A. Thermal Properties of Boron Nitride Nanotubes

The scientific aim of this use case (Fig. 2) consists in MD simulations of boron nitride nanotubes (BNNT) with exceptional physical properties, which are a prerequisite for their wide practical applications in the future. The current user community in the field of nanotechnologies and materials science includes 2 user groups (4 end users) from IMP and IPMS (Frantsevich Institute for Problems in Materials Science, Kiev, Ukraine). MD simulation workflow (Fig. 3a) was used for investigation of thermal stability of BNNT with different chirality, size and perfection of structure (Fig. 3b). The actually used workflow (Fig. 3a) was the part of the general workflow (Fig. 2a). It included the following subset workflows: 1) LAMMPS + R-package — for MD simulations and plotting some post-processed values; 2) LAMMPS + R-package + AtomEye — for the same operations + visualization of intermediate states of atoms.

From the physical point of view this work allowed us to derive some conclusions as to dynamics (Fig. 3c) and temperature dependence of the BNNT collapse for various physical parameters (chiralities, sizes and perfection of structure) and the simulation parameters (type of potential, boundary conditions, thermal conditions, etc.). The obtained results extend the
known experimental data on the thermal stability of BNNTs and described elsewhere [17].

**B. Elastic Properties of Graphene**

The scientific aim of this use case (Fig. 4) was MD simulation of nanoindentation of monolayer graphene membrane in an atomic force microscope. Its user community in the field of nanotechnologies includes 2 user groups (5 end users) from IMP and SPM/RS-Centre (Centre of scanning probe microscopy and resonance spectroscopy, Kyiv, Ukraine). MD simulation workflow (Fig. 4a) was used for investigation of nanoindentation processes with different velocities, size of membranes and perfection of structure (Fig. 4b). The actually used workflow (Fig. 4a) was the part of the general workflow (Fig. 2a). It included the following subset workflows: 1) LAMMPS + R-package — for MD simulations and plotting some post-processed values; 2) LAMMPS + R-package + AtomEye + FFmpeg — for the same operations + visualization of intermediate states of atoms, and final video visualization.

From the physical point of view it was shown that graphene membrane deforms and collapses within a wide range of physical parameters (strain, applied force, indentor shape, indentor velocity, etc.) (Fig. 4c). These results confirm and enrich the known experimental data of nanoindentation and described elsewhere [16].

**C. Mechanical Properties of Metal Nanocrystals**

The scientific aim of this use case (Fig. 5) was MD simulation of relaxation behavior of nanocrystals after uniaxial tension (Fig. 5b) [13], [14]. Its user community in the field of nanotechnologies includes 2 user groups (3 end users) from IMP and KNU (Kyiv National University, Kyiv, Ukraine). MD simulation workflow (Fig. 5a) was used for investigation of relaxation behavior of nanocrystals for different crystals (Al, Cu, and Si), physical conditions (tensile rate, size and orientation of the nanocrystals) and methodological parameters (potential type, boundary conditions, and others). The actually used workflow (Fig. 5a) included the biggest parts of the general workflow (Fig. 2a). The simulations included the following subset workflows: 1) LAMMPS + R-package — for MD simulations and plotting some post-processed values,
like stress-strain dependencies (Fig. 5c); 2) LAMMPS + R-package + AtomEye + FFmpeg — for the same operations + visualization of intermediate states of atoms and final video visualization; 2) LAMMPS + R-package + AtomEye + FFmpeg + debyer — for the same operations + processing data for X-ray and neutron diffraction and their plotting, like XRD-plots (Fig. 5d).

From the physical point of view it was shown that the oscillatory relaxation of internal stresses is associated with the periodic rearrangement of the metastable defect substructure (Fig. 5b). These results confirm and illustrate the known experimental data and described elsewhere [15].

V. DISCUSSION

The use cases presented here demonstrate that the general purpose and application specific workflows can be created in LEGO-style way by combining separate components-bricks, which corresponds to some real world operations, physical methods, and experimental devices even (see Table II). Actually, such workflows can be easily designed and used in practice for MD simulations of complex natural processes in materials science, physics, and nanotechnologies. Each simulation workflow can be considered as a virtual counterpart (virtual lab) of the real experimental process, and the SG itself can be used as a hub of such virtual experimental labs in materials science.

| Simulation result (method needed) | Virtual lab, i.e. set of components used | Real lab, device |
|-----------------------------------|-----------------------------------------|-----------------|
| Atomic structure in bulk (TEM)    | LAMMPS [2] + AtomEye [5]                | JEOL R005       |
| Stress-strain plot (AFM)          | LAMMPS [2] + R [3]                     | JEOL AFM        |
| CN, RDF plots (neutron, X-ray scattering) | LAMMPS [2] + R [3] | Synchrotron facility |
| X-ray diffraction spectrum (XRD)  | LAMMPS [2] + debyer + R [3]            | Swiss Light Source [21] |
| Neutron diffraction spectrum (ND) | LAMMPS [2] + debyer + R [3]            | Large Hadron Collider |

Some conclusions were made as to the possible future improvements of the workflow and resource management. The main problem was concerned with different demands for resources for various job parameters. For example, double increase of model size leads to nearly proportional double increase of the required memory and runtime. Sometimes this was in conflict with the available policies, like queuing rules, when the job runtime should be lower than the queue walltime. The typical example for the use case A (Fig. 3) with the physical simulation time of 1 ns is shown in Table III for one of the used PBS-clusters, namely NTU HPCC (http://hpc.real.ru) of NGI-UA. In this connection, the GUI-feature, which was pre-designed in SG for selection of a queue for PBS-clusters, was very useful, because it allowed to construct workflows optimized for runtimes lower that the prescribed queue walltime (or when the walltime is a multiple of the job runtime). This allowed to increase the overall efficiency, but the more progress could be reached by additional measures in SG, for example, by optimal job distribution over various queues (which should be automatic, and not manual, like here), and by polling all available queues among all available clusters with selection of the pool with the empty or more efficient resources.

![Fig. 5](image-url) Use case for MD simulations of mechanical properties of metal nanocrystals: (a) the part of the general workflow (Fig. 5a), which was actually used here; (b) visualization by AtomEye (defect substructure evolution in Al nanocrystal, after 15 ps, atoms in ideal lattice — green color, stacking faults — red color), (c) data post-processing and plot output by R-package (oscillation of the internal stress $P_{xx}$ for different nanocrystal sizes), (d) data post-processing and plot output by R-package (X-ray diffraction analysis).

Other means for increase of SG-efficiency, like dual-layer hardware and software management in cluster systems [18], will be also of interest and great value and should be taken.
into account. Additional important aspect is related with the failed jobs, which should be re-run. The problem is that for the big number of jobs (each for different set of parameters) it was not very easy to filter the faulty jobs and re-run them again. At this stage, it was performed manually by means of job management system, but it would be more efficient, if this will done automatically by SG after obtaining the corresponding report.

VI. CONCLUSIONS

Science Gateway Portal (like “IMP Science Gateway Portal”) for MD simulations of complex behavior of various nanostructures can be relatively easily created on the basis of WS-PGRADE technology for workflow management. These resourceful MD simulations can be effectively carried out in the heterogeneous DCI by means of gUSE technology for parallel and distributed computing in various heterogeneous DCIs. Several typical scientific applications were considered as use cases of their porting and practical using. Several advantages should be emphasized: (1) the modular-based (“LEGO-style”) principle of workflow creation and management; (2) easy access to heterogeneous DCI and software; (3) division of user roles (administrators as operators of portals; power users as principal scientists; end users as scientists, students); (4) short learning curve for usual scientists without extensive knowledge in computer science. Some ways for further improvement were briefly outlined in relation to the better performance and efficiency. Finally, the “science gateway” approach – workflow manager (like WS-PGRADE) + DCI resources manager (like gUSE) at the premises of the portal (like “IMP Science Gateway Portal”) – is very promising in the context of its practical MD applications in materials science, physics, chemistry, biology, and nanotechnologies.

ACKNOWLEDGMENTS

The work presented here was partially funded by EU FP7 SCI-BUS (SCientific gateway Based User Support) project, No. RI-283481, and partially supported in the framework of the research theme “Introduction and Use of Grid Technology in Scientific Research of IMP NASU” under the State Targeted Scientific and Technical Program to Implement Grid Technology in 2009-2013.

REFERENCES

[1] P.Kacsuk, Z.Farkas, M.Kozlovsky, G.Hermann, A.Balasko, K.Karoczki, I.Maron, WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities, Journal of Grid Computing, 10(4): 601-630, 2012.
[2] S.Plimpton, Fast parallel algorithms for short-range molecular dynamics, Journal of Computational Physics, 117(1):1-19, 1995.
[3] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 (http://www.R-project.org) 2008.
[4] S.Plimpton, M.Jones, Pizza.py Toolkit (http://pizza.sandia.gov/) 2005.
[5] J.Li. AtomEye: an efficient atomistic configuration viewer, Modelling and Simulation in Materials Science and Engineering, 11(2): 173-177, 2003.
[6] T.Fahringer, A.Jugravu, S.Pillai, R.Prodan, C.Seragiotti, H.L.Truong, ASKALON: a tool set for cluster and Grid computing, Concurrency and Computation: Practice and Experience, 17(2-4): 143-169, 2005.
[7] T.Glatard, J.Montagnat, D.Lingrand, X.Pennec, Flexible and efficient workflow deployment of data-intensive applications on grids with moteur, International Journal of High Performance Computing Applications, 22(3): 347-360, 2008.
[8] E.Deelman, D.Gannon, M.Shields, I.Taylor, Workflows and e-Science: An overview of workflow system features and capabilities, Future Generation Computer Systems, 25(5): 528-540, 2009.
[9] WS-PGRADE Portal User Manual, 2013 (http://sourceforge.net/projects/guse/).
[10] Yu.Gordienko, L.Bekenov, O.Gatsenko, E.Zasimchuk, V.Tatarenko, Complex Workflow Management and Integration of Distributed Computing Resources by Science Gateway Portal for Molecular Dynamics Simulations in Materials Science. Proc. Third Int. Conf. “High Performance Computing” HPC-UA 2013, (Ukraine, Kyiv, 2013), 148-155.
[11] Yu.Gordienko, O.Baskova, O.Gatsenko, Enabling High-Performance Distributed Computing to e-Science by Integration of 4th Generation Language Environments with Desktop Grid Architecture and Convergence with Global Computing Grid. Proceedings of Cracow Grid Workshop ’10, Cracow, Poland, pp. 234-243, (https://www.researchgate.net/publication/261361640) 2011.
[12] O.Gatsenko, L.Bekenov, E.Pavlov, Yu.Gordienko: From Quality to Quality: Massive Molecular Dynamics Simulation of Nanostructures Under Plastic Deformation in Desktop and Service Grid Distributed Computing Infrastructure, Computer Science, 14: 27-44; doi:10.7494/csci.2013.14.1.27, 2013.
[13] Yu.Gordienko, Molecular Dynamics Simulation of Defect Structure Evolution and Mechanisms of Plastic Deformation in Aluminum Nanocrystals, Metallofizika i Noveishie Tekhnologii, 33(9): 1217-1247, 2011 (in Ukrainian).
[14] A.Gatsenko, A.Baskova, B.Bandalak, V.Tatarenko, Yu.Gordienko, Application of Desktop Grid Technology in Material Science, 5th Int. Conf. “Distributed Computing and Grid-technologies in Science and Education” GRID’2012, Dubna, Russia, pp.123-129, (http://www.researchgate.net/publication/258669295) 2012.
[15] Yu.Gordienko, E.E.Zasimchuk, L.P.Steblenko, A.N.Krit, D.V.Kalinichenko, The Oscillatory Nature of Internal Stress Relaxation due to Self-organization of Defect Substructure in Nanocrystals, Int. Conf. NANSYS-2013, Kyiv, Ukraine, (http://www.researchgate.net/publication/259338433) 2013.
[16] A.S.Gatsenko, E.E.Zasimchuk, V.A.Tatarenko, V.L.Karbovsky, Yu.Gordienko, Elastic Properties and Strength of Graphene Under Nanoindentation, Int. Conf. NANSYS-2013, Kyiv, Ukraine, (http://www.researchgate.net/publication/258872193) 2013.
[17] Yu.Gordienko, L.P.Sartinska, Thermal Stability of Boron Nitride Nanotubes, Int. Conf. NANSYS-2013, Kyiv, Ukraine, (http://www.researchgate.net/publication/259338325) 2013.
[18] S.Stirenko, O.Zinenko, D.Gribenko, Dual-layer hardware and software management in cluster systems, Proc. Third Int. Conf. “High Performance Computing” HPC-UA 2013, (Ukraine, Kyiv, 2013), 380-385.
[19] S.J.Haigh, H.Sawada, K.Takayanagi, A.I.Kirkland, Exceeding conventional resolution limits in high-resolution transmission electron microscopy using tilted illumination and exit-wave restoration, Microscopy and Microanalysis, 16(04), 409-415, 2010.
[20] F.Baudelet, R.Belkhou, V.Briois, A.Coati, P.Dumas, SOLEIL a new International Journal of High Performance Computing Applications, 540(1), 42-67, 2005.
[21] B.D.Patterson, R.Abela, H.Auderset, Q.Chen, F.Fauth, F.Gozzo, et al., tation of A: Accelerators, Spectrometers, Detectors and Associated Equipment, Nucl. Instrum. Meth. A: Accelerators, Spectrometers, Detectors and Associated Equipment, 50(1), 184-187, 1995.