

Concept of a Cloud Service for Data Preparation and Computational Control on Custom HPC Systems in Application to Molecular Dynamics

Dmitry Puzyrkov^{1,*}, Sergey Polyakov^{1,2,**}, Viktoriia Podryga^{1,3,***}, and Sergey Markizov^{1,****}

¹*Keldysh Institute of Applied Mathematics of RAS, 4, Miusskaya sq., 125047, Moscow, Russia*

²*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 31, Kashirskoe shosse, 115409, Moscow, Russia*

³*National Research Center "Kurchatov Institute", 1, Akademika Kurchatova pl., 123182, Moscow, Russia*

Abstract. At the present stage of computer technology development it is possible to study the properties and processes in complex systems at molecular and even atomic levels, for example, by means of molecular dynamics methods. The most interesting are problems related with the study of complex processes under real physical conditions. Solving such problems requires the use of high performance computing systems of various types, for example, GRID systems and HPC clusters. Considering the time consuming computational tasks, the need arises of software for automatic and unified monitoring of such computations. A complex computational task can be performed over different HPC systems. It requires output data synchronization between the storage chosen by a scientist and the HPC system used for computations. The design of the computational domain is also quite a problem. It requires complex software tools and algorithms for proper atomistic data generation on HPC systems. The paper describes the prototype of a cloud service, intended for design of atomistic systems of large volume for further detailed molecular dynamic calculations and computational management for this calculations, and presents the part of its concept aimed at initial data generation on the HPC systems.

1 Introduction

The study of the evolution of complex processes without ad hoc simplifications is a scientific problem of the highest interest. A powerful mathematical method for the description of the evolution of systems consisting of large numbers of particles (molecules, atoms) obeying the laws of the classical mechanics is the method of molecular dynamics (MD) [1]. The number of particles in the investigated systems can vary from dozens of individual particles to billions of particles. These particles interact "each with each" and this generates a high computational complexity in the large-scale MD systems. Such problems can be solved on high-performance computing (HPC) systems like supercomputers

*e-mail: dpuzyrkov@gmail.com

**e-mail: polyakov@imamod.ru

***e-mail: pvictoria@list.ru

****e-mail: arh11msn@yandex.ru

and GRID systems. In this paper, we discuss the concept and implementation of a dedicated web service intended to help scientists in HPC MD simulations especially as it concerns the data preparation and control of the computing process.

2 Formulation of the problem

The geometry of the system of particles under study can be very complicated. Not every MD simulator provides a means for describing such a geometry, and in most cases the task of consistent particle distribution rests on the shoulders of third-party applications.

The question arises of the method of designing such a particle system. In particular, to create meta-description files of geometry and physical conditions, a special editor is required, which will allow creating the necessary geometric model and immersing the necessary particle conglomerates into it.

In addition to the design problem, there is also the problem of managing the available computational resources and automated task launch in the user-accessible environment. This is required in order to maximize the efficiency and ease of use of the HPC systems and provide to the user a single access point for a set of such computers. This task has already been partially solved with the KIAM Job_Control [2] job management system, but a close integration of the design and accounting environment with this system is required.

In order to successfully implement a service which allows the user to solve the tasks of creating convenient computational domains for MD modeling as well as managing such calculations, the following components are necessary: a graphical geometry designer; applications for processing meta-description files; a computational tasks management system.

3 Existing solutions

As a system for HPC access and monitoring we can mention the “FlyElephant” platform [3]. FlyElephant provides access to different HPC clusters and helps users to send their computing tasks to them. For now, the platform support has access to 3 clusters and it can work with SLURM. This platform is not aimed on MD calculations in general, so any special tasks, such as data generation for simulators should be done by other software.

In the KIAM MolSDAG web service the HPC systems will be accessible by user credentials enabling the use of any cluster resources the researcher has access to. Apart from SLURM, the KIAM Job_control system supports the SUPPZ queue control.

4 The concept of cloud service architecture

In this paper, the concept of constructing and implementing the KIAM MolSDAG cloud service (Molecular Structure Designer and Generator) is presented, which in the long term will solve the problems posed above.

The main idea of this software package is to give to the final user a simple and universal system which will provide him with an interface for creating complex geometry, creating initial data sets based on this geometry and performing MD simulations.

The main components of the service are the editor of the geometry of the computational domain, the database of microstructures and their properties, a set of various parallel applications for generation atomistic data, as well as access modules to various file storages.

4.1 CAD-application

The graphical geometry designer is a CAD-like web application using WebGL (Web Graphics Library) for drawing a three-dimensional design area. In its function, first of all, the construction of a geometric model of a projected microsystem. This service has direct connection with the database of molecular dynamic calculations (DBMC), which allows the use of metadata of the already existing materials in the database.

4.2 MD data generators

In view of the possible large scale of the microsystem being created, at the stage of designing the geometry of the computational domain, there are no real procedures for generating and filling the region with particles which are processed. After the user has finished editing the structure, the output of the editor is a text-friendly, human-readable file in the YAML format. This file contains a meta-description of the projected particle system, including all properties and parameters necessary for its generation and simulation. Using this file, the KIAM MolSDAG scenario management system initiates the process of creating a script for generating initial data. The output script of this system is shown in fig. 1. It is a python script that describes the interactions between multiple generation programs. On the calculation run it will be executed by Job_control task runner module, providing actual generators execution on a target computational resource.

```
#job_control management system import
import wmcs
nitrogen_gen = wmcs.get_application("mxwgen")
nikel_gen = wmcs.get_application("fccgen")
unite_gen = wmcs.get_application("unite")
def applications():
    return [nitrogen_gen, nikel_gen, unite_gen]
def run():
    nitrogen_gen.run()
    nikel_gen.run()
    nitrogen_gen.wait()
    nikel_gen.wait()
    unite_gen.run()
    unite_gen.wait()
```

Figure 1. Scenario file for KIAM Job_control management subsystem

As a result of the execution of this task on the target system, the computational domain shown in fig. 2 will be created. The simulation task will use this data as an initial dataset.

4.3 Database of molecular dynamic calculations

The main purpose of such a database is to reduce the time required to generate and calculate typical structures. For example, for the task of simulating a binary metal-gas microsystem of nickel and nitrogen till the thermodynamic equilibrium is achieved, the materials, already reduced to thermodynamic equilibrium, can be used as initial data. This approach helps significantly reduce the time required for

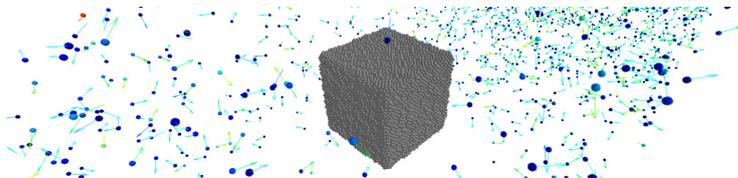


Figure 2. Visualized output of the generator scenario

simulation. By saying “database” in this context, we mean a set of programs for accounting various geometries, binary files and meta-descriptions of materials. The main components of this service are the database itself, a management system that provides API for interaction with the database, a set of modules for accessing distributed data [4], and a server-side client application. Replenishment of DBMC is performed by user calculations done with the help of the developed service.

5 Conclusion

The paper is devoted to the development of the KIAM MoSDAG cloud service and presents part of its concept, aimed for initial data generation on the HPC systems. The service is being developed for graphical design of large-scale microsystems and performing calculations. The purpose of creating such microsystems is a detailed molecular dynamic studying the processes of their evolution and the properties appearing in this operation. After graphical design of large-scale microsystems, the proposed system generates a Python language scenario and executes a generator programs based on this script. The generation process assumes both the creation (generation) of microsystems on the basis of new data, and their arrangement based on previously generated and studied microstructures. The developed service will contain in the future a large number of microstructure generators and programs for filling areas of complex shapes with particles of various types, as well as programs for combining pre-calculated data from various sources. Individual elements of the service have already been implemented and confirmed the effectiveness of the overall service concept.

Acknowledgments

The research was supported by RFBR (projects no. 15-07-06082-a, 15-29-07090-ofi_m) and by the Academic Excellence Project of the NRNU MEPhI under contract with the Ministry of Education and Science of the Russian Federation No. 02.A03.21.0005.

References

- [1] D. Frenkel and B. Smit, *Understanding Molecular Simulation. From Algorithm to Applications* (Academic Press, New York, 2002)
- [2] D. Puzyrkov, V. Podryga, S. Polyakov, and M. Iakobovskii, CEUR Workshop Proceedings **1787**, 416–422 (2017)
- [3] *FlyElephant – All-in-one automated workflow platform*, <https://flyelephant.net/>
- [4] D.V. Puzyrkov, V.O. Podryga, and S.V. Polyakov, Vestnik UGATU **20**, 175–186 (2016)