SHAPE Application Form

Project Title	Development of Chameleon Monte Carlo code for HPC: Toward Realistic Modelling of Complex Composite Systems
Company Name	Scienomics
Company Address	16 rue de l'Arcade 75008 Paris
Company Website	http://scienomics.com/
Company Number	
Number of Employees	10 - 20
Contact Name	Robin Chaudret
Contact Position	Research Scientist
Contact Email	Robin.chaudret@scienomics.com
Contact Phone	+33 6 61 43 20 28

Has the company worked with PRACE before?

No

What does the company do?

Scienomics is a French IT company established in 2004 specialized in the development Materials Modeling and Simulations Software. Scienomics' flagship "Materials Processes and Simulations platform (MAPS)" offers a unique combination of Modeling and Simulation Technology incorporating recent advances in theoretical and computational methods. Today scientists using MAPS worldwide are capable to simulate the properties and the behavior of several types of materials such as polymers, nanomaterials and composite and many others in order to achieve optimum design.

Scienomics with its headquarters in France and R&D Centers in Germany and Greece, and sales and support offices in UK and USA has established selected partnerships with the best-of-the-breed in academia through various partnership in France, Europe and abroad (e.g.: Université Pierre et Marie Curie, IDRIS, ENS Lyon, Fraunhofer Institute for Manufacturing Technology and Advanced Materials, Demokritos, Sandia National Labs, University of Shanghai, Pacific Northwest National Lab…). Finally, with high quality science and industrial solution oriented applied research, Scienomics has garnered collaborative projects with companies in the areas of energy and green chemistry, composites, alternate fuels, catalysis, polymers, defense, automotive, fine chemicals, etc.

Project Abstract

In order to perform de novo materials development and/or optimization scientists need to work with reliable models at the atomistic level. Increasing complexity of modern materials make these models become excessively complex and large and therefore exhibit characteristic that cannot be studied by standard techniques. In order to overcome this limitation Scienomics developed a software, called Chameleon, which combines Monte Carlo approach with modern algorithms for relaxing such materials, such as chain altering moves. Chameleon is now in its second phase of development and is currently able to simulate many different systems, further code optimization and validations is however still needed. The goal of this project is to improve Chameleon capabilities in several areas and test them on crucial industrial problems.

The access to a supercomputing center will allow Scienomics to improve and/or validate three different aspects of the Chameleon code:

1- Improve Chameleon performance for single processor simulations
2- Improve Chameleon parallelization:
    a. increase the part of the code parallelized
    b. improve the parallelization scaling for large number of processor
    c. test hybrid implementation
3- Scienomics will leverage the large supercomputer capabilities to run both long simulations and a statistical number of simulations in parallel to validate the ergodic theorem

These developments and validations will then be applied to optimize the polymer formulation for the mechanical properties of graphene/polymer composites. Such systems are crucial in many industrial sectors.

These simulations will involve the generation of very large systems. Scienomics will then exploit powerful machine memory capabilities to test the limitations of the MAPS platform capabilities to generate and visualize these very large systems (over 1 million atoms). Based on these tests, improvement of MAPS capabilities, such as porting MAPS to GPU/CPU architecture or improving MAPS code to speed visualization of large system for example, will then be considered.

Industrial relevance and potential business impact

The MAPS platform developed by Scienomics is fundamental to chemical and materials design and process improvements in industry ranging from aerospace to personal care, semiconductors to pharmaceutical development, chemical to manufacturing. MAPS provides unique insights into complex chemistry that can never be obtained by any experimental technique or from engineering modeling only. Chameleon was initially specifically developed to allow our industrial and academic customers to study systems which cannot be reliably simulated otherwise. 'Unphysical' chain altering moves relax a periodic system through very efficient phase space sampling. Such software has the potential to become a key selling asset

for Scienomics, but also to be viewed as a scientific breakthrough for polymer simulations. However, Chameleon performances are currently limiting the size and type of systems treated. Indeed it can currently only be run sequentially and therefore simulations sometimes last for months. Such characteristics are therefore clearly limiting its applicability to problems of real industrial interest.

Improving its speed and parallel capabilities will therefore allow Chameleon being used on systems it is developed for, greatly increase the number of industrial domains it can be applied to and model more realistic phenomena. Such developments will not only allow Scienomics to reach completely new markets but it will also increase service quality and customer satisfaction by allowing them to model more realistic systems quickly and efficiently by lowering simulation times. The systems we intend to study as part of this project are polymer/graphene composite systems, where the polymer can for example be a polycarbonate or a polyamide.

Since MAPS platform is the main product developed by Scienomics, any development to improve its speed-up and treat larger scale systems will greatly benefit Scienomics customers who will be able to generate more realistic systems and study more complex phenomena. Such improvement will therefore greatly contribute to attract new industry customers and improve customer satisfaction as well.

Proposed high-level Work Plan

Chameleon is a software interfaced with MAPS platform and developed by Scienomics. The first part of the project will therefore be to install both of them on the computer center. Chameleon is currently in its second development phase which means that it is currently able to simulate many different systems but several code optimizations and validations are still needed. The following task plan has been identified not only to take advantage of engineers expertise to allow to tackle the problem at all different aspects (improving sequential scalability, improving parallel capabilities, improving implementation of the code and port it for GPU and CPU/GPU architectures) but also to use the supercomputing resources validate the software and apply it on real industrial problem.

Once MAPS and Chameleon installed is on the server, the first goal will be to find out how Chameleon speed-up can be increased. The two different following tasks will therefore be given high priority and started first:

- Task 1: Change Chameleon code to improve its sequential speed-up (SME effort 0.5 PM, PRACE 0.5 PM)
- Task 2: Change Chameleon code to improve parallelization (SME effort 1 PM, PRACE 2 PM)

The next step will then be to test the validity of the ergodic hypothesis on Chameleon results

- Task 3: Validation of ergodic theorem (SME effort 1 PM, PRACE 0 PM)

The application of Chameleon to a crucial application as outlined above will allow to verify the improvement of Chameleon developments done as part of this project:

- Task 4: Application of Chameleon (SME effort 2 PM, PRACE 0 PM)

Finally MAPS performances for large scale systems will be tested and improved:

- Task 5: Test and improvement of MAPS performance for large systems (SME effort 0.5 PM, PRACE 0.0 PM
- Task 6 : Final report (white paper/publication) (SME effort 0.5 PM, PRACE 0.5 PM)

Total: SME 5.5 PM - PRACE 3.5 PM

Technical and business requirements

Compute Resource

Existing architecture	x86/x64 linux cluster
Preferred architecture	InfiniBand interconnexion Large memory capabilities IDRIS clusters
Parallelisation strategy	MPI, OpenMP
Storage (Gbyte)	1 To
Third party software	Numerical chemistry softwares (open-source) + some licenced software already installed on IDRIS clusters
Typical run	100 hours on 32 cpu 20 hours on 256 or 512 cpu (in function of the development)
Core hours	1,000,000 hours (one million hours)
Memory	20 Go per core

We believe that working on IDRIS clusters would allow a fast and easy start of the project. Indeed we are already in close contact with one of the members of the IDRIS support team (Christophe Narth). Besides this existing working relationship the third party softwares that we might need are already installed on the IDRIS cluster.

As part of this project we plan to model real industrial problem based systems. In order for these systems to be realistic, their size must be very large. Resources needed for the calculations (memory) as well as simulation times will therefore be accordingly long. As mentioned in the compute resource request a typical run will range between 3200 and 5000 core-hours. At least 100 runs will be needed to verify the ergodic theorem. Additionally the production runs will involve simulating three different type of composites and there 3 times more simulations will be needed. We believe the 1,000,000 core-hours will be needed for that project.

Non-technical resource

Confidentiality	No