To Compute/Calcul Canada : White Paper for Computational-Based Materials Research

Discipline-Based Group

The following list represents a group composed mostly of scientists and engineers working in the area of materials. Given the summer period, it was not possible to obtain answers from everyone. Those who have answered told us that this document represents in broad terms their needs for computing infrastructure, although individual needs do vary.

UBC
Jörg Rottler University of British Columbia jrottler@physics.ubc.ca ; tsc-230-01
Gren Patey <patey@chem.ubc.ca>
Matthias Militzer matthias.militzer@ubc.ca

Alberta
Massimo Boninsigni (Edmonton) m.boninsegni@ualberta.ca

Manitoba
Byron Southern Byron.Southern@umanitoba.ca ; eah-180-01

McMaster
An-Chang Shi shi@mcmaster.ca
Erik Sorensen sorensen@mcmaster.ca; zuz-406-01
Jeffrey Hoyt hoytj@mcmaster.ca

Guelph
Robert Wickham (Guelph) rwickham@uoguelph.ca

Waterloo
Roger Melko rgmelko@uwaterloo.ca>
Pierre-Nicholas Roy pnroy@uwaterloo.ca
Michel Gingras mjp.gingras@gmail.com
Jeff Chen jeffchen@uwaterloo.ca
Mikko Kartunnen mkarttu@gmail.com

Sherbrooke
David Sénéchal david.senechal@usherbrooke.ca
André-Marie Tremblay andre-marie.tremblay@usherbrooke.ca esc-193-01

Polytechnique
Alain Rochefort alain.rochefort@polymtl.ca

Université de Montréal
Normand Mousseau normand.mousseau@umontreal.ca ; knp-440
Laurent Lewis laurent.lewis@umontreal.ca
Michel Côté michel.cote@umontreal.ca

McGill
Nikolas Provatas provatas@physics.mcgill.ca
Martin Grant martin.grant@mcgill.ca
Hong Guo guo@physics.mcgill.ca
Kirk Bevin kirk.bevan@mcgill.ca; qdq-860

St. Francis Xavier
Peter H. Poole ppoole@stfx.ca rhw-593-01

Memorial
Martin Plumer (Memorial) plumer@mun.ca; vtq-722-01
Jolanta B. Lagowsk jolantal@mun.ca xdg-523-01
Ivan Saika-Voivod saika@mun.ca dfi-395-01

The main authors of this document are, by alphabetical order of last name
Michel Côté michel.cote@umontreal.ca
Michel Gingras mjp.gingras@gmail.com
Nikolas Provatas provatas@physics.mcgill.ca
André-Marie Tremblay andre-marie.tremblay@usherbrooke.ca

Science description

Much of today's industry relies on advanced materials whose development from discovery to commercial product takes decades. The United States is funding the "Materials Genome Initiative" to "speed up our understanding of the fundamentals of material science, providing a wealth of practical information that entrepreneurs and innovators will be able to use to develop new products and processes". In particular, "The initiative funds the development of computational tools, software, new methods for material characterization, and the development of open standards and databases." This type of fundamental research is as important for Canada as it is for the United States and the world. The students that will participate in the development of such fundamental knowledge will be especially skilled to recognize and capitalize on the opportunities brought about by this knowledge and to understand the advantages and limitations of the tools necessary to apply the fundamental research.

Current use of Advanced Research Computing

We collectively make use of all major Compute Canada (CC) systems on computational simulations relevant to materials science research. Collectively, our resource allocation is more than 10 000 cores-year.

Non-CC resources that we use are mostly those obtained through Canada Research Chairs applications that were awarded several years ago. Many of these installations have been integrated in existing CC resources and are shared with the community.

For current level of usage storage, compute, cloud, portals/gateways, accelerators, please aggregate the current usage of our groups.
Our community uses mostly "homemade" software packages. These are software developed specifically to study unique never-before-answered questions about materials properties and processes. Often, they constitute forefront computational physics & materials developments. As such, the graduate students involved in such research have in the course of their physics-based research developed highly technical information-science (IS) and computational method expertise that make them uniquely positioned to enter the computing and IS job market. That being said, there is an increased trend towards the use of special-purpose libraries, as well as in standardizing some of the world-recognized software that some of the members of this team have developed as part of their Compute Canada-sponsored research.

Some research groups in our community are also involves in large international projects aimed at the development codes for Material Discoveries such as the softwares for band structure calculations Abinit (http://www.abinit.org) and Wein2k (http://www.wien2k.at/) and the libraries for model-based problems ALPS (http://alps.comp-phys.org/) and TRIQS (http://ipht.cea.fr/triqs/). These are highly optimized codes which can make efficient used of modern computational resources. They are being actively developed and used on by groups all over the world.

Future growth

The growth in our field is driven mostly by the need for, and success of, realistic and experimentally relevant real materials simulations. Materials are studied on multiple length and time scales and the methods vary according to those scales.

At the smallest microscopic level, quantum mechanics reigns. Solving exactly a general quantum mechanical problem is exponentially difficult in both time and storage as the number of atoms considered grows. However, modern methods based on density functional theory and on methods developed for model Hamiltonians, combined with various Monte Carlo methodologies allow one to treat the most difficult problems with

sufficient accuracy to predict, for example, electronic transport properties of transition metal oxides. These materials, containing electrons in d or f orbitals are very difficult to treat with conventional methods, exhibit interesting thermoelectric, superconducting, magnetic, ferroelectric and multiferroic properties – some of which have potential for significant applications (if not paradigm-shifting in the case of superconducting or magnetic materials, for example). Even when quantum mechanical effects are not dominant, large levels of magnetic frustration or inhomogeneities in the solution of statistical mechanics problem or of finite-difference equations necessitate vast resources.

At the next level up, the use of molecular dynamics allows researchers to use potentials developed from quantum mechanical techniques to study the interactions of atoms on nano-meter and nano-second length/scales in order to understand defect, interface and grain boundary energies, as well as the physics governing microstructure formation of defects at the nm scale. These computational investigations form a current research forefront as they bridge the gap between atomistic simulations (described above) and the exploration of new phenomena and ultimately technologies that forms the basis of nanotechnology.

On the nanometer-micron scale the uses of classical density functional theory using correlation functions derived from or motivated by molecular dynamics are coarse-grained to derive so-called “phase field theories” based on real and complex mathematical quantities called “order parameters”. These microscopic models allow researchers to incorporate elasticity and crystal plasticity – effects inherently atomistic in their origin – on diffusive time scales on which the most relevant phase transformations governing materials microstructure occur. Such models have been instrumental in our understanding of self-assembly and microstructure formation in metal alloys, semiconductors and liquid crystal systems.

On length scales above those described above, the grand challenge of computational materials science is to connect continuum theories of heat, momentum and mass transfer that couple to the aforementioned microscopic theories – directly or via “well-informed” parameterization. This is another research frontier where ultimate progress will open an unprecedented window for multi-scale modeling of materials processes in various industries

To meet the challenges articulated by the above multiscale scientific problems, however, the most crucial link is the availability of a large amount of accessible compute cycles and data storage. The fastest possible level of growth, especially in computing power, is needed. Realistically, it is crucial for our community to update the CC facilities to at least keep up with Moore’s law. This is the only way we can remain competitive on the international scene. Permanent storage requirements must grow to remain commensurate with the increased speed in computing power since this is the main source of data generation.

Technical Details of Future Need 2016-2021

- Data
  - Our data is expected to grow about the same rate as our CPU usage. As data grows, we will have to pay special attention to the way we manage this data
    - We should continue to improve the infrastructure to easily share our results in a secure way between our Canadian colleague but also our international collaborators.
    - The structure of our data should be standardized and should contain all of information needed for the reproducibility of the results.
  - The largest data sets are generated through Monte Carlo simulations of three-dimensional (3D) material system grids. Typically, over a terabyte of data per group per year is generated (?). This data is transferred from the calculation to the disk at a high rate so the disks must be close to the CPU. However, over a four-year time scale roughly, the data has been analyzed and reused for various "post-processing" statistical analysis and the size of the data that must be archived over the long term is much smaller.
  - Part of this data and some of the programs used to generate will continue to be shared among several international research groups and communities. There are already web sites that store crystal structures and density functional results for band structure calculations for example. There are also several groups in Canada that have developed, and continue to develop density functional theory (DFT)/molecular dynamics/phase field codes for execution via cloud interfaces, some of which being done by the private sector. We expect that this trend will continue, as will the need for larger storage for wider use by international communities. We foresee new types of data bases that might include detailed numerical solutions of model systems obtained with methods that are considered essentially exact and industrially relevant. These can be used then as benchmarks for much faster but more approximate new methods. These new types of databases would be based for example on Structured Query Language (SQL) and accessible through web interfaces.
- Computation
  - Computational materials physicists/chemists require both parallel and serial processing. Communication between processors is not generally a bottleneck, however. Much of our work involves studying various values of controlled external parameters that drive the physics of interest, such as temperature, pressure, stress. This exploration of parameter space can be done with several jobs, each of which involving between 100 and 1000 cores. A few gigabytes of memory per core usually suffice, but we need as many cores as possible to make a thorough exploration of parameter space that allows for the discovery of new phenomena and guide the experimental search of these
  - GPU's have begun to be used for specialized applications, for example in DFT code such as in the Abinit code. The need is expected to grow, but not before these systems become more general purpose machines. Other

acceleration hardware solutions will be explored as they are introduced but we expect that our needs to be satisfied with “conventional” computational resources for the foresee future up to 2021.
    - Some test with some part of our software show a clear potential with GPGPU. Even if the software and library for using GPGPU, coprocessor (Xeon Phi) and FPGA are improving in the last years, there is still a lot of human time needed to migrate our actual software to this new technology. To deal with this fast evolving exotic hardware for the next year, the training and help for optimizing codes will be as important as the hardware himself.
- Software
    - New software and middleware is very likely to be needed when new types of data bases of the type mentioned above become more widespread in our group.
- Networking
    - It's advisable to keep the big data near the big compute center and avoid sparse data in each university. In the future, more resources may also go in the "cloud". On the other hand, we will also always need a good access to the data for post-treatment and visualization purpose. To achieve this goal, the network should be constantly improved to give an excellent connection between the data center the researchers. In the past, efforts have been made to connect data centers between them but the effort to improve connection to the researcher in his lab to the CANARIE network is crucial.

Specific Need of our community for 2016-2021

As mentioned above, our main computational needs are CPU. Considering our present usage of around 10 000 cores-year, we would need the equivalent resources of 50 000 cores-year *of present technology* in 2021 to remain competitive. Evidently, more will be better as we expect our community to grow in the coming years. Half of those CPU should be serial type machines with low communications and the other half with better communication hardware at the level of Infiniband but even for those highly connected machines, we favor CPU over communications.

We would like to stress that, as mentioned above, the importance of GPU, FPGPU and GPGPU for our field is evolving rapidly. Specific needs for these will be discussed in more details in further documents.

Finally, we indicated that our community is heavily involved in code development. Also, more and more of these codes are part of large scale projects where code is shared. In this context, the importance of support personnel cannot be overemphasized.