Enclosure 1b. Category 1 Application form – English version

APPLICATIONS ARE PREFERABLY DRAWN UP IN ENGLISH. AN ENGLISH TRANSLATION HAS TO BE ENCLOSED WITH APPLICATIONS SUBMITTED IN DUTCH.

The application form is available in English on the website https://vscentrum.be/.

Title of the application: **Quantum contributions to the polarizability**

Name and first name of the applicant: **Wouters Sebastian**

Institution: **Ghent University**

Research group / department: **Center for Molecular Modeling**

Title / position: **Post-doctoral researcher**

e-mail address: **sebastian.wouters@ugent.be**

Total computing time that is needed, in node-days: **1568 node-days**

Total disk storage that is applied for: **300 GB**

1. Title of the research project (with IWETO link if available) within the framework of which computing time is applied for:

The title of the FWO grant of S.W. is "Variational Renormalization Group methods applied to molecular systems"

2. Short description of the research project within the framework of which computing time is applied for (max. 1 A4 in Arial 12):

To study the dynamic behaviour of large chemical systems, its interactions are often parameterized with classical force fields. Traditionally, a force field is a sum of atomic two, three or four body-interactions. Polarizable force fields make better approximations of the interatomic interactions by explicitly accounting for electronic polarization. [e.g. see Verstraelen et al. J. Chem. Phys. **138**, 074108 (2013)] To develop such models, one must go beyond the basic dipole polarizability and break down the linear response in local atomic contributions [Krishtal et al., J. Chem. Phys. **125**, 034312 (2006)]. Such level of detail is needed to describe how two nearby molecular fragments mutually polarize each other, leading to the additional attractive forces that are of interest for force-field simulations.

In this work, we want to assess this partitioning by comparing it with the exact static polarizability for several well-chosen points along the ground-state dissociation curves of homonuclear first-row dimers: Li2, Be2, B2, C2, N2, O2 and F2. The total and partitioned static dipole polarizabilities can be obtained with the finite-difference method, by means of several ground-state calculations [Wouters et al., J. Chem. Phys. **136**, 134110 (2012)]. The partitioning will provide additional insights in how to model the quantum mechanical contribution to linear response in polarizable and reactive force fields. In order to obtain reliable results along the entire dissociation path, a high-level electronic structure method is needed.

An exact electronic ground-state calculation corresponds to finding the smallest algebraic eigenvalue of the corresponding Hamiltonian, a large symmetric matrix. The linear dimension of this matrix grows exponentially with the number of single-particle degrees of freedom, which renders the exact solution for all practically interesting cases infeasible. Fortunately, certain low-rank approximations of the exact solution converge quickly to

the exact solution with increasing rank. One such low-rank approximation is the matrix product state (MPS), which can be optimized by means of the density matrix renormalization group (DMRG). By exploiting the symmetry group of the Hamiltonian, an additional significant lowering of the computational cost is achieved. We have our own high-performance implementation of DMRG for quantum chemistry, CheMPS2, which was released recently [Wouters et al., Comput. Phys. Commun. **185**, 1501 (2014)].

3. Financing institution or channel, financing the research project in full or in part (FWO, BOF, IWT, EU, ...): Please attach the confirmation letter as enclosure (see instructions in enclosure 3).

FWO

4. Promoter of the research project:

Prof. Dr. ir. Toon Verstraelen

5. Persons mandated by the Applicant to compute on the TIER1 within the framework of the present project: Please provide for every person:

- name and first name
- institution
- research group / department
- title / position
- experience with TIER1/TIER2 infrastructure in Belgium and abroad

Wouters Sebastian

Ghent University, Center for Molecular Modeling, Post-doctoral researcher

Experience with compiling, debugging, and profiling (PerfExpert, Intel VTune Amplifier) software on

1. the Stevin Supercomputer Infrastructure at Ghent University
2. SHARCNET (Compute Canada national HPC)
3. TIGRESS, the HPC at Princeton University

Related links:

- https://github.com/SebWouters/CheMPS2
- http://dx.doi.org/10.1063/1.3700087

- http://dx.doi.org/10.1016/j.cpc.2014.01.019
- http://arxiv.org/abs/1405.5642

Verstraelen Toon

Ghent University, Center for Molecular Modeling, Associate Professor

Experience with compiling, debugging, and profiling software on

1. the Stevin Supercomputer Infrastructure at Ghent University
2. SHARCNET (Compute Canada national HPC)
3. the HPC system of the DAMP group (Prof. G. Maurin, Univ. Montpellier).

Contributions to the easybuild project spearheaded by the HPC team of the Ghent University. Technical support for colleagues working with SCOOP on TIER-1.

Related links:

- http://molmod.ugent.be/software
- http://dx.doi.org/10.1063/1.4791569
- http://dx.doi.org/10.1063/1.3187034
- https://github.com/tovrstra

6. Description of the computing task, justification for the computing time, disk storage and memory that are applied for, and description of the software tools required (max. 3 A4 in Arial 12). Please clearly provide the following in this regard:

- the number of nodes/cores that are applied for per computing task, with a subdivision of the computing time in sub-tasks indicating the sequence of the sub-tasks
- whether these tasks use diversification (OpenMP, MPI, hybrid OpenMP/MPI …)
- the estimated memory use of a computing task (maximum 64GiB/node)
- whether a vSMP system will be used
- the requirements for disk storage (estimated volume in GiB and the total number of files), more specifically for:
  - required input files (data set, parameter files, etc.)

- SCRATCH volume used during the performing of the computing tasks
- result files

We want to study the exact and partitioned static polarizabilities of 7 first-row dimers at 8 well-chosen points along their ground-state dissociation curves with CheMPS2. For each case, 7 calculations are needed: one for the ground-state energy, as well as six for the shifts in ground-state energy due to small finite-field operators. Based on these 7 calculations, the total polarizability as well as its partitioning can be determined with the finite-difference method.

To obtain relevant results, we need to use a triple-ζ Gaussian basis set with polarization functions (cc-pVTZ), which yields for each of the first-row dimers a total of 60 orbitals as single-particle degrees of freedom. A single calculation is independent of the number of electrons, and depends only on the number of orbitals, the desired accuracy (always mE_h in quantum chemistry), and the excitation spectrum. One geometry of C_2 in the cc-pVTZ basis can be converged to mE_h accuracy in 72 to 120 hours on a single node of the delcatty cluster (using OpenMP) of the Stevin Supercomputer Infrastructure at Ghent University, depending on the excitation gap at the particular geometry. Because the nodes of delcatty are practically identical to those of the TIER1 (same number and type of cores), this yields an estimate of on average 4 node-days for each of the 7 x 8 x 7 calculations, or 1568 node-days in total. A schematic overview of the required CPU time:

Estimate for one CheMPS2 calculation	
On delcatty (1 node, 16 cores @ 2.66 GHz)	72 - 120 hours
On TIER1 (1 node, 16 cores @ 2.66 GHz)	4 days on average
Total number of CheMPS2 calculations (7 x 8 x 7)	392
Number of CheMPS2 calculations per dimer per geometry (1 ground state + 6 perturbed)	7
Number of geometries	8
Number of dimers	7
Total number of node days (4 x 392)	1568

Each calculation can be performed independently of the other calculations, on a single node. The current version of CheMPS2 only supports OpenMP parallelization. One ground-state calculation may therefore exceed the walltime limit (3 node-days). Still, this is no problem in practice because CheMPS2 can create checkpoints to continue calculations at a later point.

A checkpoint will have a typical size of 200-500 MB for the cc-pVTZ basis set, much smaller than the 45 to 60 GB memory which will be consumed during the calculations. The input (matrix elements) and output (reduced two-particle density matrix) files will be of size 100-110 MB. So a total of 500 + 110 + 110 = 720 MB disk per calculation is required. For the 280 calculations which will be performed, a total of 720 MB * 392 calculations ≈ 300 GB disk storage will suffice.

In a DMRG calculation, renormalized operators are temporarily stored to disk (59 files for the 60 orbitals of the first-row dimers). Per calculation for the first-row dimers in cc-pVTZ basis set, about 400-800 GB local scratch will be required at runtime.

The required computational resources are summarized in the following table:

CPU time	1568	node-days
RAM Memory per node	45 to 60	GB
Local scratch space per node	400 to 800	GB
Storage for input and output files	300	GB

7. Please indicate why the TIER1 is the appropriate machine to perform the computing task (max. 1/2 A4 in Arial 12):

Given the large computational workload, this project requires a high-capacity HPC infrastructure. Running such calculations on a TIER2 would take too long. The CheMPS2 software needed for this project is already well tested and tuned for performance on the TIER2 cluster delcatty, which has identical nodes as the TIER1. Hence we can immediately start with production runs and we can guarantee an optimal usage of the computing time on the TIER1 infrastructure.

8. Summary of the software required to perform the computing task, and possible installation and compilation instructions (max. 2 A4 in Arial 12). Please clearly provide the following per item in this regard:

- a reference to the software's web page
- the software licence system (open source, GPL, etc.)
- if there is no free academic use of the software, state which licence makes the installation and the use valid on the TIER1 by the Applicant (+ add a copy of the signed licence)
- if need be, which licence server will be used (name + IP address)
- whether the software is already available on the TIER1 (see https://vscentrum.be/nl/tier1-rekenen) and, if this is not the case, compilation and installation instructions (possibly with reference to existing TIER2 installation)

Software: CheMPS2

- developed in-house
- profiled with PerfExpert and optimized
- https://github.com/SebWouters/CheMPS2
- http://dx.doi.org/10.1016/j.cpc.2014.01.019
- License: GNU GPL v2 ; free open-source software

CheMPS2 is not yet available on the TIER1 and requires

1. GNU Scientific Library: http://www.gnu.org/software/gsl/ (easily installable with Easybuild)
2. HDF5 (which is installed)
3. Intel MKL (which is installed)
4. OpenMP (which is installed)

CheMPS2 has a CMake build system, and instructions are provided on the GitHub webpage. S. W. has experience with compilation and profiling on the Stevin Supercomputer Infrastructure (TIER2) at Ghent University. Only the GNU Scientific Library should be provided, and we can compile CheMPS2.

9. Period during which the task is to be performed:

July - September 2014

10. Describe the results that were obtained within the framework of computing time that was attributed during the past two years on the TIER1 or on other TIER1 or TIER0 supercomputers (max. 2 A4 in Arial 12):

CheMPS2 has only been used on the TIER2 Stevin Supercomputer Infrastructure so far. In the first quarter of 2014, it was profiled and extensively used on the delcatty cluster, with nodes identical to the TIER1 nodes. Neither S. W. or T. V. has previously worked on the TIER1. However, several of our close colleagues at the Center for Molecular Modeling have successfully completed TIER-1 projects and T. V. was involved in technical support for some of these projects.

Should you have any questions or encounter any difficulties during the electronic submission of an Application, please contact by e-mail:
Associatie KU Leuven: leen.vanrentergem@kuleuven.be ; jan.ooghe@kuleuven.be
Associatie Universiteit Gent: hpc@ugent.be
Associatie Universiteit Hogescholen Antwerpen: stefan.becuwe@uantwerpen.be
Associatie Universiteit Hogescholen Limburg: geertjan.bex@uhasselt.be
Universitaire Associatie Brussel: rosette.vandenbroucke@vub.ac.be ; rosette.vandenbroucke@gmail.com
For the other institutions: dane.skow@herculesstichting.be