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://www.vscentrum.be/en/access-and-infrastructure/project-access-tier1

Title of the application:

Error estimates for ab initio predictions of surface energy and work function

Name and first name of the applicant:

De Waele Sam

Institution:

Ghent University

Research group / department:

Center for Molecular Modeling (CMM)

Title / position:

PhD Fellow

e-mail address:

sdwaele.dewaele@ugent.be

stefaan.cottenier@ugent.be

Total computing time that is needed, in node days:

1165

Total disk storage that is applied for (in GiB):

2500 GiB

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

"A firmer framework for the reliability claims of solid-state density-functional theory."

FWO-project G0E0116N, promotor S. Cottenier.

2. Describe your research project in short. Explicitly mention the scientific questions that you are planning to address and the overall scientific goals of the project. (max. 1 A4 in Arial 12):

The use of density-functional theory (DFT) to predict materials properties becomes ever more widespread. With this increased use comes the demand for better quantitative determinations of the accuracy of DFT results. The aim of this work is to compare calculated surface energies and work functions with experimental data, thus gaining insight into how accurately surfaces are modeled with DFT. The two most prevalent functionals, the local density approximation (LDA) and the Burke-Perdew-Ernzerhof parameterization of the generalized gradient approximation (PBE-GGA), will be used to obtain numerically very precise DFT-predictions for surface energies and work functions for a large test set of materials. The comparison with experiment will be made by performing linear regression, which results in a measure for the accuracy of the theoretical prediction.

Because the primary goal is comparison of *ab initio* theory to experiment, the elementary materials for which reliable experimental data are available are the materials of choice for this study. In total, 3 different surfaces of 53 materials will be simulated, each with both the LDA and GGA-PBE functional. This results in a total number of 318 simulated surfaces. The model parameters, most importantly the sampling in the first brillouin zone and the number of atoms in the slab unit cell, need to be determined for all these surfaces as well. This is done by selecting a few electronically diverse materials for each crystal structure and performing rigorous convergence tests for that small set. Subsequently, general settings that render satisfactory computational precision can be selected for all materials of the crystal structure. This strategy helps to reduce the total wall time drastically.

Performing this research will provide an error bar for a calculated surface energy or work function. The error analysis is not just an evaluation of a DFT functional, but ultimately determines the usefulness of DFT-calculated surface properties for experimentalists. Especially for surface energies, for which accurate experimental methods are hard to come by and extremely material-sensitive, an error bar is an essential addition to a quantitative result.

3. Provide an abstract (10 lines) for scientific communication on the website in layman's terms. See also item 12 of this application form.

Density-functional theory (DFT) is today one of the most prevalent theoretical methods to predict materials properties. Although it is a computational method which requires no experimental input, several approximations on the quantum-mechanical level are made which can result in inaccuracies with respect to actual experimental results. Assessing this inaccuracy is vital if DFT results are to be used as guidelines for experimental work. Moreover, it can help to identify significant shortcomings of DFT. The goal of this project is to determine a quantitative error bar for two important surface properties: the work function and the surface energy.

4. 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 "EasyChair proposals submission procedure").

FWO

5. Name and email address of the promoter(s) of the research project:

stefaan.cottenier@ugent.be

6. Persons mandated by the Applicant to compute on the Tier-1 within the framework of the present project: Please provide for every person:
   - name and first name
   - institution
   - research group / department
   - title / position
   - experience of using HPC resources in the past (Tier-0/Tier-1/Tier-2 infrastructure in Belgium and abroad)

Ir. Sam De Waele (vsc40941)
Ghent University
Center for Molecular Modeling
PhD Fellow
Over 2 years of experience with Wien2k, VASP, gaussian and CP2K on TIER2

Dr. ir. Kurt Lejaeghere (vsc40323)
Ghent University
Center for molecular modeling
BOF Postdoc
Over 5 years of experience with VASP on both TIER2 and TIER1

Prof. Dr. Stefaan Cottenier (vsc40026)
Ghent University
Center for Molecular Modeling
Assistant Professor
Experience with a wide variety of DFT packages since the opening of the Ghent VSC on both TIER2 and TIER1.

7. Explain why this project needs to run on a Tier-1 system, why the machine you have requested is suitable for the project and how the use of the system will enable the science proposed (max. ½ A4 in Arial 12).

The challenge of this project does not lie in massive parallelization, but in efficiently dealing with a large number of single-node jobs. There will be periods of intensive computation, followed by periods of human analysis in order to prepare the next round of calculations. TIER1-access is essential for a steady flow of data, minimizing the bottleneck effect of problematic cases, and ensuring sufficient focus during the intended short duration of this project.

8. Justify the number of node days requested. This should include information such as: number and nature of computing tasks, software used, and the sequence in which they will be performed. Indicate for each typical computing task the required resources:

   - wall clock time (note that 3 days is the maximal wall clock time for any job; checkpointing should be used for longer run times)

- memory (maximum 64 GiB/node)
- number of nodes
- number of CPU cores
- disk space (estimated volume in GiB and the total number of files); make a clear distinction between usage of Tier-2 DATA/HOME partitions and the Tier-1 SCRATCH partition
- number of tasks, and an indication of how many such tasks would be submitted concurrently.

This information should take the form of a table (an example is provided as Table 2 in the appendix). Provide additional descriptions of the computing tasks and comments as needed. Resource estimates should be preferably based on the results of actual calculations on Tier-1 (via, e.g., a Starting Grant) for system/problem sizes that are on par with those of the intended computing tasks (e.g., same mesh sizes, actual molecular system, …). If not, provide the name, architecture, #cores, memory, etc. of the machine that was used to obtain these results and explain how you have calculated/rescaled the wall clock times, number of cores, etc.

The resource estimates for the tasks at hand were estimated by performing the calculations on the **delcatty** cluster (ppn=16, physmem=62.9GB, swap=20.0GB, vmem=82.9GB). The estimates are valid for a “general” material. Potassium which has a VASP-recommended PAW accounting for 9 valence electrons is deemed to be a good representative of such a general material. The first three tasks in table 1 are only performed for 16 materials, as these are the test materials from which the general model parameters are derived. The relaxation task is done for the 37 other materials on the basis of those rigorous tests. Again using the test results, for all 53 materials the calculations are repeated with the exact same settings but using the LDA functional. After all surfaces are relaxed, for each of them a final point calculation is performed with tetrahedron smearing. This smearing method is unfit for the relaxation of metal surfaces, but yields the most accurate single point calculation.

Table 1: Estimates of required node days and storage

Task	# tasks	Walltime (days)	# node days / task	Max memory (GiB) / task	#cores per node	#nodes / task	Tier-2 storage (GiB) + inodes	Tier-1 storage (GiB) + inodes
k-point test	16*3*15 = 720	0.054	0.054	4	16	1	720 * 0.2 GiB =144 GiB 720*1 inodes	720*1GiB = 720 GiB 24*72 = 17280 inodes
Layer test	16*3*15 = 720	0.133	0.133	8	16	1	720 * 0.2 GiB =144 GiB 720*1 inodes	720 GiB 17280 inodes
Relaxation test	16*3*6 = 288	1.771	1.771	6	16	1	288 * 0.2 GiB = 56 GiB 288 inodes	288 GiB 6912 inodes
Relaxation	(37 + 53) * 3 = 270	1.771	1.771	6	16	1	54 GiB 270 inodes	270 GiB 6480 inodes
Final Point	53*2*3 = 318	0.133	0.133	6	16	1	63.6 GiB 318 inodes	318 GiB 7632 inodes
Total	2316	**1165**					461 GiB 2316 inodes	2305 GiB 55584 inodes

9. Describe the software required to perform the computing task(s). Please clearly provide the following per item in this regard:

- a reference to the software's web page
- the software license system (open source, GPL, etc.)
- if there is no free academic use of the software, state which license makes the installation and the use valid on the Tier-1 by the Applicant (+ add a copy of the signed license)
- if need be, which license server will be used (name + IP address)
- whether the software is already available on the Tier-1 (see https://www.vscentrum.be/cluster-doc/software/tier1-muk) and, if

this is not the case, compilation and installation instructions (possibly with reference to existing Tier-2 installation)

Provide the results of scaling tests that were conducted with this software, preferably on Tier-1 (using, e.g., a Starting Grant) for system/problem sizes that are on par with those of the intended computing tasks (e.g., same mesh sizes, actual molecular system, …). If not, provide the name, architecture, #cores, memory, etc. of the machine that was used to obtain these results.
Provide both a table and scaling plot such as table 1 and plot 1 in the appendix (max. 2 A4 in Arial 12).

Used software:

VASP 5.4.1
http://www.vasp.at

The license is attached at the end of the document.
The software is already available on the TIER1 infrastructure.

Scaling tests:

As mentioned in section 7 and evident from section 8, single-node jobs form the bulk of this research. Since all TIER1 jobs are full node jobs, it is necessary to determine the most efficient use of a full node and to prove the use of a full node is actually cost-effective.

To derive the optimal method of parallelization for a single node, benchmark tests are carried out for a potassium {100} surface. A slab of 10 atomic layers (which entails a 10-atom unit cell, a 90-electron system) is used with a 13x13 grid in reciprocal space (28 irreducible points). Two methods of parallelization – band and k-point – are varied and combined to achieve the lowest CPU time (Fig. 1 and Table 2). It is found that, within a single node, band parallelization is the most effective method.

FIGURE 1: Comparison of parallelization methods for one full node

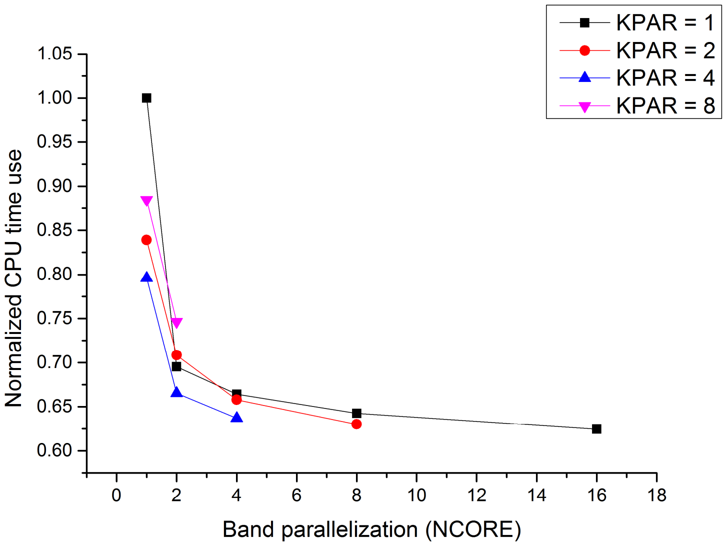

TABLE 2: Absolute timings (seconds) of different parallelization methods

KPAR \ NCORE	1	2	4	8	16
1	702.89	488.77	466.84	451.57	438.88
2	589.79	497.94	462.4	442.89	
4	559.44	467.55	447.75		
8	621.9	524.59			

To motivate whether it is actually beneficial to use an entire node for the surface calculations, the first cause of concern is the necessary memory. Already 8 GiB (see Table 1) of memory is needed for a 15-layer potassium slab unit cell. If the material has the hexagonally closely packed (HCP) crystal structure, needs a dense grid in reciprocal space and has a PAW with more valence electrons (Tl, for example), half a node will not suffice. Additionally, we show in Fig. 2 (and Table 3) that the time gain for increasing core usage (all with band parallelization) within a node is nearly linear. This proves that the full-node use of the TIER1 cluster is indeed cost-effective.

FIGURE 2: The decrease of CPU time by increasing the number of cores

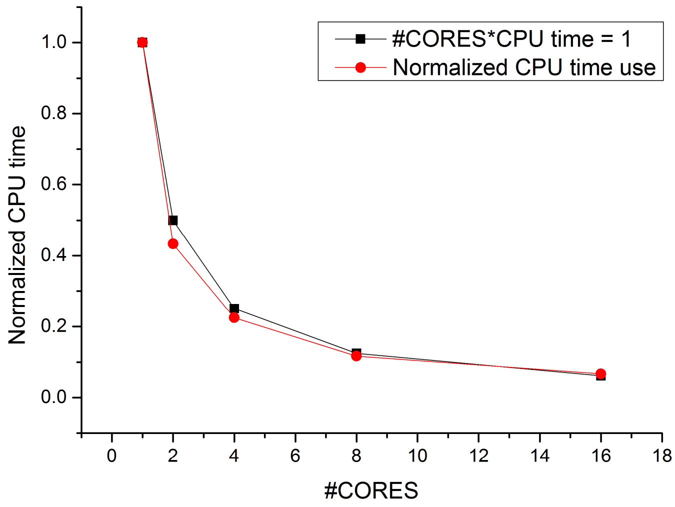

TABLE 3: Speed-up as a result of increased number of cores (band parallelization)

# Cores	Absolute time (s)	Speed-up
1	6499.39	1
2	2816.35	2.307735
4	1459.76	4.452369
8	761.41	8.535992
16	438.88	14.80904

10. Describe how you will manage the resources requested in the period during which the task is to be performed. What usage pattern do you anticipate (similar usage on monthly basis, bursts, …)?

For every of the 16 types of crystal structures in this project, the workflow starts with convergence testing for k-points (3*4*15 = 180 jobs launched at a time), then testing for slab thickness (180 jobs at a time) and finally testing for amount of layers required for full relaxation (6*3*4 = 72 jobs at a time). When the optimal settings are derived from those tests, relaxations and final point calculations can be performed for all materials with that crystal structure.

This work flow results in the launching of large batches of jobs, a large part of which will finish quickly. These periods of heavy usage will be followed by periods of moderate usage in which the more demanding calculations will be finished and idle periods in which the tests are evaluated.

We expect this process to take about 3-4 months for all 16 crystal structures (53 materials), resulting in 318 surface energies and 318 work functions.

11. List the granted computing time allocations to the promoter(s) of this research project, on the Flemish Tier-1 system, as well as other Tier-1 and Tier-0 systems. Also, describe the scientific output obtained within the framework of computing time that was granted during the past two years on the Flemish Tier-1 or on other Tier-1 or Tier-0 supercomputers. DOI links are sufficient.

Ongoing projects:

Computational Discovery of Quaternary Zintl Phases (3076 node days) T. Crepain

Finished projects:

Ab initio prediction of acoustic anisotropy of Fe, Ni, and FeNi in the Earth's inner core (3106 node days) J. Jaeken

Assessing the accuracy of a screened hybrid functional for property predictions of elemental solids (1742 node days) K. Lejaeghere

Hunting for new quaternary Zintl phases of the 1:1:1:4 stoichiometry (1000 node days) K. Dumon

High-throughput determination of vacancy trapping enthalpies for the improvement of electronic device production (4420 node days) M. Sluydts

High-throughput screening of ternary tungsten alloys with DFT (4000 node days) K. Lejaeghere

Exploring surface doping and adsorption in semiconducting systems using a high throughput ab initio methodology for applications in nanoelectronics and photonics (4680 node days) M. Sluydts

Scientific output:

DOI: http://dx.doi.org/10.1021/acsnano.5b06965

DOI: http://dx.doi.org/10.1021/ja509941g

DOI: http://dx.doi.org/10.1080/10408436.2013.772503

DOI: http://dx.doi.org/10.1103/PhysRevB.89.014304

DOI: http://dx.doi.org/10.1103/PhysRevLett.111.075501

12. Are the applicants of this application bound by a confidentiality agreement? If so, the title and the abstract of this application will not be published on the website of the FWO / Flemish Supercomputer Center.

no

Should you have any questions or encounter any difficulties during the electronic submission of an Application, please contact by e-mail:
Associatie KU Leuven: hpcinfo@kuleuven.be
Associatie Universiteit Gent: hpc@ugent.be
Associatie Universiteit Hogescholen Antwerpen: hpc@uantwerpen.be
Associatie Universiteit Hogescholen Limburg: geertjan.bex@uhasselt.be
Universitaire Associatie Brussel: rosette.vandenbroucke@vub.ac.be
For the other institutions: caroline.volckaert@FWO.be

SOFTWARE LICENSE AGREEMENT FOR THE USE OF VASP5.2 BY ACADEMIC INSTITUTIONS

The Universität Wien, Austria (UW in the following) and
Ghent University, Belgium (UG in the following) [1]
conclude the following agreement:

(1) The UG acquires a non-exclusive academic license for the use of the software-package VASP (Vienna ab-initio simulationprogram) for ab-initio local-density-functional total-energy and molecular-dynamics calculations, versions VASP5.2 and VASP4.6, by the research group Functional Nanomaterials (FUNNANO)[2]. Under this licence the use of the software is restricted to a maximum of six researchers or students, all belonging to this research group and to the same organisatorial unit and working at the same location. The licence does not cover the use of VASP by external collaborators working at other institutions.

(2) The license covers access to the source-code, the program documentation and to the data-base for ultrasoft pseudopotentials and PAW-potentials. UW reserves the exclusive property of the software. It declines any liability for the software and any responsibility for the results of calculations produced with the program. The license does not cover any maintenance service for the software or support for its implementation.

(3) The license is not transferable to another research group of UG without the written agreement of UW. UW reserves the right to refuse authorization of such a transfer. A transfer to a research group not belonging to UG is excluded.

(4)The UG guarantees that the software or parts thereof shall not be made accessible to third parties without the explicit written consent of UW. Access to the code and to the data-base shall be made available through an account of the UW. The UG guarantees that the password for this account will be known only to one contact-person and shall not be communicated to temporary co-workers or guests. All installations of the source code, the executable or the data-base must be copy-protected and accessible only to the authorized users.

[1]Please insert here the name of the institution concluding this agreement with UW. This institution must be a legal person and the agreement must be signed by an authorized representative of this institution. Define the acronym (replacing .....) under which this institution is referred to in the text of the agreement.

[2]Please insert here the name and affiliation research group for which the license is acquired

SOFTWARE LICENSE AGREEMENT FOR THE USE OF VASP5.2 BY ACADEMIC INSTITUTIONS

(5) If VASP is used as the basis of further methodological or software-development, UG agrees to make these additions available to UW. UW will also be entitled to include these additions in further releases of VASP.

(6) In future publications of work performed using VASP, the use of the software shall be properly acknowledged, e.g. in the form
"The calculations have been performed using the ab-initio total-energy and molecular-dynamics program VASP (Vienna ab-initio simulation program) developed at the Institut für Materialphysik of the Universität Wien [1-3]."
[1] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11 169 (1996).
If the PAW-version is used, reference will be made to
[2] G. Kresse and D. Joubert, Phys. Rev. **59**, 1758 (1999).
If special features implemented in VASP will have been used, reference should be made to the relevant publications as listed on the VASP home-page.

(7) The UG accepts to pay to UW a licence fee Euro 4.000,- (fourthousand Euro). The licence fee is strongly discounted and applies only to academic institutions with undergraduate teaching.

(8) The licensee will use VASP exclusively for non-profit research. If VASP is used in contractual research in cooperation with or for industry or for military institutions, the financial conditions will have to be re-negotiated.

(9) UW declares that it has the full power and authority to grant the rights granted in this agreement without the consent of any other person, and that the license and use of the software by the licensee will not in any way constitute an infringement or other violation of any copyright, proprietory right or any other rights of any third party.

(10) Any disputes arising from the license agreement are subject to the laws of the Republic of Austria.

(11) The terms of this agreement shall prevail any terms or conditions of the licensee.

SOFTWARE LICENSE AGREEMENT FOR THE USE OF VASP5.2 BY ACADEMIC INSTITUTIONS

For the Universität Wien:

Jürgen Hafner
Fakultät für Physik, Universität Wien
Sensengasse 8/12, A-1090 Wien, Austria

Date

For the UG

Name (in print): Michel Waroquier
Institution: Faculty of Sciences, Ghent University

Address: Technologiepark 903, BE-9052 Zwijnaarde, Belgium

Date: 26 January 2010

For the research group entitled to use VASP5.2:

Name (in print): Veronique Van Speybroeck (FUNNANO)