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:

Analysis of the off-stoichiometry of Fe_3N_{1+y} with Density-Functional Theory in the Fe-N system.

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:

4320

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

4378

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 accurate prediction of bulk free energies is essential in computationally-driven materials design. Within the framework of density-functional theory, the quasi-harmonic approximation (QHA) has become a very common method to calculate the vibrational contribution to these free energies. The goal is to construct phonon spectra at different volumes, thus account for thermal expansion and contraction. The equilibrium volume can be derived at each temperature by minimizing the free energy with respect to the volume. Especially for temperature below ½ of the melting point, the QHA has shown to be very successful.

Taking the application of the QHA one step further than bulk calculations, we intend to calculate the free energy of different nitrogen concentrations. As such, we are able to determine the stability of different off-stoichiometries of the Fe_3N nitride, where the atomic concentration is often higher than 25%, the thermodynamically most stable concentration. In the framework of making reliable predictions with DFT calculations, this work will mainly focus on how the numerical imprecisions present in the entire QHA approach affect the interstitial free energy. This builds on our similar previous tier-1 project (see below), which focused on $Fe_{16}N_2$ and Fe_4N. That project is now in its final stages. In the case of Fe_3N, however, non-dilute concentrations of additional nitrogen interstitials will be present. This requires a wider variety of different crystal structures to be evaluated.

We will be using VASP with the exchange-correlation functional PBE to obtain forces from atomic displacements, which are needed to construct the dynamical matrix from which the phonon spectra are derived. For the setup of the unit cells where an atom has been displaced the phonopy software package is used. Also, in the first post-processing step, phonopy is used to extract the forces from the calculations, construct the dynamical matrix and diagonalize it.

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 has become one of the most prevalent methods to determine the structure and stability of crystalline solids. To obtain temperature-dependent free energies, however, which are essential in real-world materials science, the excitation mechanisms have to be taken

explicitly into account. The quasi-harmonic approximation (QHA), where phonon spectra are calculated at different volumes to account for thermal expansion or contraction, has become a very interesting approximation as a trade-off between accuracy and computational resources. The aim for this project is to assess the numerical precision of vacancy free energy predictions for the Fe_3N_{1+y} nitride.

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
3 years of experience with Wien2k, VASP, FLEUR, gaussian and CP2K on Ghent TIER2, Tier1 Muk, Tier1 BreNIAC (Pilot phase participator)

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).

Obtaining vibrational free energies within the quasi-harmonic approximation for crystals with a large unit cell, as is necessary for point defects, requires

large amount of computational resources. However, a high degree of parallelization is possible. To apply the quasi-harmonic approximation, we need to obtain phonon spectra at different volumes. The different volumes can be handled in parallel. A single phonon spectrum, at a specific volume, is constructed by performing a number of symmetrically inequivalent atomic displacements. The force response for all these displacements is obtained from different VASP calculations, and thus fully parallelized. Finally, as can be seen in the scaling tests (see further), the VASP calculations for the supercells needed in this work show a very good efficiency for parallel calculations. Intranode scaling is almost optimal and scaling up to 16 nodes has an efficiency of better than 77%.

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.

   (max. 1 A4 Arial 12).

   All calculations are very similar, the only difference is the atomic displacement and the volume. We have chosen an upper bound of 9 hours walltime, based on calculations on an Fe_3N supercell, where symmetry was completely broken by displacing 6 atoms. We aim to investigate 5 different stoichiometries: Fe_3N, $Fe_3N_{0.903}$, $Fe_3N_{0.8}$,

$Fe_3N_{1.091}$ and $Fe_3N_{1.176}$. They require an average of 384 displacements to obtain the phonon spectrum, which has to be done for 6 different volumes. This amounts to a total number of 11520 calculations, each one requiring 1.9GB. We will have to trim the excess output off these calculations to limit the required storage and move them to tier 2 for post-processing quickly. Nevertheless, batches of 6*384 are tough to separate. Once the calculations are moved, the only sizeable file that requires storage is the vasprun.xml file (Tier 2, 1002 MB). After post-processing a specific stoichiometry, these can be compressed and moved to data. The 2.3 TB storage needed on the VO Scratch will be the maximum requirement.

	Walltime	Node days	Max Memory	# cores per node	#nodes per task	Tier2 Storage	Tier 1 storage
1 Force calculation	0.375 days	0.375 days	28*2.8 = 78.4	28 (full)	1	1.002 GB	384*1.9*6 GB
Total (5*6*384 = 11520)	4320 days	4320 days				384*6*1.002 = 2308 GB	4378 GB

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).

VASP 5.4.1

- http://www.vasp.at

- License: attached at the end of the document

- The software is already available on the TIER1 infrastructure.

We have performed scaling tests for a Fe_3N 64-atom supercell, where 6 atoms have been displaced to break the symmetry. This system is a very good test-case for the force calculations that will be performed.

Good intranode scaling was observed, with an efficiency of 73% when moving from 4 cores to full node (figure 1 and figure 2). Using fewer cores to map the full intranode behavior was impossible because of memory requirements. . This makes our calculations suitable for tier 1 usage.

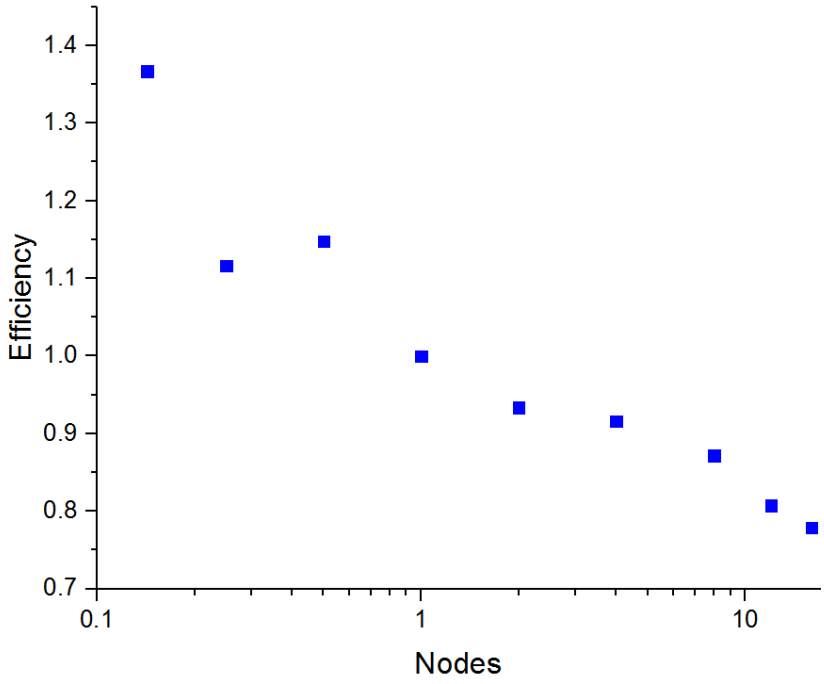

Figure 1: Efficiency (cpu time use multiplied by number of cores used, normalized to a full node) as a function of number of nodes. NCORE was set to number of cores (with a maximum of 14) and KPAR was set to the number of nodes.

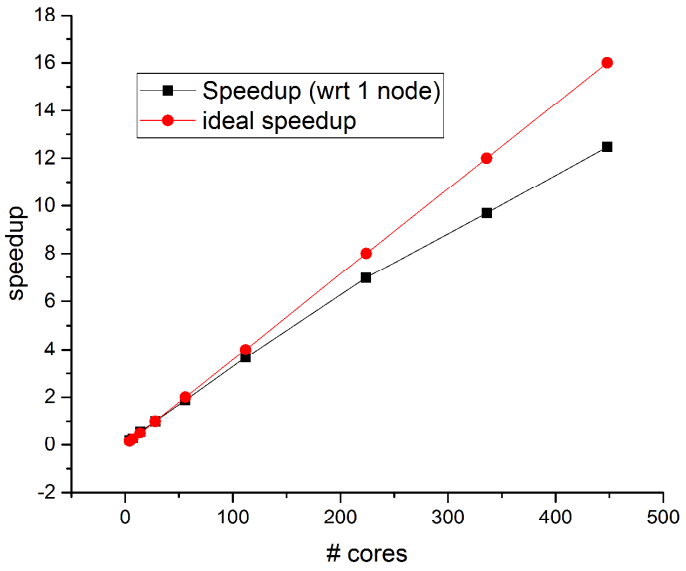

Figure 2: Speedup as a function of used cores.

Although the longest of our calculations will only take 9 hours, it is also possible to speed them up at a very good efficiency (Figure 1 and Figure 2). Using 4 nodes yields an excellent efficiency of more than 90%. Scaling tests are summarized in table 1.

# nodes	# cores	Wall clock time (hours)	Speedup (with respect to 1 node)	Efficiency
1	4	44.65	0.195520717	1.368645017
1	7	31.23	0.279538905	1.11815562
1	14	15.19	0.574720211	1.149440421
1	28	8.73	1	1
2	56	4.67	1.869379015	0.934689507
4	112	2.38	3.668067227	0.917016807
8	224	1.25	6.984	0.873
12	336	0.9	9.7	0.808333333
16	448	0.7	12.47142857	0.779464286

Table 1: Summary of the scaling tests for the Fe_3N supercell system

To determine the optimal use of a single node, we carried out parallelization testing: both band parallelization (NCORE setting) and k-point parallelization (KPAR setting) were tested. The former increases memory efficiency and yields a moderate speed-up, whereas the latter yields a better speed-up, but requires quite a lot of extra memory.

KPAR/NCORE	1	2	4	7	14	28
1	18.43	11.24	9.96	9.72	**8.71**	9.18
2	14.68	10.69	N/A	9.38	**8.43**	x
4	N/A	N/A	9.09	8.99	x	x

Table 2: CPU timings (hours) for different parallelization settings on 1 node

KPAR/NCORE	1	2	4	7	14	28
1	4.16	3.46	3.05	2.87	**2.8**	2.78
2	5.97	5.2	N/A	4.66	**4.61**	x
4	N/A	N/A	8.21	8.2	x	x

Table 3: Memory usage (GB) for different parallelization settings on 1 node

The fastest setting is NCORE=14 and KPAR=2 (Table 1). However, because of the nature of k-point parallelization, the memory cost is quite high (Table 2) in comparison to the mild speedup from NCORE=14 and KPAR=1. Because we would prefer to use less than 4GB of memory per core, setting KPAR=1 will be the preferred route. Mainly because the number of +4GB memory nodes on BrENIAC is limited.

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, …)?

The usage will come in bursts. All the calculations for a set of phonon spectra will be submitted at once (scripted). This will result in hundreds of multi-node jobs being requested at once, that will work their way through the queue. If this workflow is problematic for the schedulers on Tier 1 BrENIAC, the cluster operators can certainly request us to adapt it. With a little more management, the submit process can be broken down in smaller batches of jobs.

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.

Finished projects:

Error estimates for ab initio predictions for point defects in the Fe-N system. (3022) S. De Waele

High-throughput screening of an unknown quaternary crystal space (4056) M. Sluydts

Error estimates for ab initio predictions of surface energy and work function (1000) S. De Waele

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

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

Scientific output:

DOI: https://doi.org/10.1103/PhysRevB.94.235418
DOI: http://dx.doi.org/10.1021/acs.chemmater.6b03368
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