Enclosure 1b. Category 1 Application form 2015 – 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 Dutch on the website https://vscentrum.be/.

Title of the application: Hunting for new quaternary Zintl phases of the 1:1:1:4 stoichiometry

Name and first name of the applicant:

Karel Dumon

Institution:

Ghent University

Research group / department:

Center for Molecular Modeling

Title / position:

Master student (faculty of engineering and architecture)

email address:

karel.dumon@ugent.be

stefaan.cottenier@ugent.be

Total computing time that is needed, in node days: 1000

Total disk storage that is applied for (in GiB): 1843 GiB

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

I-III-VI_2 Semiconductor Nanocrystals, from a fundamental understanding of the hot injection synthesis to a novel colloidal quantum dot material family.

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):

A large majority of all possibly existing binary solids have effectively been discovered and described. For ternary solids, the estimated number of unknown ones is still larger than the number of known ones. With quaternary solids, one enters a Terra Incognita where only a negligible fraction of all possible solids has been described. This is nicely illustrated by the so-called Zintl phases – crystals formed by combining s- and early p-elements, showing a chemical bond that has characteristics of ionic as well as of covalent bonding. It is a combinatorial exercise to find that 7840 different quartets can be formed when picking elements from the s- and early p-groups of the periodic table. For each quartet, several plausible valence-balanced stoichiometries can be suggested, and for each stoichiometry several (as yet undetermined) plausible crystal structures are possible. Even with conservative estimates for the latter two degrees of freedom, one rapidly ends up with a search space that contains several hundreds of thousands of potentially existing quaternary Zintl phases (QZP).

How many QZP are experimentally known? According to the Inorganic Crystal Structure Database, which contains 160.000 experimentally known crystals, as few as 50 can be labeled as a true QZP. Only a couple of new QZP are discovered each year.

The present proposal aims at dramatically speeding up the QZP discovery rate, by performing a high-throughput screening for potentially stable QZP in the class with 1:1:1:4 stoichiometry and the $CsPbPSe_4$ crystal prototype (6 of such QZP are currently experimentally known, which makes it the largest quaternary type). We want to determine by quantum simulations the formation energy of all 4608 possible QZP of this type, in order to identify the most stable ones for subsequent experimental synthesis attempts. This combined computational/experimental strategy should be much more efficient compared to a purely experimental approach, and is expected to boost the number of experimentally known QZP in the near future.

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

There is a large group of less well-known solids that share properties with the ionic solids as much as with the covalent solids: Zintl phases. Scientists have reasons to expect interesting and useful materials among the Zintl phases built from four (or more) elements – the quaternary Zintl phases (QZP). The 50 true QZP that are experimentally known are negligible compared to the hundreds of thousands of QZP that might potentially exist. The major bottleneck is that there are so many possibilities to examine. The goal of this project is to boost the discovery rate of QZP by searching through this huge space in the computer rather than in the lab, such that future experiments can concentrate on the most promising cases.

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 4 "EasyChair proposals submission procedure").

FWO project G.0760.12

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

Prof. Dr. Stefaan Cottenier (stefaan.cottenier@ugent.be)

5. Billing address to which the payment invoice will be sent to:

Stefaan Cottenier
t.a.v. Wim Dewitte
Center for Molecular Modeling
Technologiepark 903
9052 Zwijnaarde

6. 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 of using HPC resources in the past (Tier0/Tier1/Tier2 infrastructure in Belgium and abroad)

Dumon Karel
Ghent University
EA17 / Center for Molecular Modeling
Master thesis student
Experience with the HPC UGent TIER2 clusters (during the previous academic year in the context of courses, and intensively during the first semester of the present academic year for his master thesis research)

ir. Sluydts Michael
Ghent University
Center for Molecular Modeling
PhD Fellow
Three years of experience using the VASP and wien2k ab initio packages on both TIER2 and TIER1 (vsc40479)

Dr. ir. Lejaeghere Kurt
Ghent University
Center for Molecular Modeling
BOF postdoctoral fellow
experience with local CMM clusters, local clusters of the Computational Materials Physics Group of the University of Vienna, HPC UGent TIER2 clusters and the TIER1 machine muk (vsc40323)

Prof. Dr. Cottenier Stefaan
Ghent University
Center for Molecular Modeling
Assistant professor
experience with local CMM clusters and with Leuven and HPC UGent TIER2 clusters (vsc40026)

7. Explain why this project needs to run on a Tier1 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 proposed work consists of 4608 similar jobs, which require 0.2 node days each. The work flow of such a job has been thoroughly tested on TIER2. We are now ready for production runs. Due to the relatively large number of jobs, this is not feasible on TIER2. The proposed work will be the core of the master thesis of Karel Dumon (to be defended in June 2015), and of the publication based thereupon. The present rather modest request for TIER1 time is explicitly meant as an exploration. If the results of this search turn out to be positive, larger TIER1 proposals will follow to explore much larger regions of the QZP search space.

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 Tier2 DATA/HOME partitions and the Tier1 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 Tier1 (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).

	Node day calculation				
Computational task	# of such tasks	Wall clock time (days) per task	# Tier1 nodes per task	# node days per task	# CPU cores per task
Course volume optimization	4608	0.01	1	46	16
Full geometry optimization	4608	0.15	1	691	16
Structural finetuning and postprocessing	4608	0.04	1	184	16
Buffer for individual cases with problems				79	
total				**1000**	

The memory usage per task is 2 GiB (per node)

MPI-parallelization over a full node is used (see scaling test), no vSMP needed

Temporary storage: up to 10 GiB per task, needed only during the execution of that task.

Permanent storage: at most 0.4 GiB per task, leading to 1843 GiB for the entire project. A considerable part of this can be deleted after the final data analysis.

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

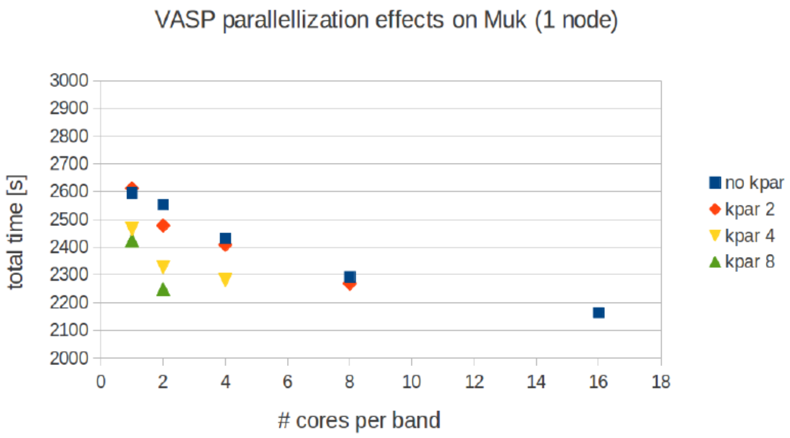
   - a reference to the software's web page
   - the software license system (open source, GPL, etc.)

- if there is no free academic use of the software, state which license makes the installation and the use valid on the Tier1 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 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)

Provide the results of scaling tests that were conducted with this software, preferably on Tier1 (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
http://www.vasp.at/
non-exclusive academic licence (see attachment)
available on TIER1

A lot of experience about the scaling behavior of VASP on MUK is available, for various types of jobs and levels of parallelization. The picture underneath shows a scaling test for a job of the kind that is used in this proposal. Two types of parallelization are combined with each other, in different ways: MPI parallelization and k-point parallelization (=distributing the job into almost independent serial parts). This test shows that the shortest wall-time is achieved for full MPI-parallelization over all 16 cores of one node. As the major challenge in this proposal lies in the large number of short-duration jobs (0.2 node-days if parallelized over 16 cores), it is useful in this case to parallelize more.



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, …)?

    A steady flow of jobs is expected, concentrated in the first 2 months.

11. List the granted computing time allocations to the promoter(s) of this research project, on the Flemish Tier1 system, as well as other Tier1 and Tier0 systems. Also, describe the scientific output obtained within the framework of computing time that was granted during the past two years on the Flemish Tier1 or on other Tier1 or Tier0 supercomputers. DOI links are sufficient.

The people involved in this proposal have been granted TIER1 time for several other recent research topics:

> **Ab initio screening of suitable tungsten alloys as first wall material in nuclear fusion reactors** (4000 node days)
>
> Mechanical properties of lots of hypothetical binary and ternary tungsten-based alloys are calculated from first principles, in order to identify those alloys with the most suitable combination of properties to serve as a first wall component in future fusion reactors. Parts of this study have been published in the PhD thesis of Kurt Lejaeghere. New publications on a larger dataset are in preparation.

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

Site selection and interaction with vacancies were calculated from first principles for 71 elements of the periodic table as impurity in the industrially important semiconductors Si and Ge. This leads to insight in the systematics of impurity-vacancy interaction, and complements the only partial experimental knowledge. Interpretation of this dataset allowed to single out a few impurity elements that can have a beneficial effect on lowering the vacancy concentration during germanium single crystal growth. A publication (first author Michael Sluydts) has been submitted.

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 Hercules Foundation / 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: marc.luwel@herculesstichting.be

Appendix: Example tables and plots

Table 1

# nodes	# cores	absolute timing (s)	speedup	# cores x timing
16	256	189.6	1.0000	
32	512	99.0	1.9154	
64	1024	55.6	3.4088	
128	2048	30.8	6.1376	

Plot 1

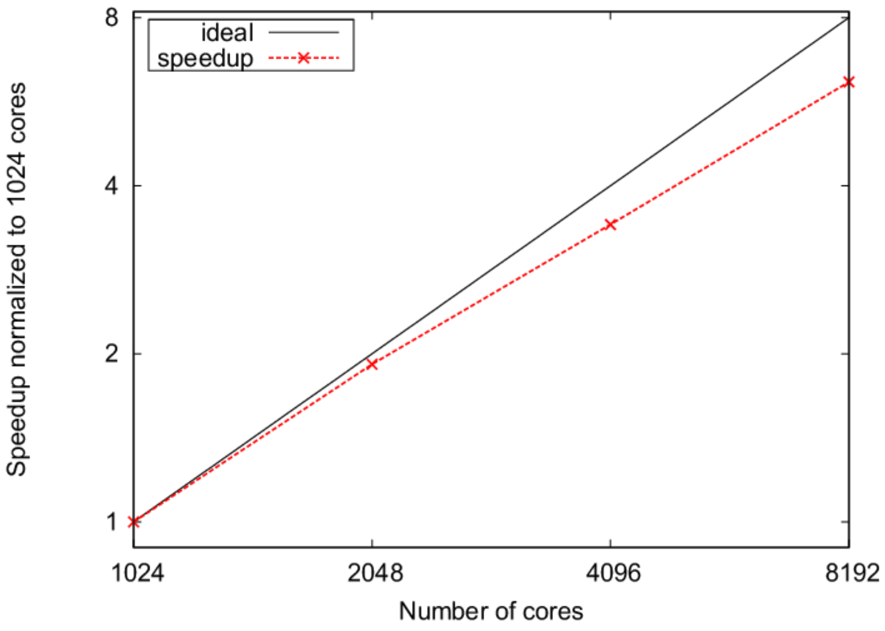


Table 2

	Node day calculation							Storage volume estimate	
Computational task	# of such tasks	Wall clock time (days) per task	# Tier1 nodes per task	# node days per task	# CPU cores per task	Memory usage (GiB) / node per task	OpenMP / MPI / hybrid / vSMP	Tier2 DATA/HOME volume (GiB) + number of files	Tier1 SCRATCH volume (GiB) + number of files
Task1	A	B	C	= A x B x C					
Task2	A	B	C	= A x B x C					